To the Graduate Council:

I am submitting herewith a dissertation written by Sheridan Clinton Barker entitled "A Study of the Perceptions of Belmont University Faculty About the Evaluation of their Teaching." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Leadership Studies in Education.

Norma T. Mertz, Major Professor

We have read this dissertation and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and

Dean of the Graduate School

A STUDY OF THE PERCEPTIONS OF BELMONT UNIVERSITY FACULTY ABOUT THE EVALUATION OF THEIR TEACHING

A Dissertation

Presented for the

Doctor of Education

Degree

The University of Tennessee, Knoxville

Sheridan Clinton Barker

August, 1995

DEDICATION

This dissertation is dedicated to my wife, Vicki; my children Lauren and Bradley and to my parents, Lorraine and Hubert. Their longsuffering, patience, understanding and love have made this all possible. They have made numerous sacrifices of time and resources which have enabled me to pursue this dream. I will be forever grateful.

ACKNOWLEDGEMENTS

Several people have provided the assistance and support which have made the completion of this project possible.

Sincere appreciation is extended to the writer's doctoral committee who have both individually and collectively been willing to encourage and guide: to Dr. Malcolm McInnis, who served as my adviser in course selection and who has been nothing but supportive in this dissertation process; to Dr. Herb Howard, whose kind words and thoughtful comments meant the world to me; to Dr. Kathy Davis, who has been an inspirational instructor and the best of constructive critics; and especially to Dr. Norma Merta, major professor, for her tireless patience, ready wit and thoughtful tutelage. She is a mentor in the truest and finest sense of the word.

Grateful acknowledgement is also made to my secretary, Anita Newport, whose help with the Mini-Tab computer program was my salvation.

ABSTRACT

This study was undertaken in order to investigate faculty attitudes toward student, administrative, peer and self-evaluation of their instruction by examining the attitudes of higher education faculty at one institution in which the four types of instructional evaluation were in use. The study sought to reveal their perceptions on each type and to compare their perceptions of how each type of evaluation was used versus how each evaluation type should be used in the ideal.

The investigation involved a survey of 131 full-time instructors at Belmont University, Nashville, Tennessee, which had an evaluation system in place that encompassed the four most widely used types of evaluaton--student, administrative, peer and self-evaluations. None of the Belmont instructor population surveyed devoted more than one-third of their time to administrative duties. Ninety-two surveys were returned, 89 of which were usable, providing an overall response rate of 67.9 percent.

Major findings of the study revealed that:

1. Belmont faculty were slightly more dissatisfied than satisfied with the way evaluation was conducted, citing inconsistency in execution of three of the four types of evaluation, and overemphasis on the fourth type, student evaluation. Dissatisfaction also stemmed from misgivings about the objectivity and reliability of student and peer evaluations. Nonetheless, Belmont faculty believed that all four types of evaluation should be conducted.

2. Belmont University faculty indicated that they were not certain if student, administrative and peer evaluation led to instructional improvement, doubted that

student, administrative, peer and self-evaluations led to improved programs/majors,and questioned the efficacy of every type of evaluation in making administrative decisions save administrative evaluation.

3. Discrepancies existed between how Belmont faculty perceived evaluation was used versus how they believed it should be used, and in the perceived disposition of information gleaned from the four types of evaluation.

4. Most Belmont faculty responses were consistent with the extant literature, except in the perception that administrative evaluation was objective, in the perception that administrative evaluation would foster instructional improvement, and in the perception that peer evaluation was fair, objective, reliable, and effective in improving teaching, improving programs/majors and in making administrative decisions.

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

CHAPTER I

OVERVIEW OF THE STUDY

Introduction

Evaluation of teaching is almost as old as teaching itself. As far back as Antioch (AD 350), a father dissatisfied with the teaching his son was receiving could file a complaint and ultimately transfer his son to another teacher if the original instructor was found derelict in his duties (Doyle, 1983). Between this ancient example and the 20th century, while evaluation of teaching in higher education has been viewed as important (Eckert, 1950), necessary (Wilson, 1942), even as a critical component in affirming the quality of teaching (Doyle, 1983; Sheehan, 1975), there is little evidence of systematic evaluation of teaching in higher education (Marsh, 1989), and almost no examinations of the perceptions of higher education faculty about evaluation of their teaching.

Interest in evaluation of faculty in higher education began as a response to inquiries by parents and legislators in the 1940s and 1950s about the quality, efficiency and soundness of instruction (Wilson, 1942; Stuit & Ebel, 1951). Earliest approaches to faculty evaluation in higher education were conducted by multi-member faculty evaluation committees or administrative evaluation committees which sought to affirm the quality of higher education instruction (Eckert, 1950). Ultimately, individual student evaluations of instruction were added as an evaluation approach in the hope that instructional improvement could be confirmed by the instructional "consumers" themselves (Stuit & Ebel, 1952).

With burgeoning college and university enrollments in the 1960s and 1970s, interest in faculty evaluation heightened as student activists and even legislators stressed accountability, and colleges and universities sought to confirm the quality of their instruction (Kerlinger,1971). To the emphasis on evaluating teaching for instructional improvement was added evaluation for administrative decision making (Stufflebeam, 1969; Cronbach, 1963). The use of student evaluations for personnel decisions increased dramatically during the 1970s (Kirschling, 1978), and the emphasis on quality and accountability in public primary and secondary education (Goodlad, 1983) was carried over into higher education (Boyer, 1987).

The two primary purposes of evaluation of instruction in higher education in the 1980s were instructional improvement and administrative decisions (Marsh, 1987), and the situation in the 1990s shows no signs of changing.

Historically, evaluation of higher education faculty teaching has been attempted in one of four ways, or in a combination of the four. They are: student evaluations, administrator evaluations, colleague or peer evaluations and self-evaluations. These four ways differ in frequency of use, and in the nature and quantity of research available.

Student Evaluations

Student evaluations have been the type of evaluation most often used (Brandenburg & Ory, 1981), and the most frequently and systematically researched (Centra, 1981; Seldin 1984; Braskamp, Brandenburg & Ory, 1984). Marsh (1987), in reviewing literature on student evaluation of instruction from the 1920s to 1987, cited over 2,000 studies on student evaluation--its uses, validity and credibility, and

estimated from those studies, that student evaluations have been done or are now being done at close to 90 percent of American colleges and universities.

Student evaluation is the appraisal of a course and/or instructor by students, usually done on a class-by-class basis, at the end of the college term (Seldin, 1990). The most common format for student evaluation is a rating form which asks students to rate the instructor's knowledge of subject, organization/presentation skills, enthusiasm and adequacy of relations with students (Simpson, 1966). The primary purpose of student evaluation is reputed to be instructional improvement (Costin, Greenough and Menges, 1971), but there is evidence that student evaluations have been used for administrative decisions on raises, retention and promotion of professors, and even on tenure decisions (Seldin, 1973).

Interestingly enough, given the frequency of their use, there have been no direct studies of faculty perceptions of student evaluations, and what is "known" about their perceptions is drawn largely from incidental comments and anecdotal materials. These comments and anecdotal materials suggest that while some instructors believe student evaluations can be leading indicators of effective teaching (Benton, 1982; Wilson & Gaff, 1975), others believe students are not capable of discerning good teaching (Eble, 1971) and that student evaluations stop short of measuring long-term effectiveness of instruction (Dressel & Marcus, 1982; Seldin, 1984).

Administrator Evaluations

Administrator evaluations, the second most frequently used type of faculty evaluations, have been on the increase since the beginning of the last decade (Bolton, 1983), occurring about half as often as student evaluation. Administrator evaluation is

the appraisal of teaching and course effectiveness by department or college heads, deans, even presidents, accomplished generally through classroom visits and narrative written assessments (Seldin, 1984). The stated purpose of administrative evaluation of higher education instruction is instructional improvement (Hughes, 1975), but increasingly, institutions are reporting the use of administrative evaluation for the purpose of administrative decision-making, e.g. tenure, promotions, raises and even termination (Doyle, 1975).

Research on faculty reaction to administrator evaluations is sparse, and what research does exist is basically anecdotal (Goodwin & Smith, 1983). While believing that administrators, as trained professionals, are more capable of informed evaluations than students (Hughes, 1975), college and university faculty appear to view administrator evaluations with suspicion as potentially punitive because of their use in personnel decisions such as denial of tenure and/or promotion and as justification for small raises in response to poor evaluations (Seldin, 1984).

Peer Evaluations

Peer or colleague review has gained in acceptance and use, but ranks third in frequency behind student and administrative evaluation (Whitman & Weiss, 1982). Peer or colleague evaluation is the appraisal of instructional effectiveness of one professor by another or several professors. The peer evaluator either visits the class(es) of a colleague and makes, normally, a narrative assessment of teaching by direct observation, or examines the teaching materials of a colleague, e.g. tests, syllabi, handouts and texts (Whitman & Weiss, 1982), or both. The expressed purpose of peer evaluations is instructional improvement (Ferren & Geller, 1983), although there have

been instances in which peer review has been used for the purpose of administrative decisions (Kulik & McKeachie, 1975).

There is a dearth of research on faculty perceptions regarding colleague or peer evaluation, and what is known about peer evaluation comes largely through secondary anecdotal information (Cross, 1986). A few studies have noted negative faculty perceptions, including the idea that faculty feel hurt by negative feedback from colleagues (DeNisi, Randolph, & Blencoe, 1983). The seriousness with which colleague raters approach their task heads the list of known positive reactions of faculty regarding peer evaluation (Doyle, 1983).

Self-Evaluations

Self-evaluation of teaching by higher education faculty ranks fourth in frequency of occurrence (Marsh, 1987). Self-evaluation is the instructor's self-diagnosis of the effectiveness of his/her teaching as accomplished by responding to specific written questions about teaching effectiveness (Simpson, 1966). The purposes of self-evaluation are to improve teaching and learning, to help combat instructor burnout and dropout, and to help instructors more clearly define their professional identities (Simpson, 1966).

While it has gained in popularity among faculty, self-evaluation remains controversial (Seldin, 1984). Except for a few studies of the efficacy of self-evaluations in personnel decisions (Holzbach, 1978; Braskamp, Brandenburg, & Ory, 1984), studies of self-evaluation are virtually non-existent (Skipper, 1975), and what little is known about them has been gleaned from serendipitous anecdotal findings. These findings suggest that there are negative reactions to possible leniency of self-

evaluations (Holzbach, 1978) and its potential for being too subjective (Braskamp, Brandenburg & Ory, 1984), and positive reactions to it as a way of combatting potential burnout (Simpson, 1966).

As suggested in the literature, most of what is known about faculty reaction to student, administrator, peer and self-evaluation has been gleaned from anecdotal material. Such material suggests that acceptance of the four approaches/methods is low (Hughes, 1975; Marsh, 1987), and that there is a lack of consensus among higher education faculty about the purposes and uses to be made of evaluation (Ory , 1991), and the fairness of student and administrative evaluation (Young & Gwalamubisi, 1986).

As interesting and potentially valuable as this anecdotal material is, it does not allow for drawing conclusions about faculty reactions to evaluation. Certainly, faculty understanding and commitment to instructional evaluation is important if the purpose of improving quality of instruction is to be achieved. We know little, however, about higher education faculty understanding of or commitment to the evaluation of their teaching, and researchers have given little direct attention to these perceptions. What we need, but do not have, are studies which address, head-on and systematically, how faculty perceive evaluation, or how we might best go about identifying those perceptions.

The National Center for Research Into Post-Secondary Teaching and Learning (NCRIPTAL) studied the implications of performance appraisal on higher education faculty. They surveyed over 400 college and university professors, and found that faculty who were not consulted about the performance appraisal process felt that evaluation of their teaching was not necessarily fair and equitable. The NCRIPTAL study concluded that if faculty are to accept and act upon instructional evaluation, the performance appraisals used must be seen as credible in their eyes (NCRIPTAL, 1987).

The NCRIPTAL report also concluded that too little research had been done on faculty reaction toward evaluation.

If faculty evaluation in higher education is intended to enhance the quality of teaching; if that evaluation is to be fair and equitable; if that evaluation is to be credible in the eyes of the faculty themselves, and thus be more likely to be acted upon rather than discounted; then getting at these faculty perceptions about evaluation is important.

Statement of the Problem

A considerable body of literature has explored faculty evaluation in higher education from the standpoint of frequency and validity of student evaluations (Seldin, 1975, 1980, 1984; Centra, 1975; Braskamp, Brandenburg & Ory, 1984; Levinson-Rose & Menges, 1981; and Marsh, 1987). To a lesser extent, the literature has examined the frequency and credibility of administrator, peer and self-ratings (Doyle, 1983; Cross, 1986; DeNisi, Randolph & Blencoe, 1983; Bolton, 1983; Goodwin & Smith, 1983; Holzbach, 1978; Meyer, 1980). Few studies have explored the attitudes and perceptions of higher education faculty toward evaluation, its purposes, types, results and who is ultimately served by evaluation. And there are virtually no studies of faculty perceptions about how and what evaluation should be as compared to what it is.

Statement of the Purpose

The purpose of the study was to begin to examine the attitudes and perceptions of faculty toward instructional evaluation by investigating the attitudes and perceptions of higher education faculty at one institution, Belmont University, Nashville, Tennessee, which

had a teaching evaluation system in place which involved each of the four types of evaluation--student, administrative, peer and self. Thus, it was possible to examine the faculty's attitudes and perceptions of "what is," as well as what "should be," in terms of the four types of evaluation.

Research Questions

The research was guided by the following questions:

1. How do faculty at Belmont University perceive each of the four types of evaluation currently used to evaluate their teaching?

2. How do the faculty's perceptions of the approaches currently being used to evaluate their teaching compare with their perceptions of what should be used?

Importance of the Study

Research on evaluation of instruction in higher education has been conducted in some form for almost 70 years, but studies specific to the examination of faculty perceptions toward that evaluation have been category specific (i.e., student evaluation, peer evaluation, etc.) and any findings and conclusions about faculty perceptions have been, for the most part, secondary. This study addressed, globally and directly, faculty attitudes and perceptions toward evaluation, thus balancing the "all too often disproportionate amount of attention given to evaluating teaching by those not doing the teaching" (Marsh, 1987, p. 282).

By getting at the attitudes and perceptions of professors themselves, this research provides a perspective not currently available in the literature--a perspective that crosses disciplines and experience levels, and a perspective which is critical to the acceptance and usefulness of any faculty evaluation system. Further, the research contributes to the knowledge base about the usefulness of faculty evaluation as diagnostic feedback; about current use(s) of evaluation (e.g. tenure, promotion); and, finally, about how faculty perceive they are served by evaluation. While recognizing the limited generalizability of the findings, it provides those who seek to effectively plan and implement evaluation programs with critical information not hithertofore available.

Limitations and Delimitations of the Study

Limitations and delimitations of the study include:

a. This study was limited to one institution, Belmont University, Nashville, Tennessee, although it included the entire population of full-time teaching faculty and administrators with at least two-thirds teaching responsibility.

b. Since this study was limited to one institution, the findings are limited to that institution, and cannot be generalized to other institutions. Extreme care has been taken in drawing implications for faculty generally.

c. This study included new full-time faculty members who had not yet been evaluated under the current Belmont evaluation procedure.

d. This study did not attempt to ascertain the faculty's pre-Belmont experiences with evaluation.

Definition of Terms

Evaluation: Any type of systematic appraisal of the quality of college teaching for the purpose of teaching improvement or administrative decisions.

Faculty: Persons in an institution of higher education who have full-time teaching responsibilities, and/or who have some teaching and some administrative responsibilities, the latter not to constitute more than 33 percent of their duties.

Perceptions: Awareness of, predispositions or tendency to react specifically toward an object, situation, or value; usually accompanied by feelings and emotions (Goode, 1968). Perceptions cannot be directly observed, but must be inferred from thoughts, behaviors and attitudes.

Procedures

Since little is known about faculty perceptions of evaluation in higher education, the study was exploratory and descriptive in nature. The study was exploratory in that it only began to examine a problem for which there is little research, faculty perceptions of types of evaluation, and examined uncharted ground in comparing perceptions of evaluation procedures in use and perceptions of what evaluations should be and how these evaluations should be used. And it did so with the total faculty of only one institution. It is descriptive in that the subjects were asked to share their attitudes and perceptions, and their responses were used as descriptors of the phenomena under study.

Information and data were gathered through the use of a survey containing both open and closed questions. Survey research methodology was used to address the

research questions from the entire population of full-time undergraduate faculty (those who taught exclusively and those who had not more than one third-time administrative responsibilities) at Belmont University. The questionnaire consisted of demographic questions, questions regarding perceptions about the specific evaluation procedures currently used at Belmont, and questions about perceptions about how evaluation should be done compared to how it was done at Belmont. Groups of questions were followed by a space for more descriptive comments. Respondents were asked to mark boxes on the closed questions indicating their perceptions about fairness, objectivity and reliability of the four types of evaluation used at Belmont (student, administrator, peer and self), satisfaction with Belmont's evaluation procedures, what should be done with information gleaned through evaluation, and what faculty perceptions were concerning the purposes of evaluation.

The survey yielded mostly nominal data, although the open-ended questions yielded some qualitative informational/anecdotal material. A Mini-Tab computer program was used for all statistical analyses. The percentages of responses to each questionnaire item were calculated and chi-square tests performed to measure the difference in the observed and expected percentages at the .05 level of significance. Additionally, the chi-square statistic was used to determine if responses were significantly different due to demographic variables such as gender, age, academic rank, tenure status, teaching versus administrative status, school, highest degree earned and years teaching experience.

Organization of the Study

There are five chapters in this study. The first chapter presents basic information about the study: background, statement of the problem, the purposes, assumptions, importance of the study, organization, procedures, limitations and delimitations and definitions of terms. The second chapter provides a critical review of related literature on evaluation of instruction in higher education as practiced and perceived by faculty. The third chapter details the methodology and procedures used in the study. The fourth chapter contains the presentation and analysis of the collected data. The fifth chapter contains a summary of the study and its findings, conclusions and discussion of the findings, and recommendations for future research.

CHAPTER II

REVIEW OF RELATED LITERATURE AND RESEARCH

Introduction

The purpose of this study was to examine the attitudes and perceptions of higher education faculty about the evaluation of their teaching--how evaluation was conducted versus how it should be conducted, and what should be done with evaluation information. This chapter provides a critical review of the research and literature dealing with the evaluation of instruction in higher education. It is presented in four sections. The first section addresses definitions of evaluation and the extent to which institutions of higher learning conduct evaluation. The second section examines the need for and purposes of evaluation of teaching in higher education. The third section directs attention to the methods used in evaluating college teaching. The fourth section focuses on what is known about faculty attitudes toward evaluation of their instruction.

Evaluation Defined

Many definitions of evaluation can be found in the literature. Indeed, for approximately the last 50 years, scholars have attempted to define what evaluation of instruction is in higher education. Evaluation has been defined as: appraising faculty

service (Wilson, 1942); determining the extent to which educational objectives are met (Tyler, 1950); measuring instructional efficiency (Stuit & Ebel, 1951); finding evidence of quality classroom performance (Eckert, 1950); or of sound teaching (Stufflebeam, 1969); performance assessment (Cronbach, 1963); even competency assessment (Weimer, 1987).

Early efforts at defining evaluation were directed more toward identifying attributes of effective college teaching from the perspective of administrators. The consensus in early studies was that evaluation was a way of measuring instructional quality, e.g. how well instructors were teaching their subject matter, according to administrative standards (Wilson, 1942; Tyler, 1950; Stuit & Ebel, 1951; Eckert, 1950).

In his book, The Academic Man, a treatise on the responsibilities of the professoriate, Logan Wilson (1942), then President of the University of Texas, called evaluation the "appraisal of faculty service" as evidenced by good teaching and scholarship based on administrative review. Ralph Tyler, in his book on curriculum and instruction, perceived evaluation to be "the process of determining to what extent educational objectives are actually being utilized," (Tyler, 1950, p. 69) as ascertained by administrators singly or collectively. Stuit and Ebel (1951), reporting the findings of an administrative committee at the State University of Iowa which had polled 35 college and university academic administrators about what constituted evaluation, concluded that evaluation was a way of measuring efficiency of instruction, and in this, there was a general belief that students, as well as administrators, could be helpful in determining instructional efficiency. Eckert (1950), a professor of higher education at the University of Minnesota, suggested that a clear definition of faculty evaluation for

higher education was evidence of quality performance in the classroom as determined by an administrative committee.

In the 1960s, efforts at defining evaluation were directed more toward identifying what constituted effective college teaching from the perspective of students, colleagues and even instructors themselves. Evaluation centered on measuring the soundness of teaching, and evaluation was also beginning to be seen as a way of providing information for decision-making about tenure, rank and salary (Cronbach, 1963; Stufflebeam, 1969; and Alkin, 1969).

Cronbach (1963), in a position paper in Teachers College Record, extolled the virtues of potential course improvement through evaluation, which he called performance assessment, but emphasized that evaluative information, which he termed feedback, was also vital for administrative decisions on tenure, promotions and raises. Stufflebeam (1969), reporting on the uses of evaluation information at institutions of higher education in Illinois, found that 78 percent of colleges and universities in the state used evaluation information for administrative decisions. Alkin (1969) emphasized the increased use of evaluation for "decisions concerning rank, pay increases and the granting of tenure" (p. 7).

Two other approaches defining evaluation emerged in the 1970s. One emphasized evaluation as a means of assessing teaching effectiveness (Eble, 1971; Kerlinger, 1971; Bolton, 1973; Kulik & McKeachie, 1975; Kirschling, 1978). The other defined evaluation in terms of description and judgment about good teaching (Hildebrand, et.al., 1971; Feldman, 1976).

In his book, The Recognition and Evaluation of Teaching, Kenneth Eble (1971), aligned himself with the effectiveness assessment approach by defining evaluation of instruction in higher education in terms of "efforts at judging teaching effectiveness"

(p. 11). Likewise, Kerlinger (1971), in his summary of research into the scope of student evaluation called evaluation "assessment of outcomes of instruction" (p. 353). This position was reinforced by Kulik and McKeachie (1975), who defined evaluation as "measuring the quality of teaching performance" (p. 210). Kirschling (1978), in his book, Evaluating Faculty Performance, called evaluation the assessment of faculty performance and vitality.

On the descriptive/judgment side, Hildebrand, et. al. (1971), on the basis of their three-year study of 1,600 students and faculty at the University of California, Davis, which attempted to define and describe effective teaching, called evaluation of teaching "describing attributes associated with effective teaching" (p. 20). Similarly, Feldman (1976), who analyzed 52 studies of student ratings of college instructors, concluded that evaluation was basically descriptions--not explanations--of teaching.

In more recent years, definitions of evaluation have focused on the assessment of merit or worth of instructors' work (Kronk & Shipka, 1980; Millman, 1981; Ory & Braskamp , 1981; Dressel & Marcus, 1982; Goodwin, 1983; Ericksen, 1984; Seldin, 1984; Braskamp, Brandenburg & Ory, 1984; Marsh, 1987). This newer definition appears to be a melding and refinement of the aforementioned assessment of teaching effectiveness and description/judgment approaches to defining evaluation.

Kronk and Shipka (1980) cited the 1980's accountability movement in their handbook of evalution of faculty in higher education as the impetus for increased concern over defining and implementing evaluation. They defined evaluation as "judgments about instructional effectiveness" (p. 7) and spoke about two categories of evaluation, summative and formative. Summative evaluation focuses on outcomes and attempts to define teaching effectiveness with global, judgmental items to be used by institutions to make decisions on tenure, salary and promotions. Formative evaluation focuses on

process and attempts to define teaching in behavioral terms to provide diagnostic feedback to instructors.

Millman (1981), accepting the uses of summative and formative categories in his handbook on teacher evaluation, focused on evaluation as measuring the effects of "worthy educational practice" (p. 229), and Ory and Braskamp (1981), also accepting the summative/formative categorizations, underscored the accountability theme in their research on validity and reliability of student evaluations by defining evaluation as assessing the teaching competence of faculty.

Dressel and Marcus (1982), emphasizing on-going accountability in their book about effective college teaching, called evaluation of teaching finding out "what learners have done, are doing and continue to do" (p. 209). Goodwin's guide, Faculty and Administrator Evaluation (1983) named evaluation the "determination of the significance, worth or value of instruction, through careful appraisal in terms of explicit objectives or standards. . .to perform tasks and fultill job requirements at desired levels of performance" (p. 1).

In the same accountability vein, Braskamp, Brandenburg and Ory's guidebook, Evaluating Teaching Effectiveness (1984), called evaluation defining good teaching "through input (knowledge), process (methodology), and product (learning, attitude change, skill acquisition)" (p. 16). Marsh (1987), stressed accountability in a series of monographs, maintaining that evaluation was assessment of instructional quality based primarily on student appraisals of instruction, but buttressed by ratings of instructional quality as seen by colleagues, reviewed by administrators and as ascertained, also, by instructor self-examination.

What it means, currently, to evaluate instructors in higher education is perhaps best summed up by Maryellen Gleason Weimer in her presidential address to the 1987

convention of the American Association of Higher Education. She said that to be evaluated is, in essence, to have a performance appraisal done--by the students, administration, peers or the instructor himself/herself. She defined performance appraisal as a review of basic competencies--assessing the adequacy of instructors in the way they teach.

So then, expressions such as worth, competencies, instructional quality and performance appraisal appear more frequently in the latest literature on faculty evaluation than in earlier research (from the 1920s through the 1960s), and more recent definitions have begun to delineate formative and summative evaluation, seemingly to justify using evaluation in making administrative decisions, while still claiming that evaluation is done to improve teaching.

Whatever definition is applied to evaluation, the extent to which faculty have been evaluated in higher education is extensive and increasing. Clark (1961), in a national survey of college faculty, found that over 70 percent of his sample of 500 college educators said their teaching was evaluated in some manner, but often done infrequently and haphazardly. An extensive study (Gustad, 1961) of instructor evaluation in 585 United States colleges and universities revealed that some type of formal faculty evaluation, usually short surveys, was being done in over 40 percent of these institutions, primarily by students. Later, Gustad (1967) replicated his earlier study and found that, while formal evaluation of college and university teaching was increasing in frequency (up nine percent from 1961), there appeared to be a decline in the systematic use made of evaluation ratings.

In 1967, The American Council of Education, wishing to confirm the quality of instruction in higher education, commissioned an extensive investigation into the state of the art of faculty appraisal. This study, a survey of over 400 higher education administrators about the presence and frequency of evaluation conducted by Astin and Lee

(1966), concluded that evaluation of teaching was increasing overall. Over half of the respondents reported an on-going evaluation procedure of some type. But the study found that such evaluation was not viewed as meaningful due to inaccurate and unreliable methods.

Between 1966 and 1971, a significant increase in the systematic evaluation of instruction in higher education, especially in the use of student evaluation, was noted by Wilson and Gaff (1975). Centra (1972) reported similar findings of more extensive evaluation in research funded by the Educational Testing Service. He found that out of 60 colleges contacted nationally, only nine had no evaluation procedure in place. Kulik and McKeachie (1975) reviewed 11 studies of evaluation in higher education from 1943-1973 and concluded that evaluation of teaching in higher education had increased markedly.

Seldin (1975) repeated the Gustad and Astin and Lee surveys in 1974 with academic deans in liberal arts colleges to examine changes that might have taken place in the evaluation procedures in the eight-year interim. The deans reported increasing use of evaluation procedures (up 17 percent over the eight years), especially student evaluation. Although in the national Gustad and Astin and Lee studies fewer than 25 percent of the institutions used student ratings, 55 percent of the colleges in Seldin's survey reported their use (Cohen & McKeachie, 1980). Seldin's study also concluded that "the gathering of systematic and structured information for evaluation has dramatically increased," (1975, p. 21) especially the data gathered by evaluation committees.

Increased use of student ratings was also reported in studies that included doctoral level universities (Boyd & Schietinger, 1976), done by the Southern Regional

Education Review Board. The studies found evaluation of instruction at doctoral-granting universities increasing, especially evaluation done by department heads.

Genova, Madoff, Chin and Thomas (1976) reported that the trend was toward more evaluation in higher education, citing statewide evaluation programs in Pennsylvania, and mandatory evaluation of instructors at the University of Massachusetts. Centra's (1977) survey of a national sample of department heads at American universities, exploring how universities evaluated faculty performance (a study sponsored by the Educational Testing Service), called evaluation of teaching competencies commonplace in higher education. Seventy-one percent of the department heads indicated that students and department heads evaluated instructors at least once a year. Centra (1977) found that evaluations by higher level administration, e.g. deans or even presidents, were on the increase, as were colleague and self-evaluations.

Kronk and Shipka (1980), who reexamined the 1977 Educational Testing Service sponsored study of department heads, concluded that over half of the American higher education institutions evaluated instruction in some way; most by doing student evaluations. Peter Seldin (1984) surveyed evaluation policies and practices on campuses of all accredited, four-year, undergraduate, liberal arts colleges (private and public) in America. A questionnaire was mailed to 770 academic deans; 616 responded. He concluded that the collecting of data on evaluation in 1983 was more frequent, structured and systematic than ever before, but especially more frequent, structured and systematic than in a similar study also done by Seldin (1978). And Licata (1986) predicted that as the higher educational community entered the 1990s, greater attention to faculty evaluation could be expected, even to the inclusion of post-tenure evaluations.

Another indicator of the extent of evaluation being done in higher education is the number of research studies completed on the subject. Between 1905 and 1948, Barr

(1948) discovered only 138 studies on faculty evaluation in higher education in his literature review. Between 1968 and 1974, DeWolf (1974) found 220 studies on faculty evaluation. Marsh (1987) revealed that over 1,055 studies on faculty evaluation had been done between 1975 and 1984. Kronk and Shipka (1980) stated that beginning with the accountability movement of the 1970s, and fueled by further concern for accountability in the 1980s, "what we are seeing today is not evaluation for the first time, but more evaluation in higher education than ever before" (p.8).

Needs and Purposes of Evaluation in Higher Education

Historically, evaluation of instruction in higher education has been seen as serving formative purposes, as a vehicle for maintaining and improving the quality of teaching and learning through providing diagnostic feedback on teaching effectiveness and appropriate teaching methodologies (AAUP, 1938; Wilson, 1942; Eckert, 1950; Stuit & Ebel, 1952; Kulik & McKeachie, 1975; Hildebrand & Wilson, 1970; Blank, 1978; Young & Gwalamubisi, 1986; Benton, 1982; Erickson, 1984; Eble, 1971; Millman, 1981; Doyle, 1983; Wilson & Gaff, 1975).

In addition, evaluation of teaching in higher education has also been perceived as needed for summative purposes, which encompass evaluation for general administrative decision making, such as tenure, promotions, retentions and salaries (Young & Gwalamubisi, 1986; Kronk & Shipka, 1980). And a third purpose of evaluation identified in the literature has been for course selection and curriculum development (Eble, 1971; Genova, Madoff, Chin & Thomas, 1976).

Formative Purposes of Evaluation

As far back as 1938, An AAUP biennial survey of education in the United States based on contact with 35 institutions of higher learning, reported that they perceived evaluation to be needed to ensure that instructional quality remained high. In 1942, Logan Wilson, writing in his book The Academic Man, stated that "the most critical problem confronted in the social organization of any university is the proper evaluation of faculty " (p. 112), which Wilson said was needed to help affirm the worth of college classroom instruction. Eckert (1950) and Stuit and Ebel (1952) affirmed this need to assess the quality of performance in the classroom. Kulik and McKeachie (1975), in a study of evaluation procedures at a large mid-Western university, noted that administrators, in particular, viewed evaluation as a vehicle for improving teaching.

Hildebrand and Wilson (1970), in their monograph, Effective University Teaching and Its Evaluation, concluded that instructor evaluation was critical for the improvement of teaching and learning. In a similar vein, Blank (1978) found, after analyzing data from an American Council on Education survey of 7,350 faculty from a cross-section of 301 colleges and universities, that faculty recognized evaluation's potential for instructional improvement. Young & Gwalamubisi's later (1986) survey of 117 full-time faculty and 36 academic administrators in six Eastern Washington community colleges, supported this position. By a large margin (89 percent), faculty deemed evaluation to be needed for the improvement of instruction and for the assessment of instructional quality.

Benton (1982), who analyzed six criterion validity studies of faculty evaluations, concluded that evaluation could contribute to better teaching, especially as

evaluation of teaching helped assess the appropriateness of methodology and approaches, and thus enhanced the quality of teaching. Ericksen (1984), in his book, The Essence of Good Teaching, based on 20 years experience at the Center for Research on Learning and Teaching at the University of Michigan, concluded that evaluation could help professors "decide if they are exercising specific, correct skills for meeting particular problems, and to aid in understanding why some methodologies work" (p. 10).

Eble (1971), in a monograph, The Recognition and Evaluation of Teaching, affirmed the use of evaluation in improving college teaching. He stated: "The desired condition is to use evaluation as we use criticism we respect: to provide both the drive and direction for improving upon what we do " (p. 16).

In his Handbook for Teacher Evaluation, Millman (1981) stressed that evaluation was needed for improvement of instruction, and to encourage instructor creativity and experimentation. He also stressed: "Evaluation can help teachers improve their performance by providing data and suggestions that have implications for what to teach and how, and should be the reason for existence of any faculty evaluation" (p. 13).

The literature in the field suggests that diagnostic feedback about instructors' knowledge and organization of subject matter, skill in instruction, and personal qualities, can motivate faculty members to improve. So concluded Doyle (1983) in his book on evaluation of instruction. He believed that diagnostic feedback would contribute to the personal and professional growth of the professoriate. Sheehan (1975), writing about efforts to improve teaching through a clinic at the University of Massachusetts, concluded that diagnostic feedback, properly supplied, would result in improved teaching, and Eble (1971) argued "evaluation increases the chances that excellence in teaching will be recognized, and faculty should make the most of the feedback evaluation furnishes" (p. 17).

In their book, College Professors and Their Impact on Students, Wilson and Gaff (1975) reported that 73 percent of the 1,069 professors surveyed said that evaluation procedures which provided teachers with feedback about their teaching "held promise for improving instructors and their teaching styles" (p. 15).

Summative Purposes of Evaluation

Another reason for assessing teaching performance is to provide data for decisions on tenure, salary and promotion, termed general administrative decision making (Young & Gwalamubisi (1986). In recent years, this has become a major purpose of teacher evaluation at institutions facing financial retrenchment and demands for accountability, according to Genova, Madoff, Chin and Thomas (1976). Their book, Mutual Benefit Evaluation of Faculty and Administration in Higher Education, which is labeled as a guide for developing faculty evaluation programs for colleges and universities, stressed the heightened interest given to evaluation for administrative decisions, calling this purpose the "principal current preoccupation of faculty evaluation" (p. 8). These writers also stated that evaluation could be used to glean information to help inform external audiences about faculty performance and to conduct research on factors related to faculty performance.

Whitman and Weiss (1982), in a report for the American Association for Higher Education, and Castetter (1981) in his book, The Personnel Function in Educational Administration, further confirmed that interest in faculty evaluation for administrative decision making was high, due to increased demands for accountability.

Pollack (1976), in his study of the purposes of evaluation in technical/community colleges of North Carolina, and Martin, Overhold and

Urbana(1976), in their critique on accountability in American education, also confirmed that accountability has caused faculty evaluation in higher education (particularly student evaluation) to be used for administrative decision making.

Course Selection/Curriculum Development

Another purpose of evaluation is for use in course selection and curriculum development. In terms of course selection, evaluation aims at providing students with a guide for selecting courses and instructors by assessing their relevance and level of difficulty. Evaluation for this purpose is often administered by students and results are typically printed in a handbook distributed to students as a course selection guide. Eble (1971) concluded that these handbook/guides were useful to students, who basically want to find out if a course is worth taking or is to be perferred over another. According to Feldman (1979), in his review of 65 analyses of college student assessments of teachers and courses, such ratings served as gauges of course utility, and as aids in determining course alignments (e.g., under humanities, fine arts, etc.), and whether courses should be dropped or added. Geiss (1984), in his essay on the context of evaluation, recounted the active involvement of campus student government organizations in the collection of faculty ratings for improved course selection.

In their guidebook for evaluating teaching effectiveness, Braskamp, Brandenburg and Ory (1984) noted that evaluation as a student guide for course selection was a bonafide use of evaluation, when evaluation information was complete, and when all courses were included in published lists. Eble's (1971) monograph, Recognition and Evaluation of Teaching ,supported the idea that, what he called "open evaluations" (p. 35) or student evaluations of instruction subsequently published and made available to

students, faculty and administration could well cause some students to select certain courses and reject others. Similarly, Arreola (1987), in an essay on the role of student government in faculty evaluation, maintained that student evaluations, collected and/or published by college or university student government associations, were intended as a guide as to which faculty's courses to take and which to avoid. This, he declared, was the "essence of faculty evaluation" (p. 39), since students could vote with their registration fees in reflecting their collective evaluation of the worth of instruction.

Genova, Madoff, Chin and Thomas (1976) echoed this belief, and stated that evaluative data collected by students could even be used to evaluate curricula, programs, departments and other units.

Evaluation for curriculum development has been seen as more of a benefit to colleagues involved in course and curriculum development, rather than to instructors for their own improvement as instructors (Braskamp, Brandenburg & Ory, 1984), since this purpose of evaluation focuses primarily on course rather than instructor evaluation. Curriculum development has been mandated, in part, by public demand for accountability, as evidenced by outcomes assessment in American colleges and universities (Office of Educational Research, 1986). These institutions are anxious to prove to state governments and other funding and accrediting agencies that they are getting the desired outcomes from students (Boyer, Ewell, Finney & Mingle, 1987). Courses that are seen as means to this end are kept, strengthened, offered more frequently, et cetera, while courses determined by faculty evaluation to be less than effective in achieving desired student outcomes have been dropped or modified (Ory & Parker, 1989).

Methods of Evaluating Teaching

There are generally recognized to be four methods of evaluating instruction in higher education discussed in the literature, each of which corresponds, basically, to the sources of the evaluative information: students, administrators, colleagues (or peers) and the instructors themselves. Each method/source will be discussed individually.

Student Evaluation of Teaching

Student evaluation of teaching in higher education is directed primarily toward improving the quality of teaching, although this type is being used increasingly by college and university administrators to make decisions about tenure, promotions and salaries. Student evaluations of college/university teaching are usually conducted once a year, at the end of a term, to determine the quality of teaching experienced and to determine if the course objectives have been met (Doyle, 1975). These appraisals have taken the form of rating scales, narratives and interviews. Assessment of teaching based on student achievement tests has also been done (Eble, 1971).

According to Doyle (1975), who conducted research into student evaluation forms, the rating scale has been the format most often used, and the questions most often asked are about the instructor's knowledge of subject, enthusiasm for teaching and subject matter, and organization and clarity of materials and presentations. Doyle compared student rating forms from the University of Minnesota, Princeton and Purdue. His findings indicated that, on average, most student evaluation forms were subdivided

into sections which assessed anything from instructor approach to teaching, to grading, to how much was learned.

Student evaluation did not get much attention in the literature before the 1920s, apparently because few colleges and universities were doing student evaluations prior to that decade. In Marsh's monograph (1987) on faculty evaluation in higher education, he mentioned that Harvard, Purdue and the Universities of Texas and Washington were exceptions. According to Marsh, all four universities implemented programs of student evaluation of professors' teaching in the 1920s.

As late as 1951, Mueller conducted a national survey which showed that 37 percent of American colleges and universities had considered, but not adopted, student evaluation, but he predicted that student evaluations of college and university teaching would increase. In the 1960s, however, there was a decline in the use of student ratings of faculty. The American Council on Education, to investigate the prevalence of faculty evaluation, sponsored a survey (Gustad, 1961) which reported that only 24 percent of American colleges and universities indicated that they used formal student ratings. That figure dropped to 12 percent in a study conducted by Gustad five years later (1967). His polling of 584 colleges and universities revealed a decline in the "systematic use of student ratings" (p. 511). Gustad concluded that such a decline was probably due to lack of sound information on the validity of the testing instruments.

A revolutionary change regarding student evaluation of instruction in higher education occurred in the late 1960s. The social consciousness of American students was awakened and students sought to make their voices heard on campus. Student evaluation of teaching was one expression of their desires, according to Kerlinger (1971), who listed student demands to evaluate professors as "one of the most insistent demands" (p. 353). Since the late 1960s, according to Sheehan (1975) in an essay on evaluation and

teaching improvement, "student dissatisfaction with the impersonalized university" (p. 64) caused such demands to escalate.

Student evaluation of instruction has become more common practice since the 1970s, particularly in response to fiscal constraints and calls for teacher accountability (Miller, 1971). Twenty-nine percent of the institutions cited in a national survey conducted by Seldin (1975), reported using student ratings always in evaluating teaching performance. The percentage increased to 54 percent in the 1978 survey conducted by the same researcher (Seldin, 1980).

Student evaluation research, then, is largely a phenomenon of the 1970s and the 1980s, according to Marsh (1987). He conducted an extensive review/analysis of research into student evaluation, and concluded that student evaluation was pervasive, but varied in format from closed question surveys to exit interviews. Marsh predicted continued growth in the use of student evaluations in higher education because of increasing demands for accountability in primary and secondary education reaching upward to higher learning.

Student evaluation of teaching in higher education is the type which has been used and studied most extensively. Costin, Greenough and Menges' (1971) research into empirical studies of student ratings of college teaching, cited the prevalence of student evaluations. The authors reviewed 60 studies of student ratings of college teaching and concluded that student ratings were in widespread use (76 percent of the studies reported use of student evaluation of classroom instruction).

Building on widespread use of student evaluation in higher education, much of the extant literature concerning faculty evaluation has examined the reliability, validity and generalizability of student evaluation. Doyle's definitive text on student evaluation (1975) indicated that student ratings tended not to be as precise as they could be, and

therefore suspect, since he equated reliability with precision. He also questioned student evaluation validity, since most student evaluations were subjective, and generalizability, since student evaluations differed from one institution to the next, or from one department to the next.

Reliability

Research on reliability of student ratings has focused on the extent to which there is agreement among different student ratings of the same course and instructor (interrater agreement) or agreement among different items intended to measure the same instructor trait, i.e., "enthusiasm" (internal consistency). Remmers (1929) asked 409 students to rate 11 Purdue instructors on a 10-item scale. His research looked at the consistency between student ratings of instructors and the grades students received from those instructors. He concluded that there was no noticeable tendency for students to vary ratings of the same instructor based on marks received, nor was there significant variance among the ratings of instructor traits. This suggested that student ratings were reliable, since the student evaluations were not correlated with student grades.

In a synthesis of 31 studies, Feldman (1977), who analyzed college students' views of similar instructor traits, found little consistency between students' concepts of enthusiastic, prepared and knowledgeable instructors, and the same conceptions the instructors held of their own enthusiasm, preparedness and knowledge. This called into question reliability of student evaluations. Butler and Tipton (1976) surveyed approximately 1,800 students taught by 17 different English instructors at Virginia Commonwealth University over a period of three years to check the reliability of student

ratings of those instructors over time. In contrast, their results indicated that those English students could reliably rate the relative strength of instruction during a course, and even three years afterward.

Costin, Greenough and Menges (1971), in a review of empirical studies of student ratings over a 60-year span of time, concluded that student evaluations fell short of a complete assessment of instructor teaching contributions because of inconsistency of rating forms and outcomes over time. In contrast, Centra (1972) of the Educational Testing Service, conducted a study of 343 instructors in five colleges to gauge reliability of student evaluations by comparing student evaluations with instructor self-evaluations. He found that subject area, gender, or years of instructor experience did not affect the reliability of student ratings, and that when student evaluations were good, so were self-evaluations. He therefore concluded, since there was consistency between self and student ratings, student evaluations were basically reliable.

Hildebrand, Wilson and Dienst (1971) surveyed more than 1,600 students at the University of California, Davis over a three-year period about the effectiveness of the teaching the students received. The results of this study suggested that student evaluations were a reliable and effective method of evaluating instruction based on the positive correlation of mean scores in six components of the student evaluation forms and the same six components of colleague evaluations.

Centra's (1973) Princeton study of student and alumni ratings for 23 teachers revealed that there was substantial agreement in the perceptions of current students and alumni (of five years). Centra concluded that student ratings were fairly reliable and did reflect long-term efforts of instruction. However, Rodin and Rodin (1972) conducted a study of 293 students in large undergraduate calculus courses at a major

state university to compare how much students learned to how they scored instructors on student evaluation forms. The study concluded that good teaching was not reliably measured by student evaluation forms, because of low correlations between amount learned and how students evaluated instructors.

Aleamoni and Spenser (1973), in testing a questionnaire (The Illinois Course Evalution Questionnaire or CEQ), collected data from over 100,000 students, 2,000 course sections and 400 different courses at the University of Illinois, and at three other similar institutions to evaluate how reliable student ratings done using those questionnaires might be. The more items there were on the rating forms, the more reliable the evaluations seemed to be (from 25-55 items). When fewer items were used on forms for student evaluation of faculty and courses, the reliability declined markedly. The researchers advised caution in making judgments based on student evaluations.

One aspect of reliability is the freedom from systematic error in the evaluation process. One such error, called "the halo effect," manifesting itself as an overall expression of the rater's judgment, influences ratings of specific traits. Studying the halo effect, Holmes (1971), surveyed seven lecture classes with enrollments of more than 100 each in the College of Arts and Sciences at the University of Texas to test the correlations between anticipated grades and evaluation of the instructor. Contrary to what was expected, students anticipating lower grades did not rate instructors lower than students expecting higher grades. Holmes concluded that it did not appear that expected grades were related to a general halo effect, and he further concluded that student evaluations were therefore fairly reliable.

Treffinger and Feldhusen's (1970) study of 192 Purdue educational psychology students seemed to affirm the reliability of student evaluations. The researchers asked

students to rate general impressions of course and instructor during the first week of class. The same evaluation instrument was used to rate the same instructors during the last week of class. Treffinger and Feldhusen concluded that two distinct processes, expectation and evaluation, had been operating, and that students reliably distinguished what they expected of an instructor from what they actually got.

A volunteer experimental study at a large college of education in the Southeastern United States attempted to examine the effects of early course evaluation feedback to instructor on subsequent end-of-course evaluations by students (n=1,484) and instructors (n=78). The group of instructors that had been given feedback showed some improvement in the end-of-course evaluations. Modest data were revealed to indicate that at least some instructional behavior had been reliably changed, according to both students and instructors.

A study (Butler & Tipton, 1976) of feedback from 1,800 students to 17 instructors at Virginia Commonwealth University was done to determine whether student ratings of instructors, or selected variables, were reliable over time, and whether feedback, in terms of student evalution, affected instructor performance. Each instructor was rated by at least 50 students in at least three classes over the course of two academic years, and 14 of the 17 made positive teaching behavior changes, presumably because of feedback from student evaluations. The results of this study indicated that students can reliably rate the relative strength of instructors, and the ratings were comparable across groups of students.

Overall and Marsh (1980), in a review of 43 studies on the reliability, validity and generalizability of student evaluations, concluded that students agreed in their evaluation of the same course and instructor and, based on follow-up evaluations of graduates one year after graduation, that student evaluations of instructors were

reliable over time. In their own research, Overall & Marsh (1979) investigated long-term stability of students' evaluation of instructional effectiveness by surveying 1,374 undergraduate and graduate business administration majors from 100 different classes at a comprehensive state university. Each student provided course and instructor evaluations at the end of each class and again one year after program completion. Results showed statistically significant correlations between end-of-term and retrospective ratings. Overall and Marsh concluded that student evaluations were indeed reliable over time.

Kulik and McKeachie's (1975) review of 72 studies on the evaluation of teaching in higher education concluded that responses of individual students to commonly used rating forms were both internally consistent and fairly stable over time. The reliability of class ratings, or average instructor ratings, was fairly high when reliabilities were calculated over a wide range of instructors and ratings were made at the end of a course by two groups of students in the same class. Kulik and McKeachie asserted that reliability co-efficients would be lower if the ratings came from different courses, or even sections of the same course taught by the instructor, if the total group rated was restricted to instructors in a single discipline, or if the time when the ratings were made was long after the end of the course.

Kirschling's (1978) sourcebook for evaluating faculty performance concluded that student evaluations were reliable, provided a minimum number of students did the rating. His suggestion was 10, provided that these 10 were representative of the class as a whole; fewer than 10 students could be influenced by student idiosyncrasies. Braskamp, Brandenburg and Ory (1984) reached the same conclusion, but added the caveat that contextual factors such as required/elective status of a course, rank of instructor and class size might confound the effects.

Centra (1980), commenting on six studies of reliability or consistency of student evaluations, concluded that reliability of those ratings was good, provided, again, that at least 10 students in a class rated the instructor, and that the reason for the evaluation was for instructional improvement. He suggested 20 student evaluations per class were needed for decisions on tenure, promotions, etc. Whitman & Weiss (1982) concluded from a review of four studies on reliability of student evaluation that students agree with one another in the ratings they give an instructor (inter-rater reliability) if there are at least 15 students participating in the evaluation; that instructors who receive good evaluations at the middle of a course will likely receive high ratings at the end of the course; and that instructors are likely to receive a high rating when teaching that same course or similar course again. The researchers also concluded that there is low reliability of student evaluations across different courses, i.e., ratings for teaching a large introductory course may be unrelated to ratings for an upper class seminar.

Page (1974), drawing conclusions from eight studies on student evaluation of teaching in higher education, stated that student evaluations had generally acceptable (but far from perfect) reliability, but that their consistency was not the same for all subject areas or colleges. For example, in one study at the University of Rochester, researchers found low reliability co-efficients among humanities students but high reliability co-efficients among natural science students, with no logical explanation for the differences. He also concluded that student evaluations were sensitive to changes in instructor behavior, with students in three studies giving better ratings to instructors at the end of a course because of changes suggested by the students on mid-term evaluations.

Another form of reliability has also been discussed in the literature. Overall and Marsh (1980) compared 1,373 student ratings (commonly called "retrospective

ratings") collected from the same University of Illinois social science students a minimum of one year after graduation to the ratings done at the end of a school term. They found no practical differences in the information provided by either set of evaluations. Their conclusions seemed to validate the consistency of student ratings.

While reliability of student ratings has been explored in the research, there is no clear concensus. While some of the literature supports the reliability of student evaluation, in particular for course improvement, most studies caution the use of student evaluations for personnel decisions on the grounds reliability may not be sufficient for such uses.

Generalizability

Besides reliability, research on student evaluation of instruction has also explored generalizability. In the abstract, generalizability of a sample is the goodness with which it represents its universe. Generalizability demands that student ratings be correlated in different courses taught by the same instructor and even in the same course taught by different instructors.

Attempting to test the generalizability of student ratings, Marsh (1980) arranged student ratings of 1,364 courses from the University of Sydney, Australia, into sets of four, such that each set contained ratings of the same instructor teaching the same course on two occasions, the same instructor teaching two different courses, and the same course taught by a different instructor. He found a high correlation of ratings of the same instructor teaching different courses, and the same instructor teaching the same course on different occasions. Marsh concluded that students' evaluations

primarily reflect the effectiveness of the instructor rather than the influence of the course.

Marsh and Hocevar (1984), again researching classes at the University of Sydney, examined the consistency of the multi-variate structure of student ratings. University instructors who taught the same course at least four times over a four-year period were evaluated by different groups of students in each of four courses (n=314 instructors, 1,254 classes, 31,322 students). The researchers concluded that factor analysis confirmed not only the generalizability of the ratings of the instructors across the four sets of courses, but also the generalizability of multi-variate evaluations.

Webb and Nolan (1955) conducted a study in which 51 instructors in the Naval Air Technical Training School at Jacksonville, Florida, were rated by both students and supervisors on a teaching proficiency scale. In addition, the instructors rated themselves on the same scale. It was found that the student ratings and the instructor self-ratings were highly correlated, but that self-ratings tended to be a little higher. Centra's (1972) study of randomly selected faculty from five colleges also attempted to learn more about the relationship of teacher self-ratings to student ratings. He also concluded that student ratings correlated highly with instructor self-ratings, but again the research suggested that self-ratings tended to be a bit higher than student ratings. The thrust of these studies seems to be that, while there is sometime agreement between student and instructor self-ratings, which suggests generalizability to a degree, the general tendency is for instructors to rate themselves more favorably than students.

Student ratings have also been compared to colleague and administrative ratings. Guthrie (1954) did a study of the evaluation of teaching on the basis of faculty and student judgment at the University of Washington, Seattle, over a 30-year period. He found that student ratings of teaching had a high correlation (.63) to colleague

evaluations. Maslow and Zimmerman (1956) carried out research at Brooklyn College over a three-year period to determine how well colleague ratings of "good" teaching correlated with student ratings of "good" teaching. The researchers concluded that colleague and student ratings of good teaching agreed fairly well (.69).

As for the correlation between student and administrative evaluation of instruction, Webb and Nolan (1955), who studied ratings of 51 instructors at the Florida Naval Air Training Technical School, found that student and administrative evaluations had little correlation, even though evaluation items on rating forms were virtually the same. Marsh, Burgess and Smith (1956) compared student and administrative supervisor ratings in a large multi-section course in aircraft mechanics taught by 121 instructors at Sheppard Air Force Base. Little relationship was found between supervisor estimate of instructor effectiveness and student estimates of instructor effectiveness. It should be noted, however, that most of these studies compared students' in-class observations with peer or administrative inferences about classroom activities based on their experiences with those instructors in other settings.

Student evaluations have also been compared to alumni evaluations. To test the hypothesis that ratings of instructors do not change with the maturity of the rater, Drucker and Remmers (1951) made comparisons of the ratings of 92 Purdue University instructors by 138 alumni (of at least t10 years standing) with ratings by 251 current undergraduates. The researchers concluded that there were substantial positive relationships between the relative average ratings of instructors by students and alumni. Centra's (1973) comparison of student and alumni (of five years standing) ratings of 23 Princeton instructors revealed that student and alumni ratings correlated .75.

Gillmore, Kane and Naccarato (1978), in a sample survey of 42 pairs of classes at the University of Washington, reported the results of two generalizability studies intended to isolate teacher and course effects. Their study sought to compare the effects of teacher versus course evaluations, when the class, not the instructor, was the object of measurement. Their findings indicated students from the samples had difficulty separating course and instructor evaluations, thereby making generalizability over course and teacher quite difficult. They also found that course content did not seem to be a major factor in determination of ratings from a class. The researchers concluded that unless instructors were evaluated on the basis of at least five courses, then reliability and generalizability of ratings were questionable.

Overall and Marsh (1980) examined the contribution of the instructor, the level of the course and the type of course in end-term and retrospective student ratings in a random sample survey (n=220) at the University of Washington. They found that the individual instructor was the most important factor influencing student ratings, and predicted that the same instructor would receive similar ratings in a different course.

Benton (1982) summarized six studies of student evaluation instruments used at the University of Illinois in the psychology department to assess the generalizability of the results. He compared the evaluation instruments to see what common criteria were used in the instruments to assess effective teaching. His findings indicated that criteria on the evaluation forms at the University of Illinois which were designed to measure students' appraisal of what constituted effective teaching were basically generalizable, but he suggested that student evaluation forms, in general, because they vary from department to department and from institution to institution, should not be the sole criterion by which to judge effective teaching. As he stated: "Apparently a great deal

more goes into effective teaching than can be easily evaluated with student evaluation of instruction instruments" (p. 33).

With regard to the generalizability of student, self, colleague, administrative and alumni evaluations, it would seem that, in general, there is little similarity among data from these different sources. What little similarity exists is not strong enough to warrant substituting one evaluation type for another, because of differences in rating forms and perceptual differences among raters as to what is actually being rated.

Validity

Besides generalizability, studies on student evaluation of higher education faculty have focused on validity, the extent to which student ratings measure what they are intended to measure. Hildebrand, Wilson and Dienst (1971), in their three-year study of 1,600 randomly sampled students and faculty at the University of California, Berkeley, sought to describe and define effective teaching while examining the validity of student evaluations. The researchers surveyed students and faculty about what constituted the best and the worst teaching. Their hypothesis was that consistency between faculty and students as to what constituted the best and worst teaching would demonstrate some validity in the student evaluation forms and in what the evaluations measured. Students and faculty responses on good and bad teaching varied so significantly, however, that the hypothesis could not be proved.

Doyle (1975) stated that validity (or meaning) can be both subjective (relying on human judgment and reason) and objective (dependent upon empirical fact and quantitative technique). Crannell (1953), studied over 300 students at Miami University of Ohio . He attempted to identify subjectively valid components or factors

involved in student appraisal of instructors on a 21-item experimental rating scale. He found that three dimensions of student ratings consistently appeared--enthusiasm, knowledge of subject matter and fairness. Bendig (1954) did a similar study of subjective validity on eleven classes at the University of Pittsburg to find whether factor analytic techniques applied to student ratings on the Purdue Rating Scale of Instruction would reveal any consistent patterns of responses. He discovered that two dimensions of instructor rating--"instructional competence" and "instructor empathy" were present consistently.

Coffman (1954) did similar research at Oklahoma A and M University of 19 items on 2,000 subjects. Four significant aspects of teaching that students rated highest in judging effective teaching were extracted. They were: sympathy, organization, punctuality/normalcy and verbal fluency. The general conclusion of this research was that students were competent assessors of teaching. Cosgrove (1959) surveyed 100 student subjects at a major Mid-western university on a 900-item checklist. Students consistently named four factors: knowledge/organization, relations with students, plans/procedures and enthusiasm as subjectively valid appraisals of teaching.

Similarly, Deshpende, Webb and Marks (1970) studied 674 undergraduate students at Georgia Institute of Technology who rated 32 engineering instructors. One hundred forty-seven teacher behavior items were factor analyzed and 14 first-order factors were extracted. The results indicated that student ratings of instructor classroom behavior were subjectively valid and that certain dimensions of these behaviors were systematically related to the student-judged quality of the instructor's teaching efforts, including enthusiasm, adequacy of preparation and classroom management ability.

One of the most comprehensive of these factor analysis reports was that by Isaacson, et. al. (1964). Using an instrument composed largely of questionnaires then in use at Ohio State University, University of Minnesota and the University of Michigan, the researchers surveyed two groups of 1,260 students in introductory psychology at the University of Michigan. The students rated their instructors on a 46-item questionnaire. Six factors which were consistently identified as measures of successful teaching by both groups were: skill, difficulty of material, structure, feedback, group interaction and student-teacher rapport.

The meaning (validity) of evaluation items is further extended by their relationship with some objective measure (criterion validity). For example, ratings of the trait "verbal fluency" could be objectively validated against the number of "uhs" delivered by an instructor and recorded in a given amount of time. This is known as direct criterion validation (Doyle, 1975). Among the very few examples of direct validation of student ratings are two correlations in an extensive study by Elliott (1950). Elliott, at Purdue University, correlated student ratings of 36 chemistry recitation and lab instructors' knowledge of chemistry with the lab instructors' scores on four achievement tests taken prior to the start of a semester and one test taken after completion of the semester. For the lab instructors, the correlation was a positive but insignificant .30; for the recitation instructors, the correlation was a positive, significant .40. The conclusion was that the students in the recitation sections were significantly able to distinguish instructors who knew more about chemistry from those who knew less because of the nature of the class structure (recitation). And, across 18 sections of introductory economics, McKeachie, Linn and Mann (1971) found a significant positive correlation between ratings of instructor effectiveness in changing

student's beliefs and changes in students test scores on a test that involved distinguishing between naive and sophisticated beliefs about the substance of economics.

Also operating, according to Doyle (1975), is indirect criterion validation, which involves ramifications of the evaluation items, usually ramifications having to do with some kind of measurable change in students, i.e. student interest, student learning, etc. Apparent change in student interest as a result of experiences with a particular instructor was the criterion in a McKeachie and Solomon (1958) study of high-enrollment psychology courses over a three-year period. The study, which attempted to validate student ratings of instructors against the percentage of students who elected advanced courses was based on the premise that one criterion of an effective teacher was his/her ability to arouse interest in the subject matter. In two of the five semesters studied, instructor ratings were significantly correlated with the percentage of continuing students.

Other kinds of meaning or validity could be assigned to student evaluations by relating them to measures of student learning. An early study by Root (1931) of student learning and student evaluation combined a student rating scale with critical essays by students and a more detailed questionnaire designed to probe for reasons for any negative ratings. In the same course the next year, the same instructor reportedly tried to change his behavior as suggested by the previous year's evaluation. Ratings for the second year showed substantial improvement over the first year. On the course exam, the mean of the second year's class exceeded that of the first year's by almost points. The study concluded that there was a positive relationship between ratings and learning mesured by a course exam.

In an attempt at a multi-section validity study, Rodin and Rodin (1972) used 293 students in 12 sections of an undergraduate calculus class taught by graduate

teaching assistants. Three measures were obtained for each section: initial ability in calculus, amount learned by students, and student evaluations of instructors. Rodin and Rodin reported a negative correlation between section-average grade and section-average evaluations. The researchers concluded that students were not perfect judges of teaching effectiveness if that effectiveness was measured by how much students learned, and that good teaching may not be validly measured by student evaluations.

Cohen (1981) did a meta-analysis of all known multi-section validity studies. Across 68 mult-section courses, Cohen found that student achievement was consistently positively correlated with student ratings of instructor skill, and structure of class. Correlations were higher when ratings were of full-time instructors, and when achievement tests were scored by an external evaluator. Cohen concluded that these results provided strong support for the validity of student evaluations of teaching effectiveness.

An interesting facet of the research into the validity of student evaluations has been the "educational seduction" research. Hildebrand, Wilson and Dienst's (1971) study at Berkeley, in which they sought to describe/define effective teaching, raised the possibility that instructors' popularity, expressiveness, authoritative style, wit and other forms of showmanship influenced students' ratings, thus calling into question the validity of student evaluations. Furthermore, Naftulin, Ware and Donelly (1973), in research now referred to as the "Dr. Fox Effect," (or educational seduction), surveyed 55 subjects, all of whom were professional educators, at three different conferences, about the effectiveness of a speaker. They named him Dr. Fox, and said he was an authority on mathematics and human behavior. In actuality, Dr. Fox was a professional actor who gave a lecture on relatively inconsequential information, using non sequiturs, nelogisms and double talk. The audience was asked to evaluate the presenter's

effectiveness on a closed-question rating scale. The researchers found that audiences rated Dr. Fox as effective, knowledgeable and reliable. The results suggested that the audiences were "seduced" into rating these presentations favorably. Similar findings were revealed when Abrami, Perry and Leventhal (1982) replicated this experiment on students in college classes. The researchers concluded that students taught by instructors displaying enthusiasm, humor, friendliness, charisma and the like, may rate their instructors more favorably. Validity was said to be questionable, then, according to Abrami, Perry and Leventhal, since students focused more on instructor traits than on course content and what was learned in class.

Studies in the literature have attempted to prove the validity of student evaluations, but whether the validity of student evaluations has been sufficiently proved is open to question (Doyle, 1975). It would appear that a major deficiency in research on validity of student evaluations is the lack of reasonable theory of instruction to guide the choice and assessment of ways students should rate higher education faculty (Doyle, 1975). Another problem in proving student evaluations valid is that there is often little consistency in the wording of evaluation forms, thereby allowing for vast differences in interpretation.

Other Issues

In addition to the supposed reliability, generalizability and validity of student ratings, a central related issue regarding student appraisals is the debate over whether students are able or competent to make judgments about teaching effectiveness. Administrator and peer evaluations involve assessments by individuals with at least

comparable experience and training. Student evaluations do not, and it has been argued that students are not qualified to make such assessments.

Kerlinger (1971) argued that student evaluations were "unsystematic, lacking objectivity and control. Add to this the emotional and attitudinal filters through which they must pass, and student evaluations raise serious questions " (p. 353). Page (1975) raised the fundamental objection that student ratings were not reliable because students were too young, made their ratings carelessly, and were not able to analyze teaching in its basic elements. Supporters of student evaluation like Doyle (1985), however, see student evaluation of instruction as a "sensitive evaluation method" (p. 91), although subject to student emotionalism and immaturity, and also view student evaluation as not an end in itself, but a means for improving teaching.

Research on yet another aspect of student evaluation, evidence supporting faculty improvement as a result of student ratings, is inconclusive. Overall and Marsh (1979), after analyzing data collected from over 300 faculty members at a large university in the Pacific Northwest, discovered that instructors who received feedback from student ratings at mid-term received more favorable ratings at the end of the year. But Miller (1971), who studied 36 teaching assistants in three courses for college freshmen, seemed to reach a different conclusion. After initial ratings were done on the teaching assistants, subsequent student ratings did not usually improve. This seemed to suggest that the teaching quality did not improve as a direct result of student evaluation feedback.

Centra (1972) studied 343 teachers from five different colleges to compare student ratings with faculty self-ratings. His research asked the same closed survey questions in student ratings and self-ratings. Instructors were asked to comment on the usefulness of both evaluation types. Results were in favor of student ratings as a means of providing instructors with diagnostic feedback, and hence, the potential for improved

teaching. But Centra concluded that behavioral change was more likely to result from consideration of both types of evaluation.

Studies by Fitzgerald (1979) and Centra (1972) suggested that student evaluations had both positive and negative effects on teaching, but their value was considered to have limitations. Fitzgerald conducted a comparative study of student and peer evaluations during a three-year period at a large university in the Mid-west. He surveyed a random sample of the teaching staff, asking which type of evaluation had the greater potential to change institutional practice. He concluded that the faculty surveyed were more likely to change teaching practices as a result of peer evaluation, than from student evaluation.

Centra (1972) surveyed a sample of five colleges and universities from a population of 60 which had no formal student evaluation program. The purpose of this study was to investigate the effects of feedback from student ratings on instructional practices. Student ratings at mid-semester were compared to ratings at semester's end. Multi-variate analysis of variance results for the end of semester ratings, however, indicated no significant differences between mid-semester and end-of-semester ratings, regardless of the college, subject area of the course, sex of instructor, or number of years the instructor had taught. But a major hypothesis of this study was that student feedback could effect change in teachers who had rated themselves more favorably than had students. Results of regression analyses indicated that those instructors who had higher opinions of their instructional practices than did their students did were more likely to change teaching practices after feedback. He also concluded that student evalution seemed to have sufficient impact to warrant its continued use.

Marsh (1987) postulated, in his extensive monographs on student evaluation, that social reinforcement of getting favorable ratings from students provided incentives

for instructor improvement of teaching. Marsh also indicated that there were some factors which influenced student ratings and could compromise student ratings. These factors included: student class size and motivation (also supported by Frey's 1978 research); prior subject matter interest and motivation (Kulik & McKeachie, 1975; Frey, 1978); teacher characteristics such as teaching experience, and teaching load (Feldman, 1977); course characteristics, such as if the course was elective or required (Aleamoni, 1981); and general characteristics, such as the anonymity of student raters (Feldman, 1979), stated purpose of ratings (Feldman, 1976), and presence of instructors during ratings (Feldman, 1979).

Kulik & McKeachie (1975), who reviewed 71 studies on the evaluation of teachers in higher education, attempted to compare inferences made about quality of teaching based on observer (student) evaluations with direct measurement of student performance. The results of this study were inconclusive. As Kulik and McKeachie concluded, the most impressive thing about studies relating class achievement to class ratings of instructors was the inconsistency of results. Student ratings, in particular, they concluded, were influenced by factors other than teaching skills, e.g. enthusiasm for subject and availability to students. Kulik and McKeachie also noted that student feedback, because it could be so judgmental, would not hold much meaning for instructors.

Frey (1978), in a two-dimensional analysis of student ratings of instruction, surveyed all undergraduates of Northwestern University in an attempt to compare ratings with class size and grades. His findings questioned Northwestern's instructional ratings because of two factors, identified as instructor skill and instructor rapport, which he said should have been analyzed separately, but were not. He concluded that most similar ratings, because they tend to "lump together" factors that should be

examined separately, would be questionable. Frey's findings also indicated that student motivation for learning was a key factor in how high students rated instructors and courses.

Feldman (1979) reviewed analyses of college students' assessments of 65 courses and their instructors in a study at the State University of New York at Stonybrook. He discovered that ratings were higher for instructors and courses when instructors were present during evaluation. In 1988, Feldman analyzed 41 studies from American colleges and universities in which students and faculty specified instructional characteristics that each group considered particularly important to good teaching and effective instruction. His purpose was to explore the extent to which students and faculty differed in the criteria each used in evaluating teaching. Feldman discovered significant differences between the students' and faculties' views. Faculty placed more importance on teachers being intellectually challenging, motivating students, and setting high standards. Students placed more importance on teachers being interesting, available and helpful. Feldman concluded that faculty reservations about the meaning of student responses regarding teaching effectiveness were well grounded, based on the analysis of the conclusions of the 41 studies under review.

In 1977, Feldman reviewed and analyzed the literature on consistency and variability among college and university students in how they rated their teachers. He included 71 studies which spanned two decades, and concluded that some student evaluations could not be interpreted meaningfully because the influences of student background and experiences, and the actual behavior and attitudes of the teachers made getting and receiving objective feedback on teaching very difficult.

Aleamoni (1981), in a review and analysis of literature on six studies of extraneous variables or conditions that faculty thought could affect student ratings,

concluded that students who were required to take a course tended to rate it lower than students who elected to take it.

In summary, student evaluations of teaching in higher education are concerned with assessing the effectiveness and quality of instructor and instruction for the purposes of instructional improvement, and for administrative decision making (in some institutions). It is the method of evaluation most often employed, and more research about this type of evaluation may be found in the literature than on any other type. Most of the studies of student evaluation extant deal with the validity, reliability and generalizability of student evaluations. While student evaluation of teaching can be used for instructional improvement, and even for administrative decisions, the reliability, validity and generalizability of student evaluations in assessing overall teaching effectiveness is questionable due to attendant methodological inconsistencies and problems.

Administrative Evaluation

Administrative evaluations are done to assess teaching competency and to aid in making decisions on tenure, promotions, salaries, etc. Administrative evaluations are usually done by department heads, deans or the equivalent, about once a year. Sometimes, classroom visits are involved, but some administrators conduct informal polls of students and colleagues or review written student evaluations (Simpson, 1966). Evaluation of instruction by administrators has been second only to student evaluation in frequency of use in higher education (Seldin, 1984).

Student ratings have been compared to administrative ratings in the literature. Guthrie (1954) described a procedure developed at the University of Washington for

evaluating faculty service. Student ratings of instruction were compared to administrative ratings, based on findings of a faculty committee. There was a low correlation between these types of evaluation, owing to the subjective nature of the administrative evaluation, or so concluded the researchers, for it was found that administrative evaluation via faculty committee findings were often gleaned from anecdotal comments made by students to faculty and administration.

Webb and Nolan (1955) reported findings comparing student and supervisor ratings of 55 instructors at the Naval Air Technical Training School in Jacksonville, Florida. The supervisor's ratings were not correlated with the student ratings at all. The absense of or low correlations in student and administrative evaluation of teaching was explained by stating that student evaluations involved in-class observations, while administrative evaluations involved administrative inferences about classroom activities. The findings called into question the generalizability, validity and reliability of those supervisor ratings.

The trend in administrative evaluations has been for administrators to gather data on faculty teaching quality in a systematic, structured, formalized, cognitive way (Young & Gwalamubisi, 1986) for purposes of improving instruction and for making decisions on retentions, promotions, tenure, etc. (Seldin, 1984). There are two views regarding the role of administrators in faculty evaluation for administrative decisions. According to one view, evaluating faculty for promotion and tenure decisions is an appropriate role for administrators. Philippi (1979) in his review of the evaluation practices of higher education institutions in New England, noted that no faculty review process was complete without administrative input, but he acknowledged the difficulty in administrators evaluating faculty. A second view of the role played by administrators in evaluating higher education faculty holds that administration should have less control

over the evaluation process. Ladd and Lipset (1973), in a report on teaching, concluded that, while administrative input may add dimensions to faculty evaluation, faculty rewards in the form of raises, promotions and tenure should be based more on criteria clearly spelled out and on student, colleague and self-evaluations.

A Southern Regional Education Board evaluation project (AAHE, 1982), indicative of the growing interest in administrative input into evaluation in higher education, surveyed 30 institutions, all of which responded to the questions. The project coordinator compiled a list of seven factors contributing to the success of any college-level evaluation program based on the survey results. Active involvement of high-level administration topped the list. This study also concluded that administrative evaluation, to be most effective, should also involve educational consultation (ways to improve, new methods/approaches to try, etc.). Out of the 30 schools participating, approximately half (52 percent) said administrative evaluations were more advantageous in making personnel decisions than in improving teaching.

Seldin (1984) conducted a nationwide, random-sample survey of 770 academic deans at public and private colleges and universities to examine trends and practices in faculty evaluation in higher education. Those deans reported that administrative evaluation was the most heavily relied upon type of evaluation used in making decisions on tenure, promotion, etc. Seldin also asked them what they thought was the best way to assess faculty teaching performance. Almost by acclamation, the deans listed classroom teaching as the most important index of faculty performance.

Genova, Madoff, Chin and Thomas (1976), after surveying all degree-granting institutions in Massachusetts on guidelines for faculty evaluation, concluded that administrative evaluation was "impressionistic judgments based on incomplete data" (p. 31), because these evaluations were rarely based on classroom observations.

The frequency of administrator evaluations has increased in the last 20 years. In a national survey commissioned by the American Council on Education to investigate the dimensions of faculty appraisal (Seldin, 1975), 491 academic deans (or 83.5 percent) reported increased usage of administrative evaluations, including evaluations by department chairs and deans (a 26 percent increase from a similar 1966 study). Whitman and Weiss' research (1982) also confirmed the increased use of administrative evaluation. Their research examined 100 faculty evaluation systems in United States higher education institutions (chosen at random) in the 1970s. They noted that increased use of administrative evaluations more readily served personnel decisions than helped faculty improve teaching. Hughes (1975) analyzed seven studies of college level administrator evaluations to examine the prevalence of that practice. His research indicated that administrative evaluation was done only about half as often as student evaluation, but that administrative officers were beginning to promote administrative evaluation on their campuses. Hughes suggested that administrators openly seek the counsel and advice of faculty members in the development and evaluation of instruments and procedures. The same sentiments were echoed by Millman (1981).

According to Kulik and McKeachie's (1975) national random sample survey of 65 college and university officials which compared the results of student, administrator and peer ratings, the same rating scales and forms were often used for both student and administrator evaluation, forms which rated organization and clarity, knowledge of subject, etcetera. The researchers found that administrator ratings were nearly interchangeable with peer ratings, but neither of them was interchangeable with student ratings because neither of them agreed with student ratings completely. The implication here was that peers, administrators and students agreed with one another concerning

general impressions of faculty teaching ability, but not on specific classroom behavior associated with good teaching. These findings were echoed by Doyle (1975).

Research on administrator evaluations of college/university teaching, particularly regarding generalizability, reliability and validity, has been limited, or so concluded Seldin's (1975) survey of academic deans. His findings suggested that administrator evaluations were based largely on opinion, and the validity of opinion was suspect, since it was seldon based on classroom observation. Seldin discovered that 82 percent of administrator evaluations in his study consisted of examination of student ratings, bolstered by informal student and colleague opinions.

Administrator evaluations, while second in use only to student evaluations, are used more to make administrative decisions than to improve teaching. The reliability, validity and generalizability of administrator evaluations has been questioned in some of the literature because of their highly subjective nature. Also, administrative evaluation tends to rely more on inferences rather than on direct observation.

Peer Evaluation

Behind student and administrative evaluation, in frequency of occurrence, is colleague or peer evaluation. Colleagues evaluate each other's teaching through classroom visits, examination of course material, or a combination of these, primarily for the purpose of instructional improvement, although peer ratings are being used for administrative decision making with increased regularity (Cohen & McKeachie, 1980). Through examination of course syllabi, course objectives, reading lists, examinations and the like, peers provide input on accuracy and currentness of colleagues' knowledge and appropriateness of course content.

Visitation (utilizing a rating scale report form similar to the ones students fill out on instructors or a free-written narrative) is done usually once or twice a term, and reported on to a department head, dean, or president, according to Kulik and McKeachie (1975). Writing in the Review of Research in Education, the authors argued, based on extensive review of research, that ratings by colleagues tended to reflect the teacher's general reputation rather than independent judgments based on close observation. Whitman and Weiss (1982), in a report done for the American Association of Higher Education, in which they studied a national sample of higher education evaluation practices, cited peer review as an unresolved issue, particularly since some faculty considered classroom visits an infringement of rights. Ferren and Geller (1983), in a summary of a report about a newly organized Teaching Effectiveness Committee at the University of Maine at Farmington, cited the need for trust in colleague evaluation, and suggested that colleagues should evaluate only those teaching effectiveness criteria that they were in the best position to judge, such as course design features or intellectual breadth of colleagues. The authors said that the greatest evaluative role peers might serve was to provide professional informaton on teaching effectiveness. Cohen and McKeachie (1980) echoed the same ideas in their review of research into peer evaluation. After their analysis of 16 studies of peer evaluation in higher education, Cohen and McKeachie identified the need for integration of feedback on teaching effectiveness from student, peer and self-evaluations. The authors concluded that the vast majority of the studies they analyzed affirmed that assessing appropriate methodology for teaching in content areas and support of departmental instructional efforts should be evaluated by departmental colleagues.

French-Lazovik (1981) cited both positive and negative aspects of peer evaluation. After a review of the literature on peer evaluation, she proposed that peer

evaluations based on actual classroom visits could be indicators of effective teaching and hence a way for helping validate student evaluations. She also pointed to negative aspects of peer ratings, such as little statistical reliability, validity or generalizability. Reviewing the peer evaluation process in A Handbook of Teacher Evaluation, French-Lazovik also cited the problems of obvious bias and halo effect in peer evaluation, plus the idea that colleagues are in competition with each other. She also mentioned the potentially intrusive nature of classroom visits by colleagues.

Other researchers have maintained that peer evaluations are often indicators of teaching effectiveness (Ferren & Geller, 1983; Centra, 1980; Cohen & Mckeachie, 1980; Aleamoni, 1981), but indicated that the validity of such evaluations is suspect. Ferren and Geller's (1983) summary of efforts of a teaching effectiveness committee at a Maine university concluded that fairly executed standardized colleague evaluations can contribute to better student ratings. Aleamoni (1985), who conducted studies at the University of Illinois while developing and utilizing course evaluation instruments, indicated that peer ratings can be as effective as student ratings in assessing teaching effectiveness.

One study has examined the reliability of peer ratings. Hildebrand, Wilson and Dienst (1971) asked 119 faculty to rate 84 previously identified "good" instructors. Internal consistency correlations on five scales of evaluation items ranged from .65 to .86.

Higher education faculty are usually more receptive to peer evaluation than to student or administrator evaluation, despite the potential for either a halo effect or professorial competition, or so concluded Doyle (1983) in his handbook on faculty evaluation. Centra (1975) obtained student and colleague ratings from 78 randomly selected instructors at Princeton. When he compared the two types of ratings, he

discovered that colleague ratings were more favorable than student ratings, but less reliable than student ratings, because peer ratings were based on only one or two classroom visits.

In studies where peer ratings were not based on classroom visits (Guthrie, 1954; Maslow & Zimmerman, 1956), evaluations by peers have correlated well with student ratings of university instructors, but it is likely that peer ratings could have been based on information from students. Guthrie (1954) studied 30 years worth of evaluation data at the University of Washington based on faculty and student judgment. Ninety-nine of 121 colleague ratings correlated positively (.63) with student ratings. Maslow and Zimmerman (1956), in research at Brooklyn College which spanned three years, concluded from their study of student and faculty agreement on teaching ability, that students and faculty agreed well on who the good teachers were (.69 correlation).

Marsh, Burgess and Smith (1956) correlated student ratings, student achievement, peer ratings and supervisor ratings in a large multi-section course which included 121 instructors. Peer and supervisor ratings, which significantly correlated with each other, were not related to either student ratings or achievement, which they concluded meant that peer ratings did not have much value as an indicator of good teaching.

Cohen and McKeachie (1980), in their analysis of research into peer review, concluded that reliability, validity and generalizability of colleague evaluations were severely lacking and would need to be assured if peer evaluations were to be used for decisions on promotion, tenure and pay increases. They suggested one way to test validity of peer evaluation would be to relate colleague ratings to student achievement.

Conclusions from Centra's (1975) analysis of the beginnings of peer review research in the 1960s and early 1970s indicated that a case could be made for peer

evaluation as a means for improving instruction, but he emphasized that colleague ratings of teaching effectiveness based primarily on classroom visits would not be reliable enough in most instances to use in decisions on tenure and promotions, unless more frequent visits were done and colleagues were trained in what to observe.

Murray (1980), in a review of the literature comparing student ratings and peer ratings, found peer ratings to be "(1) less sensitive, reliable and valid; (2) more threatening and disruptive of faculty morale; and (3) more affected by non-instructional factors such as research productivity" (p. 45) than student ratings.

In summary, peer evaluation is the assessment of colleague teaching effectiveness made through classroom visits, scrutiny of course materials, or both, done primarily for instructional improvement. Peer evaluation suffers from a general lack of research, and from a particular lack of research into the validity, reliability and generalizability of such evaluation.

Self-Evaluation

Self-evaluation refers to instructors' ratings of their effectiveness by means of a scaled form, a narrative assessment, or a descriptive report of performance (Simpson, 1966). Self-evaluations are the least frequently used type of evaluation, and have traditionally been used for purposes of instructional improvement, according to Simpson (1966), who wrote the definitive textbook on self-evaluation. He argued that self-evaluations would be useful for the purposes of aiding professional development, helping instructors develop a professional identity, and even combatting burnout through instrospection which could make teaching more challenging and exhilarating.

According to Skipper (1975), in the Improving College Teaching Yearbook, self-evaluation is based on the assumption that improved teaching, in the final analysis, is the result of a series of individual personal encounters that include self-acceptance and understanding, and self-improvement, with self-evaluation falling in between. The goal of self-evaluation is to encourage instructors to closely examine what they are doing and to assess their strengths and weaknesses (Centra, 1980).

Self-evaluations can take many forms, ranging from written rebuttal-type responses to student or administrator evaluations, to individually constructed narratives. Some self-evaluations utilize the same rating scale format as student or administrator evaluations, according to Young and Gwalamubisi (1986). The two researchers surveyed 117 full-time faculty and 36 academic administrators in six Eastern Washington, rural-oriented community colleges about the prevalence of current evaluation practices, self-evaluations and what the faculty felt should be ideal evaluation practices. They discovered that self-evaluations were the third most often practiced type of evaluation, behind student and peer evaluations. The researchers also found that student, administrator, peer and self-evaluations were being done in 7 percent of the schools responding. Self-evaluations were the second highest ranked type of evaluation under the "ideal" category (behind student evaluations).

Self-appraisal of faculty performance, while gaining some popularity as a useful tool in evaluation of teaching, has been greeted with skepticism (Kulik & McKeachie, 1975; Doyle, 1975; Miller, 1971; Bailey, 1981). Faculty members presumably know most about the amount of effort they have expended and the activities pursued. However, since self-evaluations are made by a single instructor, personal disposition could influence results, according to Genova, Madoff, Chin and Thomas (1976) in their guidebook for higher education evaluation development. Among the benefits of self-

evaluation cited by Genova, et. al., was the reflection stimulated by evaluating one's own performance.

Some of the research into self-evaluation has dealt with exploring the correlation between self-evaluation and student, administrative and colleague evaluation. Webb and Nolan (1955), reporting on the results of supervisor ratings, student ratings and self-ratings by instructors at a naval technical training facility, found good agreement between student ratings and instructor self-ratings, but they doubted that self-ratings could be used administratively because of the high incentives for distortion. Sorey (1968) used 50 college teachers rated by students to study the correlation between student evaluations and instructor self-evaluations at Purdue. He found that superior instructors (as rated by students) showed more accuracy in their self-ratings than instructors rated as inferior by students. But Centra (1972), in his comparative study of five colleges about the effects of feedback from student ratings on changing instructional practices at the college level, found only a modest (.21) relationship between student ratings of instruction and instructor self-ratings.

Blackburn and Clark (1975), studying the correlation among responses from administrators, students peers and instructors themselves on evaluation of instruction at the University of Michigan over a four-year period, found little support for the usefulness of self-evaluation. They discovered little correlation between self-evaluations and administrative and student evaluations. The correlation of self-evaluations with ratings by colleagues was also low.

The study of 343 teaching faculty from five colleges conducted by Centra (1973) compared self-ratings with student ratings of college teachers on 21 items dealing with instructional practices and effectiveness. Teacher ratings had only a modest relationship with student ratings (a correlation of .21). Two additional studies in higher academic

settings revealed correlations between self-ratings and student ratings as .47 for 14 teaching assistants (Doyle & Crichton, 1978), and .49 for 51 social science faculty members (Marsh, Overall & Kesler, 1982).

Research critical of self-evaluation suggests that the phrase itself is almost a contradiction in terms, as Millman espoused in his Handbook of Faculty Evaluation, since evaluations are typically conceived as external measures of competence as judged by administrators, peers or students. Meyer (1980) interviewed 92 subordinates at the University of Florida before they had appraisal sessions with supervisors, asking the subordinates for a confidential self-appraisal of their own performance versus performances of others in similar positions. The 92 rated themselves in the 78th percentile, on average (with 100th percentile being the highest possible rating), while their superiors rated the subordinates in the 64th percentile on average. This led Meyer to conclude that self-evaluations were often inflated.

In addition, research suggests that self-ratings tend to be more lenient than peer or student ratings. Holzbach (1978), in an analysis of supervisor, self and peer performance ratings of 107 managerial and 76 professional employees in a medium-sized manufacturing location, confirmed the greater leniency in self-ratings, and his research supported the presence of a halo effect in self-ratings.

The efficacy of self-evaluation for making decisions on salary, tenure and promotions was questioned by Braskamp, Brandenburg and Ory (1984), who argued that self-evaluation of teaching provided only contextual information for assessing teaching effectiveness. The authors based this conclusion on over a decade of research and experience in the Measurement and Research Division of the Office of Instructional Resources at the University of Illinois, Urbana-Champaign, and they questioned the

credibility of self-judgments of teaching for anything other than improvement of teaching.

Research on the reliability, validity and generalizability of self-ratings of instruction is quite limited (Simpson, 1966). No retest reliability study of self-ratings of college instruction seems yet to have been published (Bailey, 1981), but a small number of educators have issued position papers on self-evaluation. Bayley (1967) expressed little faith in the reliability of self-evaluation, arguing that the word of each instructor can hardly be accepted at face value. A similar view was expressed by Ozman (1967). Eble (1971) concurred, and concluded that administrators, in particular, questioned the reliability of self-evaluation.

Research on the validity of self-ratings of instruction appears to be very limited, especially in comparison to the validation research of student ratings of instruction (Millman, 1981). A few empirical studies have demonstrated that self-ratings show little agreement with student, colleague or administrative ratings (Centra, 1973). Centra (1973) and Blackburn and Clark (1975) both reported correlations of approximately .20 between self-ratings and student ratings. On the other hand, correlations of approximately .70 have been reported between student ratings and colleague ratings (Maslow & Zimmerman, 1956; Blackburn & Clark, 1975).

Self-evaluation validity suffers from what Doyle (1975) called "subjective validation procedures" (p. 67), meaning they must rely on human judgment or reason--especially suspect, since the human judgment and reason involved in self-evaluation is one's own.

As for generalizability of self-evaluations, Doyle's (1975) definition of generalizability--that "the generalizability of a sample is the goodness with which it

represents the universe" (p. 71)--seems to suggest that to make inferences about self-evaluations from each instructor's evaluation of himself or herself is highly suspect.

Kulik and McKeachie (1975) concluded that the problems with self-evaluation were partially a matter of numbers. Administrative, student and peer evaluations were composites, whereas self-evaluations were not composites, but ratings of a single individual, and that personal dispostions of evaluations, while potential stumbling blocks for student, administrative and peer evaluations, were insurmountable obstacles for self-evaluations.

Self-evaluation is basically instructor self-appraisal of teaching effectiveness, executed usually for the purpose of instructional improvement, and done either by rating scales or written narrative. It is the least practiced type of faculty evaluation, partly, perhaps, because of the skepticism surrounding the reliability, validity and virtual non-generalizability of such subjective information. Research is sparse on self-evaluation.

Faculty Attitudes and Perceptions Toward Evaluation

The literature and research contain few systematic studies of faculty attitudes toward evaluation of instruction (Bailey, 1981; Miller, 1971). Braskamp, Brandenburg and Ory (1984) argued that little was known about the credibility, trustworthiness or usefulness of differing types of evaluative information from the perspective of the primary users, the faculty, and stressed the need for more research in this venue. Most of the literature on faculty perceptions of evaluation involves anecdotal information and/or concentrates primarily on student evaluations. Aleamoni (1987) examined such anecdotal evidence for over six years and concluded that there

were several typical concerns higher education faculty held about evaluation (most of which concern student evaluations): (1) lack of maturity and experience of students as evaluators, (2) only peers with excellent publication, research and teaching experience were qualified evaluators, (3) student evaluations were basically popularity contests, (4) students can make critical judgments about instructors' teaching only after they have concluded a course and/or have been away from the institution a number of years, (5) rating forms used by students to evaluate instructors can be unreliable and invalid, (6) some extraneous variables (i.e., gender of instructor, if course is elective or required, the level of the course, etc.) can affect student ratings, and (7) grades students make have been highly related to evaluations the students have given instructors.

Student Evaluation

Ory and Braskamp (1981) studied the reactions of three groups of 50 ramdonly selected faculty members at the University of Illinois to determine the quality and usefulness of evaluative information. The purpose of this study was to investigate faculty perceptions of student evaluation information collected by three methods: objective questionnaire items, open-ended questions and group interviews. Faculty rated three simulated evaluation reports on their potential for accuracy, trustworthiness, usefulness and value as information for self-improvement and promotion purposes. The majority of the faculty (80 percent) regarded the evaluation information to be more credible and useful for their own self-improvement than for promotion purposes. The majority of the faculty (68.2 percent) also desired more than one type of evaluative feedback, regardless of the purpose of evaluation.

Wheeler (1972) studied faculty attitudes toward student evaluation at four liberal arts colleges. He researched receptivity to student evaluations and how administrators used student evaluation information. He surveyed the entire full-time teaching populations (n=420), ostensibly looking for differences in attitudes between tenured and non-tenured faculty. He concluded that most tenured faculty accepted mandated student ratings with reservations, and that most young, non-tenured faculty viewed student evaluation of teaching unfavorably. Wheeler's findings also indicated the presence of faculty uncertainty regarding the interpretations and weighting of data by the administration. Yet the faculties believed, in the main, that student evaluations could improve instructional quality.

Blank (1978) conducted a nation-wide American council on Education sponsored study of faculty attitudes toward the evaluation of teaching performance in higher education. Specifically, he was wanting to prove faculty support for evaluation programs, and the relationship between that support and differences in faculty characteristics, i.e. teaching versus publishing orientation, faculty status and faculty security. Responses were obtained from over 7,000 randomly selected faculty at 301 teaching institutions. Blank found that faculty members with a dominant teaching role were more likely to support the use of teaching performance as a means of evaluating faculty than research-oriented faculty members. Also, faculty members with high status and security in the institutions were more likely to support students evaluation than low-status faculty members. Blank found that 8 percent of the faculty responding favored some sort of evaluation as proof of effective teaching, and 69 percent supported student evaluation.

Attitudes of all 328 faculty members, both full and part-time, at George Mason University were surveyed about the evaluation of faculty and about publication of

individual faculty ratings by students (Gross & Small, 1979). In a survey which included open and closed questions, the authors asked how faculty reacted to being evaluated, if evaluation had any impact on teaching, and if students were capable of adequate assessment of teaching. A majority of the faculty members (76.5 percent) supported the idea of having student evaluation. Tenured faculty members were more favorable to student evaluation than were non-tenured faculty. None of the respondents, however, indicated that evaluation information led to improved teaching. Respondents were evenly split on the students' ability to evaluate instructors, but most seemed to agree that student evaluations should be used in some way to evaluate teaching effectiveness. Over 60 percent of the sample felt that publication of evaluation results did not violate their civil rights.

All full-time, tenure-track faculty members at a predominantly teaching-oriented state university were surveyed about their opinions on the validity of student evaluation. The purpose of this study was to ascertain the uses faculty made of student evaluation information (McMartin & Rich, 1979). The study concluded that faculty opinions regarding various uses of student evaluations depended upon the respondent's frame of reference regarding the valildity of the evaluation instruments. Specifically, those who believed in the validity of student evaluations had more favorable attitudes toward varying uses of student evaluation than the uncommitted group (those who did not have strong feelings regarding the validity of student evaluation).

Mahfous (1979) distributed questionnaires to 1,190 randomly selected faculty members from each category of higher education institution (including research institutions, comprehensive universities, liberal arts colleges, community colleges and specialized institutions) in Maryland. His study sought to analyze faculty acceptance of student evaluations for instructional improvement versus administrative decisions. He

also sought to compare attitudes held by faculty and by administrators toward student evaluation. He found significant differences in those attitudes. Seventy-one percent of respondents across all categories said they were in favor of having students evaluate their teaching, although anecdotal evidence from the research revealed that faculty questioned the scope of student evaluation and its use in administrative decisions. Faculty who worked at research institutions favored the use of student evaluations in personnel decisions more than did those who worked at liberal arts colleges, community colleges and specialized institutions.

At the University of Wisonsin, Lacrosse, all 300 faculty members were surveyed after a student evaluation of instruction (SEI) policy had been in effect four years (Ory & Braskamp, 1981). The policy required that SEI ratings be collected from all classes of each faculty member for one semester each year. A single overall instructor evaluation item was averaged across all students in the classes of each instructor and was used as one of the indices in personnel decisions: retention, promotion, tenure and merit ratings. The study was conducted to ascertain faculty reaction to being evaluated and to the student evaluative data being used in personnel decisions. Over 90 percent of the faculty believed the policy had reduced faculty morale and their confidence in university administration.

Faculty in the study reported changes in instructional activities as a result of the SEI process. The most frequently reported changes were reduced coursework demands on students, explicit specifications of course objectives, provision of handouts and attention to the organization of course content. All of these changes were listed as desirable by the majority (61.2 percent) of the subjects. Furthermore, a majority of the subjects (79.1 percent) believed that one or more faculty acquaintances had attempted to improve the SEI ratings by engaging in academically irrelevant or inappropriate

activities. Nearly 80 percent of the respondents approved of the use of SEI ratings for instructional improvement, but not for personnel decisions.

Some aspects of faculty attitudes toward evaluation seem consistent, based on what is extant in the literature, including the fact that faculty are guarded about favoring student evaluations (Millman, 1981). Also, faculty at research institutions seem to have more favorable attitudes toward evaluation of teaching by students than faculty members at teaching-oriented institutions (Blank, 1978). Faculty members seem to support the use of student evaluations for teaching improvement (Mckeachie, Lin & Mann, 1971), although some faculty report using such feedback, while others virtually ignore it (Payne & Hobbs, 1979). Payne and Hobbs (1979) did an experimental study of 78 faculty at the University of Georgia to determine how feedback from students affected instructional effectiveness. The researchers found "modest data to indicate that some selected behaviors had been reliably changed" (. 533). The only instructor behaviors showing any change were preparedness and encouragement to students, each of which posted a 7 percent increase in occurrence after feedback was given.

Faculty, in general, do not seem to favor the use of student evaluations for administrative decisions. Boyd and Schietinger (1976) surveyed 453 graduate deans about how they assessed total faculty performance for tenure, promotions and raises. The majority of the deans (68 percent) said they leaned toward student evaluation, but admitted that there was a lack of faculty support for student evaluation being used in administrative decisions.

Administrative Evaluation

There is little research on faculty attitudes toward administrative evaluation. Hughes (1975), in reviewing and summarizing seven studies of administrator evaluation of faculty in higher education noted that "the literature carries little in the way of specific cases of research about faculty attitudes toward administrative evaluation" (p.47). Since all seven studies alluded to problems with administrative evaluations, he conlcuded from this review that faculty perceive administrator evaluation of their teaching as being too subjective to be of much benefit.

Seldin's (1975) study of the state of the art of faculty evaluation in higher education involved polling 491 academic deans. He asked administrators what methods they most relied on for evaluating faculty, and how, generally, faculty felt about those types or methods. A majority (68.6 percent) of the deans reported using administrator evaluation to assess faculty performance, while 83.5 percent said they used student evaluations for that purpose. Findings from anecdotal responses about faculty reaction to evaluation, based on the perceptions of these deans were summarized, and conclusions were that evaluation from chairmen and deans were viewed as "less reliable" (p. 5) by faculty and of "dubious validity" (p. 5) because they were based on indirect information such as numbers of complaints about teaching performance and the relationship of the faculty to the administrators who served as appraisers.

Whitman and Weiss (1982) examined over 100 studies of faculty evaluation systems at American higher education institutions in the 1970s. They summarized their findings about faculty attitudes toward administrative evaluations this way: "For many

faculty, evaluation of their performance by administrators is threatening and the ends of evaluation are seen as punitive" (p. 32).

A study of the higher education institutions in the University of Oregon system (Thorne, Scott & Beaird, 1976), one of the few studies that sampled faculty perceptions of evaluation procedures, polled a random sample of over 1,300 faculty and administrators in colleges, universities and community colleges in Oregon about what types of evaluation were done, how evaluation was used (whether for instructional improvement or administrative decisions), and how faculty reacted to evaluation. They found that department head evaluations were one of the most influential factors in promotion and tenure decisions, but that the majority of faculty (68.7 percent) were not comfortable with administrative evaluation procedures in particular.

Goodwin's (1983) manual for constructing instruments for administrative evaluation of faculty cited negative faculty reaction to administrative evaluation based on anecdotal evidence collected, because of the "great difficulty administrative evaluators have in making objective assessments of the performance of those they are evaluating. They tend to inflate the evaluation in general" (p. 3). Similarly, Millman's (1981) handbook on teacher evaluation cautioned that instructor attitudes toward administrative evaluations were not usually positive because instructors perceived that "evaluation emerges from political rather than professional forces" (p. 297).

Centra's (1977) study surveyed 453 department heads of American colleges and universities, asking them to rate the current use and importance of 15 sources of data in the evaluation of teaching and for personnel decisions. Chairman evaluations were first; dean evaluations were sixth. But the conclusions drawn by Centra suggested that there were problems with the faculty attitudes toward the chairman and dean evaluations. He stated that faculty believed that these evaluations were not sources of data, as was the

case in most student evaluations. Rather, faculty said the chairman and dean evaluations were opinions of the evaluators, and therefore subject to negative perceptions by those being evaluated--namely the faculty.

Blank's (1978) survey of a random sample of 7,350 higher education teaching faculty was an attempt to analyze faculty members' responses to the importance of teaching performance as a means of faculty evaluation. His findings supported the idea that the majority of faculty (73 percent) supported the use of teaching performance as a criterion for promotion, but anecdotal findings indicated an overall distrust of administrative evaluations as accurate methods of evaluating teaching performance.

Sheehan's (1975) study of faculty attitudes toward evaluation conducted by the Clinic to Improve University Teaching at the University of Massachusetts, revealed that of the 339 respondents, only 8 percent felt that administrators were qualified to evaluate their teaching. Reasons given for such attitudes ranged from a fundamental distrust of administrative motives to lack of objectivity.

In addition to exploring how widely practiced and how widely accepted the four types of evaluation were, Young and Gwalamubisi's (1986) study of a random sample of 117 full-time faculty in six Eastern Washington rural-oriented community colleges found that administrative judgment was perceived to be the least desirable method because of bias.

In summary, the literature directly related to faculty attitudes and perceptions toward administrative evaluation of teaching is sparse. What research does exist reports attitudes and perceptions secondarily, as anecdotal findings, and those findings usually reveal ambivalent feelings toward administrative evaluations. While faculty recognize that administrators have the experience and maturity lacking in students, the

faculty also distrust administrator evaluations because of their subjectivity and potentially punitive use.

Peer Evaluation

Colleague, or peer evaluation, like administrative evaluation, has received little attention in terms of studies and research. Doyle's (1983) handbook on faculty evaluation stated that peer evaluation "has been less thoroughly studied, even though peer evaluation has been used in the evaluation of instruction for promotion and salary decisions" (p. 18).

Citing virtually no research on how peers react to being evaluated, DiNisi, Blencoe and Randolph (1983) conducted an experimental study of reactions to peer evaluation in undergraduate sociology classes at the University of South Carolina. They found that negative ratings produced lower performance levels, cohesiveness and job satisfaction. Anedcotal findings indicated that peers felt less threatened by peer evaluations than by superiors doing the evaluating.

Eble's (1971) account of classroom visitation at Carnegie-Mellon University, in which a random sample of the entire faculty was asked to assess peer teaching quality, discovered that the majority of both evaluators (64 percent) and evaluatees (76 percent) were uncomfortable with this type of evaluation. Evaluees, commenting in open-ended responses, likened peer evaluation to a pervasive threat. Batista's (1976) literature review of 33 studies of colleague evaluation concluded that peer evaluation can be used with considerable advantage to increase the fairness of administrative decisions on tenure, promotions, retention and merit pay, because peers are better judges of such

things as professional behavior, knowledge of up-do-date subject matter, and attitudes and commitment to peers, students and institutions.

Cohen and McKeachie's later review of literature on colleague evaluation (1980) indicated that the role of colleagues in evaluation had yet to be adequately defined, but their conclusions indicated that faculty seemed to perceive peer review as potentially useful for formative evaluation designed to improve teaching. They also concluded that colleagues were not the best judges of teaching effectiveness.

Gage (1961) sounded a negative note on peer evaluation in a review of research on appraisal of college teaching. He stated that when instructors know they are being evaluated by peers, and that the opinions of those peers "will determine promotion or salary, classroom performance may depend more on nerve than teaching skill" (p. 19).

Seldin (1990) listed some problems faculty may associate with colleague evaluations for administrative decision making, including bias in what constitutes good teaching, discomfort of colleagues in an evaluative role and excessive emphasis on methodology. It is interesting that Seldin reported that, when peer evaluation is used for improving teaching instead of for administrative decisions, faculty perceptions are likely to be much more favorable.

Ferren and Geller's (1983) summary of a report on the inpact of a newly organized Teaching Effectiveness Committee at the University of Maine, Farmington, reported positive reactions to peer evalution, but the authors' conclusions that peers were positive and enthusiastic about being evaluated by colleagues were qualified by the idea that peer evaluators were introduced as "consultants" who were to "interpret" what took place in the classroom.

An increasing amount of research has been conducted on peer evaluation, but the amount is not substantial, particularly with regard to faculty perceptions about peer

evaluations (DiNisi, Blencoe & Randolph, 1983). In their report of a laboratory experiment concerning reactions of peers to both positive and negative colleague ratings, DiNisi, Blencoe and Randolph (1983) conducted research with psychology students at the University of South Carolina. The researchers stated "virtually no research has been reported on how people react upon learning that they have been either poorly or favorably evaluated by their peers" (p. 457).

So, then, what is known from the literature concerning faculty perceptions of peer evaluation is limited, and tends to be either quite positive or quite suspect. The chief factor determining whether faculty react positively or negatively to peer evaluation seems to be how that type of evaluation is used. If faculty believe evaluation is used for summative (administrative decision making) reasons, they seem to perceive it negatively, even threateningly. On the other hand, if faculty believe evaluation is for instructional improvement (formative), they perceive it in a more positive light.

Self Evaluation

Finally, the literature revealed little about faculty perceptions of self-evaluation. Meyer (1980), who studied the psychological effects of self-evaluation on the University of South Florida faculty, learned from open-ended comments that faculty felt positive toward this form of evaluation, because it gave them a chance to really explore what they were doing.

Miller's (1971) study of instructors' attitudes toward and their use of evaluation at the University of Illinois, concluded with the recommendation that self-evaluation be implemented because of the faculty's sometimes negative feelings toward student evaluation. He believed that self-evaluation could be a worthy teaching

evaluation method, because as teachers developed greater awareness, they would be able to respond more effectively to interests of students.

Ormrod (1986), who surveyed over 100 faculty at the University of Northern Colorado about their satisfaction with a new evaluation system, found that the majority of respondents believed that self-evaluation was a better predictor of teaching effectiveness than student, peer or administrative evaluation.

Neely (1973) studied the attitudes of a random sample (N=484) of educators in Ohio toward self-evaluation. He found that their attitudes ranged from neutral to only slightly favorable. Bailey (1981), in his text on instructor self-assessment based on a five-year study of 200 instructors, revealed some positive outcomes from self-evaluation. The instructors revealed positive attitudes toward both their proficiency in self-assessment and in teaching, but they were concerned that they needed more training in self-evaluation.

Perceptions of self-evaluation in higher education have received little attention in the literature. Extant perceptions vary from the very positive to the very negative. Positive reactions center on self-evaluation's potential for intensive self-examination, while negative opinions toward self-evaluation are grounded in the belief that it is so subjective as to be rendered virtually impotent in its ability to assess teaching effectiveness.

Summary

The review of literature and research has attempted to examine evaluation in American higher education. The chapter was organized to include definitions and purposes of evaluation, research and literature relative to the four major types/forms

of evaluation currently practiced, and research and literature regarding faculty perceptions about evaluation of teaching in higher education.

Evaluation of instruction in higher education has been practiced in various forms for almost 70 years, and that practice show no signs of diminishing. As the literature has shown, evaluation was begun as a way of determining what has gone on in college and university classrooms. Specifically, evaluation was originally aimed at improving instruction and learning in the higher education classroom. This evaluation was undertaken by administrative committees earlier in the century, but those committees gave way to student evaluation of teaching.

As student populations grew, so did the demand for evaluation, and since students were and are consumers of classroom instruction, student evaluations have become the most widely used form of evaluation in higher education. As the frequency of evaluation has grown, so have the reasons why evaluation has been practiced. Originally, as the literature research indicated, evaluation was for the purpose of instructional improvement. But administrators began to see the viability of using student evaluation in making decisions on tenure, promotion, dismissals and salaries of higher education faculty. The literature indicated that this added purpose (making administrative decisions) is not always clearly indicated, however.

In addition to student evaluation, other types or forms of evaluation have arisen. Evaluation by administrators via visitation, examination of student evaluations or examination of classroom materials, for both teaching improvement and for decisions on tenure, salaries and the like has increased with some frequency. Peer or colleague evaluations, ostensibly through classroom observation and examination of course materials (tests, syllabi, reading lists, etc.) was begun in the early 1960s and reports from literature and research indicate that peer evaluations are steadily increasing. The

purpose for peer review has traditionally been for instructional improvement, and this continues to be the most prevalent purpose for executing this form of evaluation.

Self-evaluation by instructors, done systematically through specific written questions or by personal narrative assessment, is the least practiced form of evaluation in higher education, according to the research literature. This type of evaluation is gaining in frequency of occurrence, however. Although some administrators admit to using results of self-evaluations in the administrative decision making process, the most often proclaimed purpose for self-evaluation is improvement of teaching.

As for the reliability, validity and generalizability of student, administrative, peer and self-evaluations, the literature indicates that more research has focused on student evaluations than on the other three types combined. While some studies conclude that student evaluation of instruction in higher education is reliable, valid and generalizable, almost an equal percentage of studies have called into question the reliability, validity and generalizability of student evaluations. Far less research has examined the validity, reliability and generalizability of administrative, peer and self-evaluations, but most extant studies question their reliability, validity and generalizability on the grounds that the subjective nature of these types makes reliability, validity and generalizability difficult to attain.

Perceptions of higher education faculty concerning the four types of evaluation have been gleaned neither consistently nor systematically. Most studies on faculty perceptions of student, administrative, peer and self-evaluations have been findings which are primarily anecdotal, not global and direct.

Few, if any, studies have examined globally, specifically and systematically, faculty perceptions about the practice of student, administrator, peer and self-evaluations in higher education. This present study is designed to do just that. The

present study of perceptions of higher education faculty about student, administrative, peer and self-evaluation was intended to explore how faculty feel toward evaluation, its purposes, types, results and who is served by evaluation.

As the literature has shown, there is a dearth of research on faculty perceptions of the four types of faculty evaluation. And there are virtually no studies of faculty perceptions about how and what evaluation should be as compared to what is, and this present study attempts to help fill that void.

CHAPTER III

METHODS AND PROCEDURES

Introduction

The purpose of the study was to examine faculty attitudes toward student, administrative, peer and self-evaluation of their instruction by examining the attitudes of higher education faculty at one institution in which the four types of instructional evaluation were in use. The study sought to reveal their perceptions of each type and to compare their perceptions of how each type of evaluation was used versus how each evaluation type should be used in the ideal.

Since there is litle systematic research of faculty perceptions of evaluation, and no extant research comparing perceptions of evaluation in use with perceptions about how those evaluations should be used, the study was exploratory in nature. It sought to identify variables which may influence the process rather than imposing variables a priori.

The study was descriptive in that the subjects were asked to share their attitudes and perceptions, and their responses were used to describe the phenomena under study, rather than depend upon anecdotal and serendipitous information to ascertain those attitudes and perceptions.

An examination of faculty perceptions was seen as critical to learning more about how higher education faculty feel about the evaluation of their teaching--purposes of and

uses of the four types of evaluation most widely used in higher education, something we currently know little about. Additionally, the research was seen as a way to explore who is ultimately served by evaluation, and to gain insights into faculty understanding of evaluation, both of which are vital to accomplishing the goal of instructional improvement in higher education.

This chapter details the methods and procedures used in the study, including selection of the study population, development of the instrumentation, and procedures used to collect and analyze the data from which findings were made and conclusions drawn.

Selection of the Study Population

This study was designed to gather and analyze the perceptions of a total faculty of one higher education institution. Belmont University, Nashville, Tennessee, was selected as the institution because the four types of faculty evaluation under study were currently being used, thus perceptions of current practice and ideal practice could be compared. A religiously affiliated, four-year, private liberal arts university, Belmont has emphasized and rewarded teaching excellence and service above research, and its faculty have traditionally been so devoted to the cause of private, religious higher education that many have begun and ended their careers at the university as a type of testimony to such devotion. The student body is comprised largely of Southern Baptists from Middle Tennessee who traditionally have sought a small, quality denominational school with a sense of family or community, where individual attention is given to the development not only of the mind, but of the body and the spirit as well.

Also, Belmont was involved in a study of its evaluation system and sought the feedback such a study could provide. The academic affairs office at Belmont was contacted, and the academic dean and members of the ad hoc committee on faculty evaluation reviewed the research proposal and a draft of the survey. All parties agreed that the study was worthwhile and could benefit the university.

Undergraduate instructors who taught full-time and those who taught and had administrative duties that did not exceed one-third of their work load constituted the study population. Since the study examined the perceptions of virtually an entire population, no sampling procedures were required. Preliminary investigation revealed that there were 131 full-time undergraduate instructors. A complete list of names and addresses of these instructors was provided by the Academic Affairs Office at Belmont.

Instrumentation

In an attempt to determine perceptions of current and ideal practice of evaluation at Belmont, a structured survey questionnaire with both open and closed questions was developed (Appendix B). The form included three sections. Section I included eight questions designed to gather demographic information about the respondents including, highest degree earned, years of teaching experience, gender, age, academic division, academic rank and tenure status. Section II consisted of four questions designed to determine the subjects' perceptions of each of the types of evaluation currently in use at Belmont, including how satisfied they were with the procedures, how fair, objective and reliable they perceived each type to be, and whether they thought each type of evaluation should be done.

Section III explored the respondents' perceptions of how faculty evaluation is conducted and used at Belmont versus how respondents believe it should be conducted or used in the ideal. Resondents were asked to react to statements about whether each of the four types of evaluation should be used to improve teaching, improve the major or program, and/or make administrative decisions. Additionally, Section III asked respondents to indicate what was done and what should be done with evaluation information, from returning evaluation information to instructors to doing nothing with the evaluation information. Included in Section III was a question (#15) about how effective respondents perceived the four types of evaluation to be in improving teaching, improving program/major, and in making administrative decisions, which related to faculty perception of the four types of evaluation in use at Belmont.

Open-ended questions (seven in all) were placed at the end of each sub-section of related questions, and at the end of the major section divisions. The purpose of the open-ended questions was to allow respondents the freedom and flexibility to elaborate on and possibly clarify the responses made to the closed questions.

Data Collection and Analysis

A version of the questionnaire was field tested by a randomly selected sample of faculty at Carson-Newman College, a similar private, religious liberal arts institution, comparable in size, programs and in types of evaluation procedures used to the institution targeted in the study. The field testing resulted in clarification and combination of some questions, and even in the elimination of a few questions.

The refined survey form was mailed to each faculty member at Belmont University who was employed full-time in teaching undergraduate courses, and to

faculty who had some administrative responsibilities, but spent two-thirds of their time teaching (n=131). A cover letter (Appendix A) and stamped, addressed return envelope were included in the mailing. Coding of the mailing labels of the return envelopes allowed accurate, efficient follow-up of non-returns.

Respondents were asked to return the completed forms within two weeks. After a two-week response period had elapsed, follow-up postcards (Appendix C) were sent to remind participants to complete the form. After an additional response period of approximately three weeks had elapsed, telephone calls were made to non-respondents asking them to complete the questionnaire. In several cases, a second survey form was mailed to those persons who had not yet responded.

It was expected that the response rate for returned, completed questionnaires would be approximately 90 percent. The actual response rate was 70.2 percent or 92 returned surveys out of 131 mailed to faculty. Eighty-nine (67.9 percent) were usable. Two surveys were returned unanswered by non-teaching faculty members, and another was completed by an adjunct faculty member who had recently switched from full-time to part-time teaching.

When the survey instruments were collected, the data they included were entered into a computer for storage and analysis. The responses to each question on the survey were entered into the computer by groupings corresponding to the demographic variables in Section I of the survey, e.g. age, academic rank, years teaching experience, et cetera, and the data were analyzed by computer to yield percentage responses to each demographic variable, as well as statistical comparisons among the responses. The Mini-Tab package, version 5.0, was used for data analysis. The raw data were converted to percentages, and the relative frequencies were determined for the responses to each question in the survey.

Responses to the demographic variables were compared with questions in Section II of the survey--about current Belmont evaluation practices--perceived fairness, objectivity and reliability of student, administrative, peer and self-evaluations; the level of satisfaction faculty had with the four specific types of evaluation and with the Belmont evaluation procedures in general, and with whether faculty perceived that the four types of evaluation should be done at all. Because the variables such as age, gender, rank, et cetera, were categorical and statistical manipulations were intended to test for differences between groups, i.e. male versus female, chi-square statistical tests were then conducted to determine if any responses were by chance, or if any significant differences existed between responses that could be linked to age, gender, rank, years of teaching experience, school taught in, tenure status and teaching versus administrative status. These data served to answer the research question: "How do faculty at Belmont University perceive each of the four types of evaluation currently used to evaluate their teaching?"

Responses to the demographic variables were also compared to questions in Section III of the survey regarding the ideal pracatices of evaluation versus the current evaluation practices at Belmont University, e.g. what the ideal usage of information from student, administrative, peer and self-evaluation was as opposed to the actual usage of that information was, and the ideal versus real disposition of evaluation information. Chi-square statistics were performed to see if the responses were by chance, or if significant differences existed between responses due to age, gender, rank, years teaching experience, college taught in, tenure status and teaching versus administrative status. These data answered the research question, "How do the faculty's perceptions of the approaches currently being used to evaluate their teaching compare with their perceptions of what should be used?"

Answers to open-ended questions associated with the overall research questions were summarized at the ends of appropriate sections, and attention was given to any discernible patterns in open-ended responses.

The findings from this analysis are presented in Chapter 4. Findings and results of the study are organized and discussed in terms of the research questions guiding the study.

CHAPTER IV

DATA ANALYSIS AND FINDINGS

Introduction

The purpose of this study was to examine the attitudes and perceptions of faculty towards instructional evaluation at one institution, Belmont University in Nashville, Tennessee. Belmont had a teaching evaluation system in place which incorporated the four types of evaluation--student, administrative, peer and self, making it possible to examine the faculty's attitudes and perceptions of "what is" as well as what "should be" in terms of the four types of evaluation. A questionnaire was used to determine faculty perceptions about the current practice of evaluation and the ideal practice of evaluation. The 89 usable, returned questionnaires (a 67.9 percent return rate) were analyzed in terms of the research questions which guided the study:

1. How do faculty at Belmont University perceive each of the four types of evaluation currently used to evaluate their teaching?

2. How do the faculty's perceptions of approaches currently being used to evaluate their teaching compare with their perceptions of what should be used?

This chapter presents the data analysis and findings of the study. The presentation is organized and discussed in three sections. Demographic information is presented in the initial section. In the second section, research question one is considered along with

the data that answer that question. Anecdotal findings from open-ended responses relevant to research question one are also included in this section. The third section focuses on research question two and presents data that address the question, including anecdotal information gleaned from responses to relevant open-ended questions in the questionnaire.

Chi-square analyses were run to determine if there were significant differences (at the .01 level) in instructor response based on demographic variables, comparing each variable to each research question on the survey. Those variables were: gender, age, academic rank (instructor, assistant professor, associate professor and full professor), tenured or non-tenured, full-time teaching versus teaching plus carrying out limited administrative duties, school within the university (nursing, religion, humanities/education, the sciences [behavioral and natural], business and music), academic degree (doctorate, master's, bachelor's, associate's, doctoral candidate), and years of teaching experience (at Belmont, at other institutions, and total years experience). When the Mini-Tab program indicated too few responses in the cells for valid chi-square comparisons, such comparisons were not considered for purposes of this study.

Demographics

The demographic breakdown of the respondents included 36 females (40 percent) and 53 males (60 percent). The data on age of respondents were collapsed to expedite analysis from the division of five-year intervals to 10-year intervals. That translated to two percent of the respondents being in the 21-30 age group, 31 percent being in the 31-40 age range, 34 percent being between the ages of 41 and 50, and eight percent

being between 61 and 70 years of age.

In academic rank, only eight (9 percent) of the faculty were at an instructor level; 26 (29 percent) were assistant professors; 27 (30 percent) were associate professors, and the largest number of respondents, 28 (31 percent) were full professors.

In terms of tenure, 49 (53 percent) of the respondents were tenured, while 40 (45 percent) were not. Sixty-eight respondents (76 percent) were full-time instructors, while 16 (18 percent) taught primarily, but had some administrative duties (not more than one-third of the time). Five respondents indicated that they were non-teaching faculty, all librarians in this case.

The largest group of respondents, 32 (36 percent), were in the school of humanities/education, followed by 19 (21 percent) in the business school; 14 (16 percent) in the sciences; 12 (13 percent) in music; four (4 percent) in nursing, and three (3 percent) in religion.

A majority (61;69 percent) had doctorates; 22 (25 percent) held master's degrees; four (4 percent) were doctoral candidates, lacking only the dissertation; one (1 percent) had a bachelor's degree; and one (1 percent) had a two-year associate's degree.

Years of teaching experience in higher education were aggregated into five categories: 1 to 5 years of teaching experience; 6 to 10 years experience, 11 to15 years; 16 to 20; and over 20 years. Years of experience were tabulated in terms of years taught at Belmont, years taught at other institutions, and total years of teaching experience. The largest percentage of faculty (29;33 percent), had 1 to 5 years experience at Belmont, followed by 27 faculty (30 percent) who had been at Belmont 6 to 10 years. Nine faculty (10 percent) had taught at Belmont 11 to 15 years; 11 faculty (12 percent) had taught at Belmont 16 to 20 years, and 12 (13 percent) had been at

Belmont over 20 years; and 11 faculty (12 percent) had taught at Belmont 16 to 20 years.

All of the faculty had teaching experience at institutions other than Belmont. The highest number (34; 38 percent) had taught from 1 to 5 years at other institutions; 18 (20 percent) had taught from 6 to 10 years; six (7 percent) had taught 11 to 15 years; four (4 percent) had taught 16 to 20 years; and three (3 percent) had taught over 20 years at other institutions.

For total number of years teaching at Belmont and other institutions, seven (8 percent) had taught one to five years; 22 (25 percent) had taught a total of 6 to 10 years and 11 to 15 years respectively; 16 (18 percent) had taught 16 to 20 years, and 21 (24 percent) had taught over 20 years.

Research Question Number One: How do faculty at Belmont perceive each of the four types of evaluation used to evaluate their teaching?

The questions that generated the data which answered this question were Section 2, Question 9, concerning the fairness, objectivity and reliability of the four types of evaluation in use at Belmont (student, administrative, peer and self); Question 10, indicating how satisfied the respondents were with student, administrative, peer and self-evaluations; Question 12, concerning how satisfied, overall, respondents were with the current evaluation practices at Belmont; Question 11, regarding whether Belmont faculty thought that each of the four types of evaluation should or should not be done; and Question 15, regarding how effective respondents perceived each of the four types of evaluation procedures at Belmont to be for the purposes of improving teaching, improving programs/majors, and making administrative decisions. These questions will

first be examined in terms of general numbers and percentages, then they will be examined in terms of the demographic variables.

Fairness, Objectivity, Reliability

Table 1, Appendix D presents the data concerning the perceived fairness, objectivity and reliability of the four types of evaluation practiced at Belmont for the group as a whole. A majority thought that student evaluations were very fair or fair; 32 percent believed student evaluations were unfair or very unfair. In contrast, 14 percent considered administrative evaluation very fair or fair, while 68 percent thought administrative evaluations were unfair or very unfair.

Peer evaluations were regarded as very fair or fair by 45 percent of respondents, while eight percent believed peer evaluations were unfair or very unfair.

Similarly, self-evaluations were perceived as being very fair or fair by 63 percent of the respondents, while 13 percent thought self-evaluations were unfair or very unfair.

While favorable perceptions of fairness for student, peer and self-evaluations prevailed, objectivity ratings for the same three categories did not fare as well. Only 36 percent of the Belmont faculty rated student evaluation very objective or objective, compared to 58 percent who rated student evaluation non-objective or very non-objective. Administrative evaluation was perceived to be very objective or objective by 54 percent of respondents, while 30 percent thought administrative evaluation was non- objective or very non-objective. Peer evaluation was considered very objective or objective by 31 percent of the Belmont faculty, while 21 percent said peer evaluation was non-objective or very non-objective. Self- evaluation was rated very objective or

objective by 47 percent of the respondents, and non-objective or very non-objective by 33 percent of the respondents.

Reliability ratings were similar to objectivity ratings for student, administrative and peer evaluation. Thirty-seven percent said student evaluation was very reliable or reliable, while 56 percent rated student evaluation unreliable or very unreliable. Administrative evaluation was rated very reliable or reliable by 59 percent of the respondents, and unreliable or very unreliable by 26 percent of the respondents. Peer evaluation garnered a very reliable or reliable rating from 39 percent of the respondents, while 14 percent said peer evaluation was unreliable or very unreliable. Sixty-three percent said self-evaluation was very reliable or reliable, while 16 percent said self-evaluation was unreliable or very unreliable.

Satisfaction

Table 2, Appendix D presents the data concerning the respondents' perceived satisfaction with the four types of evaluation specifically, and with the overall evaluation system in general.

Belmont faculty were generally satisfied with administrative, peer and self-evaluation, but seemed ambivalent on student evaluation. Thirty-three (37 percent) indicated that they were satisfied with student evaluation at Belmont, while 34 (38 percent) said they were not satisfied with student evaluation. Forty-three (48 percent) indicated that they were satisfied with administrative evaluation, while 21 (24 percent) claimed to be dissatisfied with administrative evaluation. In the peer evaluation category, 29 respondents (33 percent) indicated that they were satisfied, while11 (12 percent) said they were dissatisfied with this type of evaluation procedure. The

majority, 45 (51 percent) indicated that they were satisfied with self-evaluation, while 15 (17 percent) were not satisfied with self-evaluation.

Faculty satisfied with the overall evaluation system at Belmont University barely eclipsed those claiming to be dissatisfied. Forty-one (46 percent) respondents expressed satisfaction with the overall evaluation procedures at Belmont, while 35 (39 percent) were not satisfied with the procedures.

Evaluation: To Do or Not to Do

Table 3, Appendix D contains data describing faculty responses to the question of whether they perceived that evaluation should be done at all. Respondents were almost overwhelming in their perceptions that all four types of evaluation--student, administrative, peer and self--should be done. Seventy-four (83 percent) of survey respondents said that student evaluations should be done. In the administrative category, 75 (84 percent)--the highest number of respondents in these four categories--said that administrative evaluation should be done. Sixty-two (70 percent) of the Belmont faculty responding thought that peer evaluation should be done, while 72 (81 percent) thought that self-evaluation should be done.

Effectiveness of Evaluation

Table 6, Appendix D contains data concerning perceived effectiveness of the four types of evaluation practiced at Belmont in improving teaching, improving program/major, and in making administrative decisions. Almost equal numbers of respondents perceived student and administrative evaluation to be effective as ineffective

in improving teaching, while peer evaluation was negatively perceived in this regard, and self-evaluation was highly regarded as a way to improve teaching. Thirty-five (39 percent) of the respondents rated student evaluations effective for improving teaching. Thirty-six (40 percent) respondents perceived student evaluations as ineffective in improving teaching.

Thirty-five (39 percent) said administrative evaluations were effective or very effective in improving teaching, compared with 40 (45 percent) who rated administrative evaluation ineffective or very ineffective in improving teaching.

Eighteen (20 percent) Belmont University faculty believed peer evaluation very effective or effective in improving teaching, while 35 (39 percent) believed peer evaluation ineffective or very ineffective in improving teaching.

Self-evaluation was rated very effective or effective in improving teaching by 48 (54 percent) faculty, and ineffective or very ineffective by 28 (31 percent) respondents.

None of the four evaluation types were perceived as particularly effective in improving program/major. Twenty-nine (32 percent) saw student evaluations as very effective or effective in improving program/major, while 45 (61 percent) of the respondents believed that student evaluations were ineffective or very ineffective in improving program/major. Similarly, administrative evaluation was considered very effective or effective in improving program/major by 31 (35 percent) and ineffective or very ineffective in improving program/major by 45 (51 percent) respondents. Fourteen (15 percent) considered peer evaluation very effective or effective in improving program/major, while 40 (45 percent) who responded rated peer evaluation ineffective or very ineffective in improving program/major.

Regarding Belmont faculty perceptions of the effectiveness of self-evaluation in

improving program/major, 35 (39 percent) of the respondents rated self-evaluation very effective or effective in improving program/major, while 40 (64 percent) rated self-evaluation ineffective or very ineffective in improving program/major.

Respondents tended to question the effectiveness of student, peer and self-evaluation in making administrative decisions. Only administrative evaluation emerged as being somewhat credible in this regard, and the respondents who favored administrative evaluation did not quite constitute a majority. Twenty-eight (31 percent) perceived student evaluations very effective or effective in making administrative decisions, while 48 faculty (54 percent) believed student evaluations to be ineffective or very ineffective in this regard. Forty-one (46 percent) respondents rated administrative evaluation very effective or effective in making administrative decisions, while 29 (32 percent) felt administrative evaluation was ineffective or very ineffective in making administrative decisions.

Twenty-one (23 percent) respondents thought that peer evaluation was very effective or effective in making administrative decisions, while 30 (34 percent) believed peer evaluation was ineffective or very ineffective in making administrative decisions. Self-evaluations were rated by 29 (32 percent) respondents as being very effective or effective in making administrative decisions, and ineffective or very ineffective by 38 (43 percent) respondents.

Gender

Gender of respondents had virtually no impact on perceptions of Belmont University faculty. No significant differences were found in terms of perceptions of the fairness, objectivity or reliability of, or satisfaction with any of the four types of

evaluation based on gender of respondents. Likewise, no gender-based differences were found in perceptions of whether the four types of evaluation should be done.

Data about effects of gender on faculty responses including chi-square analyses appear in Appendices E-1 through E-6.

Chi-square analyses revealed no significant differences in perceived effectivensss of student, administrative, peer and self-evaluations based on gender of respondents .

Age

Age of respondents seemed to have limited impact on differences in response to most of the questions of the survey. There were no significant differences in perceptions among the five age categories regarding the fairness of student, administrative, peer or self-evaluation; regarding the objectivity or reliability of student, administrative, peer or self-evaluations; regarding the perceived satisfaction level of the four types of evaluation done at Belmont University, or to the perceived overall satisfaction with he Belmont evaluation system, or to whether the four types of evaluation should be done. Likewise, there were no significant differences in perceptions due to age of respondents regarding the perceived effectiveness of the four types of evaluation in improving teaching, improving program/major or in making administrative decisions. Data concerning chi-square analyses of the impact of age on responses of faculty are contained in Appendices E-7 through E-12.

Academic Rank

The academic rank of respondents appeared to have no impact on the responses to survey questions pertaining to Research Question 1. Chi-square analyses revealed no significant differences in responses to perceived fairness, objectivity or reliability of the four types of evaluation based on academic rank of faculty; no significant differences based on academic rank to perceived satisfaction with any of the four types or with the evaluation system overall; no significant differences about whether evaluation should be done; and no significant differences in responses based on academic rank to perceived effectiveness of student, adminisrative, peer or self-evaluation for improving teaching, improving program/major or for making administrative decisions. Data supporting these findings, including chi-square analyses, are included in Appendices E-13 through E-18.

Tenure

Similarly, tenure status of repsondents seemed to have no effect on responses concerning Research Question 1. Chi-square analyses revealed no significant differences in responses based upon tenure status to the perceived fairness, objectivity or reliability of the four types of evaluation practiced at Belmont University; no significant differences in responses based upon whether faculty were tenured to the perceived satisfaction with student, administrative, peer or self-evaluations, or to the overall satisfaction level; no significant differences in responses based on tenure status as to whether or not the four types of evaluation should be done at Belmont University; and no

significant differences in faculty response based on tenure status to the perceived effectiveness of student, administrative, peer or self-evaluation for improving teaching, improving program/major, or in making administrative decisions. Appendices E-19 through E-24 contain data outlining impact on responses based on tenure status, including chi-square analyses.

Teaching/Administrative Status

Whether faculty taught exclusively or taught in addition to having some limited administrative duties had a limited impact on responses. Question Five in Section 1 of the survey concerned faculty status of teaching exclusively versus teaching and having no more than one-third of the time devoted to administrative duties. Chi-square analyses revealed neither significant differences in response to fairness, objectivity or reliability of any of the four types of evaluation done at Belmont based upon teaching status of respondents, nor to the satisfaction level of the four types of evaluation practiced at Belmont University. Concerning overall satisfaction, chi-square analyses did reveal, however, some differences in response based on whether faculty taught exclusively or taught and had some administrative responsibilities. Thirty-two (35 percent)--significantly more--faculty who taught exclusively said, overall, Belmont evaluation procedures were satisfactory, than the eight (9 percent) who taught and had some administrative responsibilities. Also, a very unsatisfactory overall rating was given Belmont's evaluation system by eight (9 percent) faculty who taught exclusively, significantly more than the one (1 percent) faculty member who taught and had some additional administrative duties.

Comparisons of teaching/administrative status using chi-square analyses

revealed that, on the question of whether the four types of evaluation done at Belmont should or should not be done, significant differences in response were present in all four categories--student, administrative, peer and self-evaluations. Significantly more of the respondents who taught exclusively (57;64 percent) said student evaluation should be done, compared with 16 (18 percent) faculty who taught and who had some administrative duties--almost the opposite of what one would expect.

Administrative evaluation fared similarly. Fifty-nine (66 percent) respondents whose duties were solely pedagogical believed administrative evaluation should be done, significantly more than the 14 (16 percent) who taught and had some administrative duties--again, almost the opposite of what one might expect.

Concerning peer evaluation, significantly more faculty who taught, only, (49; 55 percent) thought peer evaluation should be done, than did the 12 (13 percent) faculty who taught and who had some additional administrative tasks. Once again, this would seem to be almost the opposite of one could anticipate in such a situation.

On the question of whether self-evaluation should be done, significantly more (55; 62 percent) respondents who taught exclusively said self-evaluation should be done than did the 16 (18 percent) respondents who taught and had some administrative duties. Nine (10 percent) respondents who taught exclusively--significantly more--thought self-evaluation should not be done, than did the zero respondents who had some administrative duties added to their teaching responsibilities.

Chi-square analyses revealed no significant differences based on teaching versus administrative status to the perceived effectiveness in improving teaching, improving program/major, or in making administrative decisions of student, administrative, peer or self-evaluations. Data comparing instructor responses by teaching/administrative status to the survey questions regarding Research Question 1 are contained in Appendices

E-25 through E-30.

School

The school in which faculty taught seemed to have moderate levels of influence on faculty responses. Survey Question 6, Section 1 of the survey dealt with the school within the university faculty taught in, including the schools of nursing, religion, humanities/education, the sciences (behavioral and natural), business and music. No significant differences were found in terms of the perceived fairness, objectivity and reliability of student, administrative, peer and self-evaluation; in response to the question of whether or not the four types of evaluation practiced at Belmont should be done; or in the perceived effectiveness of the four types of evaluation in making administrative decisions. Only in the perceived fairness of one of the evaluation types--administrative evaluation--did chi-square analyses reveal significant differences in response based on the school in which faculty taught. Significantly more respondents (10;11 percent) in the business school rated administrative evaluation very fair than did the faculty in the sciences (6; 7 percent). No other significant differences in perceived fairness of administrative evaluation compared to school were found.

Significant differences in responses about the perceived effectiveness of only one of the four types of evaluation--self-evaluation--in improving teaching compared by school were revealed by chi-square analyses. Significantly more humanities/education faculty (17; 18 percent) believed that self-evaluation was effective in improving teaching than did the five (6 percent) business faculty who responded. No other school comparisons of perceived effectiveness of self-evaluation in terms of teaching improvement were significant.

Chi-square analyses revealed significant differences in responses to the perceived effectiveness of administrative and peer evaluation in improving program/major when compared by school, but no significant differences were found using the same comparison for student and self-evaluations. Concerning administrative evaluation's perceived effectiveness in improving program/major with regard to school in which faculty taught, only in the case of humanities/education faculty compared to business faculty was there significant difference. Significantly more humanities/education faculty (12 ;13 percent) believed administrative evaluation was effective in improving program/major than did the three (3 percent) business faculty responding to this question. No other significant differences in perceived effectiveness of administrative evaluation in improving program/major relative to school in which faculty taught were found.

Peer evaluation was rated effective in improving program/major by significantly more faculty in humanities/education (11; 12 percent) than in nursing and religion faculties respectively (2; 2 percent for each). No other significant differences in perceived effectiveness of peer evaluation in improving program/major relative to school in which faculty taught were discovered. Data comparing instructor response by school are found in Appendices E-31 through E-36.

Degree

The highest degree earned by respondents seemed to have a moderate influence on faculty response. Question 7 in Section 1 of the survey asked respondents to record their highest earned academic degree. Chi-square analyses revealed no significant differences in responses relative to highest academic degree earned by Belmont faculty in terms of

perceived fairness, objectivity, or reliability of student, administrative, peer or self-evaluations; in terms of satisfaction with student, administrative, peer or self-evaluation, or to the perceived overall satisfaction with the Belmont evaluation system; or in terms of the perceived effectiveness of any of the four types of evaluation in improving teaching, improving programs/majors, or in making administrative decisions.

There were significant differences based on highest degree held revealed by chi-square analyses among Belmont faculty regarding whether peer and self-evaluations should be done at Belmont, but no significant differences revealed concerning whether student or administrative evaluations should be done. There were also significant differences in the perceived effectiveness of only one of the four types of evaluation--student evaluation--in improving teaching and in improving program/major.

Forty-four (49 percent) of the Belmont faculty holding doctorates thought that peer evaluation should be done, significantly more than the 13 (15 percent) faculty with master's degrees. Significantly more (15; 17 percent) of the faculty with doctorates thought that peer evaluation should not be done than the three (3 percent) of the faculty with master's degrees.

In the self-evaluation category, significantly more (51; 57 percent) of faculty respondents with doctorates said they thought self-evaluation should be done, compared to 16 (18 percent) of the faculty with master's degrees who concurred. Significantly more respondents with doctorates (8; 9 percent) thought that self-evaluation should not be done than the zero with master's degrees who agreed.

Twenty-nine (33 percent) of the respondents with doctorates rated self-evaluation effective in improving program/major, significantly more than the the seven (8 percent) faculty holding master's degrees. Appendices E-37 through E-42 contain

information on instructor responses based on highest degree held.

Teaching Experience-Belmont

Belmont teaching experience had no apparent impact on respondents' answers to survey questions. Survey Question 8a, Section 1 examined faculty responses based on the years of experience in teaching at Belmont. Chi-square analyses revealed no significant differences in responses based upon years teaching experience at Belmont regarding perceived fairness, objectivity or reliability of student, administrative, peer or self-evaluations.

No significant differences were observed based on chi-square analyses either in the perceived satisfaction level with student, administrative, peer and self-evaluations with regard to years teaching experience at Belmont University, or in the perceived overall satisfaction level with regard to years teaching experience at Belmont.

Chi-square analyses revealed no significant differences in faculty response based on years teaching at Belmont University to the perception of whether the four types of evaluation should be done, and chi-square analyses revealed no significant differences based on years of Belmont teaching experience to the perceived effectiveness of any of the four types of evaluation--student, administrative, peer or self--in improving teaching, improving program/major, or in making administrative decisions. Data contained in Appendices E-43 through E-48 examing the impact of years teaching experience at Belmont on instructor response.

Teaching Experience-Other Institutions

Teaching experience at other institutions had negligible impact on instructor response. Survey Question 8b, Section 1 of the survey explored faculty response based on years of teaching experience at institutions other than Belmont University. Chi-square analyses revealed no significant differences in responses based on years teaching experience at other institutions with regard to fairness or objectivity of student, administrative, peer or self-evaluations, or to the reliability of student, peer or self evaluations; no differences in terms of satisfaction with evaluation individual evaluation procedures or with overall evaluation procedures; in terms of whether the four types of evaluation should be done, or in terms of the effectiveness of the four types of evaluation in improving teaching, improving program/major, or in making administrative decisions were noted. However, regarding the reliability of only administrative evaluations, chi-square analyses did reveal significant differences among Belmont faculty responses pertaining to years teaching experience at other institutions. Nine (10 percent) faculty with 1 to 5 years teaching experience at schools other than Belmont rated administrative evaluation unreliable, significantly more than the one (1 percent) Belmont faculty member with 6 to 10 years teaching experience at institutions other than Belmont who rated administrative evaluation unreliable. Appendices E-49 through E-54 contain data examining the impact of teaching experiences at other institutions on Belmont faculty responses.

Total Years Teaching Experience

Total years teaching experience had no effect on instructor response. Demographic Question 8c, Section 1 of the survey asked for the ranges of total years of teaching experience (at Belmont and all other institutions combined). Chi-square analyses revealed no significant differences in instructor responses based on total years teaching experience to the perceived fairness of student, administrative, peer or self-evaluation; no significant differences in response based on total years teaching experience to the perceived satisfaction with student, administrative, peer or self-evaluation, or in the responses of Belmont faculty based on total years teaching experience in higher education to the perception of overall satisfaction with Belmont's evaluation practices were revealed.

No significant differences in instructor response comparing total years teaching experience to the perception of whether student, administrative, peer or self-evaluation should be done were revealed, either, and no significant differences in instructor response based on total years teaching experience in higher education to the perceived effectiveness of student, administrative, peer or self-evaluation for the improvement of teaching, the improvement of program/major, or for the making of administrative decisions were noted. Appendices E-55 through E-60 contain data on the effects of total years teaching experience on instructor response.

Anecdotal Responses : Research Question One

Anecdotal material pertaining to Research Question One about the perceived fairness, objectivity and reliability of the four types of evaluation practiced at Belmont generated the largest number of anecdotal responses. Twenty-seven wrote comments pertaining to student evaluations. The anecdotal comments on student evaluation represented quite a cross-section of teaching disciplines. Eight reacted to student evaluations in primarily negative veins, citing evaluation questions that did not pertain to courses (e.g., questions on textbook selections), lack of objectivity of student evaluations, and the format of student evaluations (circling numbers to indicate perceptions, as opposed to giving specific written comments). One respondent added that written comments on student evaluations tended to be extreme in either direction--positive or negative--while another said only in written comments on student evaluations is there a full assessment. The rating type questions on student evaluations were categorized as difficult to understand due to lack of knowledge of rating distributions across campus. Another respondent questioned whether students had sufficient information on which to make the judgments they were asked to make on instructor evaluations. Yet another indicated a belief that student evaluations were based on some things beyond the instructor's control, e.g. attendance. One respondent said students were mainly concerned with personalities in their ratings. Another said the student evaluation forms at Belmont needed other categories to gather information which would be more helpful to instructors. And one instructor called for a complete re-tooling of the current student evaluation program at Belmont.

Perhaps the most profound comment on student evaluation came from the faculty

member who said: "Students must be questioned several times a semester, because they tend to allow the specific courses they are taking to color their generalizations. The best student evaluations come from those I've had in several courses over a period of four or five times."

There was no discernible pattern of responses in the seven anecdotal responses directed toward administrative evaluation. Interestingly, administrative evaluation appeared to be perceived as done with little consistency by five respondents. One instructor wrote: "Evaluations from department chairs are usually based on observation of classes as well as review of some other information." That was ostensibly a positive comment. But the same instructor went on to say: "Evaluation from deans, however, is based on little data." Another instructor said: "Administrative review can be colored by factors I consider to be invalid or beyond my knowledge."

Judging, in part, from the five anecdotal responses on peer evaluation, it would appear that peer evaluation at Belmont University is sporadic and not done in all schools. In the sciences, business and humanities/education, responses indicated that peer evaluation was not done. Apparently, some type of peer evaluation is practiced in the schools of religion, nursing and music. From those respondents who practiced peer evaluation, reactions were mixed. One respondent said the only peer evaluation was when an instructor asked a colleague to write a recommendation. Another said peer evaluations done at Belmont were non objective. One response indicated that peer evaluation was used for recommendations for tenure and/or promotion. Another respondent stated there was no formal, systematic form of peer evaluation at Belmont, but added: "Some faculty members expect to be heard and consulted and expect to have their wishes followed." But on the positive side, one faculty member said: "Peers solicited for counsel and evaluation on specific issues are the best forms of evaluation."

Finally, the four open-ended responses on self-evaluation indicated, as with administrative and peer evaluation, inconsistency in execution. One respondent said there is only infrequent formal self-evaluation. Another said self-evaluation was not objective, while yet another said Belmont self-evaluations were "useful for my purposes." The most cogent response on self-evaluation was, perhaps: "I am harder on myself than others are, but I try, as a rule, to tone down the negative, since people read self-evaluations as biased toward the positive."

Three respondents chose to react globally to evaluation in general, and those responses, by and large, were critical. One instructor said evaluation "leads one to do things which make one look good, and this is not necessarily good education." Another respondent (non-tenured) said: "I do not accept the concept of any of these evaluations. Evaluation should be eliminated." But the most reasoned response came from a faculty member who said: "Most evaluations are not objective. I think you need to stress that evaluation relates to teaching only (if that is what is intended) and not other factors such as service or research."

Anecdotal material from Question 10 on satisfaction with the four types of evaluation practiced at Belmont was the least prolific. Only seven respondents wrote additional written comments under this question, and four of the seven merely affirmed instructor satisfaction. Three responses linked to Question 10 seemed worthy of mention. One respondent said: "Regarding administrative, self and peer review: I would like to have true honesty. Regarding student evaluation, I would like evaluation criteria to be more specifically applicable to my work. Also, averages we are measured against are absurd and not relevant."

Another respondent wrote: " I am quite satisfied that I initiate most of the really productive evaluation. If I don't want to improve, then evaluation by students, etc. will

not help. I also find that trusted peers dialogue about teaching and offer suggestions." The third comment was: "We are trying. I like that."

Question 12 on overall satisfaction with Belmont evaluation, while out of numerical order, will be discussed now because of its obvious tie in with Question 10. Question 12 yielded 21 anecdotal responses, with only two of the total indicating that respondents were satisfied, overall, with the evaluation procedures. The two positive comments went on to suggest that improvements were much needed in Belmont's evaluation system.

The majority, 19 of 21 respondents, were dissatisfied, overall, with evaluation practices at Belmont. Criticisms dealt with matters ranging from student evaluation forms (a confusing scale) to bad timing (the last two weeks of the semester), and from student evaluation being over done, to too much time elapsing between when evaluation was done and when feedback was received (as long as four months). The longest narrative response was on student evaluation, and it contained this perception: "Student evaluations assess how contented students are with the course. Moreover, they ask students to evaluate things which they are not competent to judge, such as adequacy of coverage, knowledge of subject by teacher, and content of course. The administrators actually compare the numerical averages of student evaluations from one course to another, believing that you can compare averages (I learned the fallacy of this the second week of freshman statistics!)."

Anecdotal feedback on administrative evaluation was sparse--only three faculty responded--but the feedback tended to be more detailed, and negative in tone. One respondent said it was difficult for administrators to be honest in their evaluation of faculty because departments and schools were so small. The following comment, though, was particularly caustic: "Administrators don't know what do do. My department chair

reads my self-evaluation and echoes that in a report in his evaluation. The dean reads the chair's evaluation and my self-evaluation and echoes that in a report in his evaluation. They never observe my teaching or examine other aspects of my work. They just know if they have heard complaints."

Another faculty member indicated that administrative evaluations are a clue to knowing "what the boss wants," but expressed the desire that evaluations be reciprocated--that faculty should evaluate administrators. He or she added: "The dean's evaluations are based on student comments and numerical ratings, and on the chairman's evaluation. If you are not in tight with the dean and department chair, you are up s--t creek."

Interestingly, no anecdotal comments about self-evaluations, specifically, were given, and only one respondent made a comment on peer evaluation, which was that it needed to be more objective.

There were some comments which tended to sum up perceptions about satisfaction with evaluation in general. One faculty member expressed that all evaluation needed to be viewed more broadly and never targeted against a faculty member through application. This same faculty member suggested that a thorough and open appeals process should be in place. Another respondent indicated a belief that evaluation, overall, is not a thorough indication of true performance, and expressed dismay that there was no help given at Belmont to correct deficiencies revealed in evaluations. Another faculty member stated that evaluations were not thorough enough, overall, and not clearly related to the tenure/promotion process. This same respondent believed that an instructor at Belmont could have good evaluations and not receive tenure, while another instructor could have poor evaluations, yet receive tenure.

One instructor noted: "I used to (at another school) do my own evaluation with

my students--at mid-term and at finals--to find out what was happening in class. I also used to observe others' teaching and invite my peers to observe and respond to my teaching. All of this was more meaningful and effective than all the types of formal evaluation administered at Belmont. Now the formal process takes so much time and effort that my students and I have no time or energy left to devote to another procedure." Another wrote, regarding the purpose and utility of Belmont evaluation procedures: "I would like to have a sense of how all types of evaluation are used at Belmont, especially regarding tenure. I would also like the possibility for greater honesty, but I recognize that this is unrealistic. I sometimes feel, for all the time and nerves expended, that a lot of this feels like just going through the motions."

Anecdotal material gleaned from Question 11 on whether evaluation should be done revealed that all 21 responses indicated a belief that student, administrative, peer and self-evaluations should be done. There was a consensus that revisions were needed in all types of evaluation, but especially in student evaluation. As for the revisions suggested, most of the anecdotal information indicated that the format of student evaluations needed changing, away from a "like/dislike" orientation, for example, and also away from questions which students are less qualified to answer, e.g. "instructor understands material." One respondent said student evaluations should be more in-depth to include evaluation for advising, for example. Another said that student evaluations should be done less frequently, particularly for tenured faculty. One respondent suggested that student evaluations should be monitored by another faculty member.

Concerning the perception that some evaluation types served certain purposes more than others, one faculty member wrote: "Administrative evaluation should pertain to tenure/promotion issues only. Real evaluation of teaching--in any meaningful way--can only be done by the instructor, his or her peers, and his or her students." Another

respondent said: "Self-evaluation, I think, should be focused more toward addressing concerns we have, not just about teaching, but about things affecting our teaching." As somewhat of a summary regarding Question 11, one instructor said: "I would like for anyone who evaluates my work to observe it in process. Sometimes, this does not happen."

There were 15 open-ended responses to Question 15 on effectiveness of the four types of evaluation practiced at Belmont University in improving teaching, improving program/major, and in making administrative decisions. Nine of the 15 respondents expressed that they did not really know how effective any of the four types were in making administrative decisions. One faculty member expressed the belief that each type of evaluation is considered and probably has some bearing in administrative decision making, but wondered how much of a bearing each had in that decision making. Another respondent indicated that there was no rationale for administrative evaluation, and further indicated a belief that under the previous vice-president for academic affairs, it was that dean's own personal likes and dislikes that determined tenure/promotion decisions. Another faculty member wrote: "I think administrators see their evaluations as helpful and effective in reaching their decisions, but I am not sure how fair that is, since administrative evaluations, especially, may reflect issues unknown to the faculty members." Another faculty member indicated that administrative evaluation from department chairs can be reassuring and helpful, and that informed peer evaluation can be reassuring and helpful. Yet another respondent believed self-evaluation could improve programs/majors, but only indirectly.

Finally, one respondent summed up all evaluation types and their effectiveness in this perception: "Peer and self-evaluations can be extremely helpful for all three purposes; student evaluation may be helpful. Administrative evaluation is essentially

useless. Administrators don't have time to evaluate. Generally, they create more problems than they solve when they become involved in the classroom."

Summary

The data answering Research Question One seemed to suggest that Belmont University faculty considered administrative evaluation, overall, to be the most fair, objective and reliable type, although self-evaluation, surprisingly, was rated the fairest and most reliable type of the four. Peer evaluation and student evaluation fared similarly to each other, with marginal objectivity and reliability ratings, and peer evaluation edging out student evaluation in fairness. The only variable which affected the fairness ratings appeared to be age. Faculty in the 31-40 age group believed most strongly in the fairness of student evaluation.

Objectivity ratings were affected only by gender. Males tended to doubt the objectivity of student evaluation more than females. Reliability ratings were affected by gender, and number of years teaching experience at other institutions. Males consider student evaluation less reliable than females. The fewer the years teaching experience at other institutions, the more unreliable faculty perceived administrative evaluation to be.

Belmont faculty were more dissatisfied with student evaluation than any other type, and the one variable affecting satisfaction levels was teaching exclusivity. Faculty who taught exclusively were more satisfied with the overall evaluation procedures than faculty who taught and had some administrative duties.

Belmont faculty seemed to agree that all four types of evaluation should be done, but, surprisingly, faculty who taught exclusively thought all four types should be done

significantly more often than their counterparts who taught and had some administrative duties. And faculty with terminal degrees favored doing peer and self- evaluation more than faculty with other degrees. Student evaluation and peer evaluation were considered relatively ineffective in improving teaching, while self-evaluation was rated effective in improving teaching most often by Belmont faculty, and ratings were affected by the variables of gender, school and highest degree earned. Males believed less in the potential for student evaluation to improve teaching than females, but the numbers were only slightly more against the potential for improved teaching than for its potential. Males tended to believe self-evaluation was effective in improving teaching more than females; more humanities/education faculty believed self-evaluation effective in improving teaching than did faculty from other school, and more faculty with doctorates thought that student evaluation had the potential for improving teaching.

Self-evaluation was believed to be the most effective type of evaluation for improving program/major, and one demographic variables, apparently, had an impact on instructor responses. Administrative evaluations were considered the most effective types of evaluation for making administrative decisions; student evaluations the least effective type. Gender was the only variable which affected responses to these questions. More males believed that student evaluation was ineffective for making administrative decisions than did females.

Research Question Number Two: how do the faculty's perceptions of the approaches currently being used to evaluate their teaching compare with their perceptions of what should be used?

Survey Questions 13 and 14 in Section 2 were designed to elicit responses which would answer Research Question Two. Question 13 was divided into four sections, corresponding to the four types of evaluation practiced at Belmont University--student, administrative, peer and self-evaluation. These questions will first be examined in terms of general numbers and percentages, then they will be examined in terms of the demographic variables.

Usage of Student Evaluation

In Question 13 part one, respondents were asked if student evaluations at Belmont: were used to improve teaching, program/major and/or to make administrative decisions; should be used to improve teaching, improve program/major and /or to make administrative decisions; were and should be used to improve teaching, improve program/major, and/or to make administrative decisions. There was also a "Don't Know" response possibility in each of the above categories. Table 4, with percentages of total responses not tested against any demographic variables, displays data on Belmont faculty perceptions of the usage of the four types of evaluation. Respondents thought that student evaluation should be used to improve teaching and improve program/major, but were against using student evaluation for making administrative decisions.There was not a wide disparity between the actual and ideal uses of student evaluation to improve

teaching; there was more of a disparity between responses on actual and ideal uses of student evaluation to improve program/major and to make administrative decisions. The largest number of respondents (41;46 percent) said student evaluations should be used to improve teaching; 35 faculty (39 percent) believed student evaluations were used and should be used to improve teaching; and six (7 percent) said student evaluations were being used to improve teaching.

Almost half the respondents (44; 49 percent) said Belmont student evaluations should be used to improve program/major, while 19 (21 percent) checked that student evaluations both were and should be used to improve program/major. Four (4 percent) respondents indicated that they perceived student evaluations were used at Belmont to improve program/major.

Eleven (12 percent) surveys recorded that student evaluations should be used in making administrative decisions; 27 (30 percent) indicated that student evaluations both were and should be used in making administrative decisions. Twenty-eight (31 percent) perceived that student evaluations were used at Belmont to make administrative decisions.

Usage of Administrative Evaluation

Part Two of Question 13 asked Belmont faculty how administrative evaluation was used and should be used at the university. Respondents believed that administrative evaluation should be used more to improve teaching, and to improve program/major and less in making administrative decisions. There was a rather wide disparity between the actual and ideal responses on the usage of administrative evaluation to improve teaching, improve program/major and to make administrative decisions. Thirty-eight (43

percent) respondents said administrative evaluations should be used to improve teaching, while 19 (21 percent) respondents indicated that administrative evaluation was and should be used to improve teaching. Three (3 percent) said administrative evaluation was used at Belmont for that purpose.

Forty-two (47 percent) said administrative evaluation should be used in improving programs/majors, contrasted to the 13 (15 percent) respondents who indicated that administrative evaluations for improvement of program/major both were being used in this way and should be. Three (3 percent) faculty perceived that administrative evaluations were used at Belmont to improve programs/majors.

Ten (11 percent) believed that administrative evaluations should be used in making administrative decisions, while 32 (36 percent) of the respondents believed administrative evaluation was being used and should be used to make administrative decisions, and 16 (18 percent) perceived that administrative evaluations were used in administrative decision making.

Usage of Peer Evaluation

The third part of Question 13 dealt with Belmont faculty perceptions of the usage of peer evaluation at the university. More Belmont faculty felt that peer evaluation should be used to improve teaching than believed it should be used to improve program/major or make administrative decisions. There was also a much wider disparity between the actual and ideal perceptions of usage of peer evaluation for improving teaching, improving program/major and in making administrative decisions. Almost half (44; 49 percent) of the respondents indicated that peer evaluation should be used to improve teaching, while seven (8 percent) checked that peer evaluation both was

and should be used at the university to improve teaching, and two (2 percent) respondents believed that peer evaluation was used to improve teaching.

Twenty-two (25 percent) of the respondents checked that peer evaluation should be used to improve program/major, and only one (1 percent) respondent indicated that peer evaluation both was and should be used to improve program/major; zero checked that peer evaluation currently was used to improve Belmont's program/majors. Nineteen (21 percent) said peer evaluation should be used in making administrative decisions; only six (8 percent) perceived that peer evaluation was used for making administrative decisions.

Usage of Self-Evaluation

The last category of Question 13 concerned Belmont faculty perceptions about the use of self-evaluation for improvement of teaching, improvement of program/major, and in making administrative decisions. Most respondents thought that self-evaluation should be used most in improving program/major, and somewhat in improving teaching. Self-evaluation's use in making administrative decisions was questionable. There was some disparity between actual and ideal usage of self-evaluation to improve teaching, improve program/major and to make administrative decisions. Seven (8 percent) indicated that self-evaluation should be used to improve teaching, three (3 percent) indicated that self-evaluations actually were used at Belmont to improve teaching. Nineteen (21 percent) indicated that they believed self-evaluation both was and should be used to improve teaching.

Twenty-six (29 percent) indicated that self-evaluation should be used in improving program/major, and seven (8 percent) indicated that self-evaluations were

used in improving programs/majors at Belmont. Eighteen (20 percent) believed that self-evaluations both were used and should be used to make administrative decisions, and nine (10 percent) thought that self-evaluations were used to make administrative decisions.

Disposition of Evaluation Information

Question 14 asked respondents to indicate what they perceived was done and what should be done with information from the four types of evaluation practiced at Belmont -student, administrative, peer and self. Options for perceptions of dispositions were: 1. return to instructor, no response requested; 2. return to instructor for response; 3. not returning evaluation information; 4. nothing. Table 5 presents data on instructor perceptions of the disposition of the four types of evaluation practiced at Belmont University in percentages of total responses, not tested against any demographic variables.

Disposition of Student Evaluation Information

Surprisingly, an overwhelming number of respondents thought that nothing should be done with information from student evaluations, but a large percentage favored at least returning student evaluation information to instructors with no expectations. Wide disparity existed between the perceptions of the ideal and actual dispositions of evaluation information. Four (4 percent) believed that student evaluation information should be returned to instructors. Twenty-five (28 percent) of the respondents thought that information on student evaluation both was returned and should be returned to the

instructor, and the majority, 51 (57 percent) of Belmont faculty who responded to the survey thought that student evaluation information was returned to the instructor, but with no response required from the instructor.

Thirty-four (38 percent) thought that student evaluation information should be returned to instructors for response. Four (4 percent) indicated that they thought this procedure was practiced and should be practiced, and two (2 percent) indicated that, in their perception, information from student evaluation was being returned to instructors for responses.

Two (2 percent) respondents thought that student evaluation should not be returned to faculty, and zero said student evaluation information was not returned and should not be returned to instructors, while five (6 percent) believed student evaluation information was not returned under the current system. Four (4 percent) respondents indicated that nothing should be done with information from student evaluations, and zero respondents indicated that nothing was done with student evaluation information. A considerable majority, 85 (96 percent) believed that nothing was done and should be done with information from student evaluations.

Disposition of Administrative Evaluation Information

The second portion of Question 14 asked for respondents' perceptions of what was and what should be done with information gleaned from administrative evaluations. The options for this question were the same choices found in student evaluation: "return to instructor with no response requested;" "return to instructor for response;" "not returned to instructor" or "nothing." Respondents almost overwhelmingly favored the option of responding to administrative evaluation information over other possible

dispositions, and there was wide disparity between the perceptions of the actual and ideal dispositions of evaluation information.

In the category of "return. . .no response requested," five (6 percent) instructors indicated that administrative evaluation information should be returned to instructors with no response expected; 11 (12 percent) respondents believed that administrative evaluation information both was and should be returned to instructors with no response expected, and 18 (20 percent) indicated that administrative evaluation information was returned to instructors at Belmont with no response expected.

When asked if administrative evaluation information should be returned to instructors for responses such as instructor comments or plans for improvement, 32 (36 percent) indicated that information regarding administrative evaluation should be returned to instructors for response, while 10 (11 percent) said administrative evaluation information both was and should be returned to instructors for responses. Two faculty members (2 percent) indicated that administrative evaluation information was returned to instructors for responses.

Two (2 percent) faculty indicated that administrative evaluation information should not be returned to instructors; two (2 percent) communicated the belief that administrative evaluation information both was not returned and should not be returned to instructors, and 16 (18 percent) faculty indicated that administrative evaluation information was not returned.

Finally, Belmont faculty were asked to record their perceptions about nothing being done with administrative evaluation information. Only one (1 percent) faculty member indicated that nothing should be done with administrative evaluation information, and one (1 percent) faculty member also believed that nothing was done with administrative evaluation, and this was the proper disposition, while 11 (12

percent) indicated that they thought nothing was being done with information gleaned from administrative evaluation information.

Disposition of Peer Evaluation Information

Question 14's third section dealt with peer evaluation. Instructors were asked their perceptions as to what was and should be done with evaluation information obtained from peer evaluation--"returned to instructor/no response expected; returned to instructor for some response; not returned; or nothing." Most respondents favored the option of responding to peer evaluation, and the disparity between the perceptions of the actual and ideal dispositions of evaluation information was wide in most instances. An equal number, five (6 percent), indicated that peer evaluation information was being returned to instructors with no response expected, and that peer evaluation information should be returned to instructors with no expectations of a response. Twenty-three (26 percent) thought that peer evaluation information should be returned to instructors for a response. Only one (1 percent) faculty member believed that peer evaluation information was being returned to Belmont faculty for response. Ten (11 percent) respondents indicated that they believed evaluation information from peer reviews was returned to and should be returned to instructors with no expectation of response from faculty.

One (1 percent) respondent said that peer evaluation information should not be returned, and one (1 percent) respondent believed that peer evaluation information was not being returned and that such information should not be returned, while 11 (12 percent) respondents believed that peer evaluation information was not being returned.

Five (6 percent) faculty indicated that nothing should be done with peer

evaluation information, and 27 (30 percent) believed that nothing was being done with evaluation information gleaned from peer review.

Disposition of Self-Evaluation Information

Finally, Belmont faculty were asked their perceptions of what was done and should be done with information gleaned from self-evaluations. Most respondents favored being able to respond to self-evaluations, and wide disparity existed between the perceptions of the actual versus the ideal dispositions of evaluation information. Only two (2 percent) respondents thought that self-evaluation information should be returned to instructors with no response required. Fourteen (16 percent) believed that self-evaluations were being returned to instructors with no response requested, and that this was what should be done. Twenty-three (26 percent) of the respondents indicated that information from self-evaluations should be returned to instructors for response. Nine (10 percent) believed that self-evaluation information was being returned to instructors for responses, and that this procedure should be followed.

Only one (1 percent) respondent thought that not returning self-evaluation information was what should be done, while four (4 percent) believed that self-evaluation information was returned and should be returned to faculty. Thirteen (15 percent) respondents believed that self-evaluation information was not returned at all.

Four (4 percent) faculty members believed that nothing should be done with self-evaluation information, and four (4 percent) believed that nothing was being done and nothing should be done with self-evaluation information at Belmont, but 13 (15 percent) indicated that nothing was being done with self-evaluation information.

Gender

Gender of respondents apparently had limited impact on the perceptions of ideal versus actual uses of evaluation and evaluation information. No significant difference was found in terms of perceived uses of student or administrative evaluation, the perceived disposition of evaluation information for student, administrative and peer evaluation. On the perceptions of the usage of peer and self evaluations, chi-square analyses did reveal some significant differences in responses based on respondents' gender.

Regarding Belmont faculty perceptions of usage of peer evaluation, chi-square analyses revealed significant differences based on gender only in the "Don't Know" category. Twenty-two (25 percent) of the male faculty members, significantly more, indicated they did not know if peer evaluation was used at Belmont, compared to only five (6 percent) female faculty.

On the perceived usage of self-evaluation to improve program/major and to make administrative decisions, chi-square analyses did reveal significant differences in responses based on gender. Significantly more female faculty, 15 (17 percent) thought self-evaluation was being used and should be used to improve program/major, than the seven (8 percent) of the male faculty who concurred.

Regarding the perceptions of the use of self-evaluation to make administrative decisions, eight (9 percent) male faculty respondents believed that self-evaluation was used to make administrative decisions, significantly more than the one (1 percent) Belmont female faculty member who agreed. Fourteen (15 percent) female respondents perceived that self-evaluations were and should be used to make administrative decisions, significantly more than the five (6 percent) male respondents who

concurred.

There were significant differences noted between male and female faculty responses as to the disposition of self-evaluation information based on chi-square analyses. Eleven (12 percent) of the female Belmont faculty respondents believed that self-evaluation information was returned to instructors with no response expected, significantly more than the three (3 percent) male faculty respondents who agreed with this perception. Data concerning the impact of gender on perceived usage and disposition of evaluation types and information are found in Appendices E-4 and E-5.

Age

Age had a limited influence on the perception of the uses of types of evaluation and on the disposition of evaluation information. Chi-square analyses revealed no significant differences, based on age, to the perceived usage of student evaluation, administrative evaluation or self-evaluation. However, chi-square analyses did reveal a significant difference in perception among the five age groups on the usage of peer evaluation, but only on the perception that peer evaluation was used to make administrative decisions (there were no significant differences in responses based on age categories to the use of peer evaluation for improving teaching or improving program/major). Significantly more Belmont faculty in the 31-40 age grouping (12; 13 percent) believed peer evaluation should be used to make administrative decisions, compared to four (4 percent) respondents in the 41-50 age category, one (1 percent) in the 51-60 age group, and zero in the 61-70 age category.

Chi-square analyses revealed no significant differences in responses based on age to the perceived disposition of student, peer or self-evaluation information, but

regarding the perceived disposition of administrative evaluations, chi-square analyses revealed that a significantly higher number of respondents (7; 8 percent) in the 41-50 age category perceived that nothing was done with administrative evaluation compared to the other age categories (21-30, 31-40, 51-60, 61-70), each of which recorded only one (1 percent) response in the respective categories. None of the other disposition possibilities for administrative evaluation, when subjected to chi-square analyses, showed significant response differences based on age. Appendix E-10 contains data comparing ages of respondents to perceived usage of student, administrative, peer and self-evaluation at Belmont University.

Academic Rank

Chi-square analyses revealed no significant differences in instructor response based on academic rank to the perceived usage of student, administrative, peer or self-evaluations, or to the perceived disposition of student, administrative, peer or self-evaluation information. Appendices E-16 and E-17 contain data concerning the effect of academic rank on instructor response regarding usage and disposition of evaluation information.

Tenure

Chi-square analyses revealed neither significant differences in faculty responses based on tenure status to the perceived usage of student, administrative, peer or self-evaluation, nor significant differences based on tenure status to the perceived disposition of student, administrative, peer, or self-evaluation information. Appendices E-22 and E-23 contain data on the effect of tenure status on the responses of Belmont faculty regarding the perceived usages of the four types of evaluation practiced at the university and on the perceived disposition of information gleaned from the four types of evaluation.

Teach/Administrative Status

No significant differences were found though chi-square analyses to the perceived usage of student, administrative, peer or self-evaluation based on teaching status. Chi-square analyses also revealed no significant differences in response based on teaching/administrative status to the perceived disposition of student, administrative, peer or self-evaluation information. Appendices E-28 and E-29 contain data comparing teaching/administrative status to the perceptions of Belmont University faculty on the usage of the four types of evaluation used and on the disposition of evaluation information

School

The school in which faculty taught at Belmont University had a rather negligible impact on the perceptions regarding the use of the four types of evaluation at the university, and also on the perceived disposition of the evaluation information gleaned from the four types of evaluation. No significant difference was found with regard to student, administrative or peer evaluation's potential for improving teaching, improving program/major or for making administrative decisions, or in the disposition of student, peer or self-evaluation information. Significant differences were found with regard to self-evaluation's potential to improve teaching, and in the disposition of administrative evaluation information.

Significantly more faculty in humanities/education (18; 20 percent) thought that self-evaluation both was and should be used to improve teaching compared to business faculty (4; 4 percent). No other statistical differences regarding perceived use of self-evaluation compared by any other school was discovered, and no statistically significant differences were found regarding self-evaluation's use in improving program/major or in making administrative decisions.

For administrative evaluation, chi-square analyses revealed significant differences in responses among Belmont faculty regarding perceived disposition of evaluation information based on the school in which respondents taught in only one case--humanities/education faculty and business faculty for the category "Return to Instructor/No Response Requested" (no other categories were affected by school). Significantly more humanities/education faculty (11; 12 percent) thought that

administrative evaluation information was returned to instructors with no response expected than did business faculty (2; 2 percent). Also, signficantly more faculty in humanities/education (11; 12 percent) thought that administrative evaluation information both was and should be returned to instructors with no response expected than did business faculty (zero).

No other schools or categories of responses for disposition of administrative evaluation information were significantly different. Appendices E-34 and E-35 contain data comparing Belmont University faculty perception of the usage of the four evaluation types and of the disposition of evaluation information to the school in which faculty taught.

Degree

Highest degrees held by respondents had virtually no effect on faculty responses to usage of evaluation types and to the possible dispositions of evaluation information. No significant differences in responses to perceived usage of evaluation based on degree held were found in student, administrative, peer or self-evaluation, or in perceived usage of student evaluations for the purposes of improving teaching, improving program/major or in making administrative decisions, and no significant differences were observed based on academic degrees held to the perceived disposition of student, administrative, peer or self-evaluation.

Appendices E-40 and E-41 contain data comparing degrees held by respondents to perceptions of usage of evaluation types and disposition of evaluation information.

Years Experience-Belmont

Belmont University teaching experience had very limited impact on instructor response regarding perceived usage of evaluation types and disposition of evaluation information. Chi-square analyses revealed no significant differences in instructor response based on years taught at Belmont to the perceived usages of student, administrative, peer or self-evaluations for the improvement of teaching, the improvement of program/major or for making administrative decisions.

Appendices E-46 and E-47 contain data comparing years of Belmont University teaching experience to faculty perceptions concerning usage of the four evaluation types and the disposition of evaluation information.

Years Experience-Other Institutions

Years teaching experience at other institutions had negligible impact on instructor perceptions regarding the usage of the four types of evaluation or on the perceived disposition of evaluation information. Chi-square analyses revealed no significant differences in responses of Belmont University faculty based on years teaching experience at other institutions to the perceived usage of the four types of evaluation done at Belmont, or to the disposition of information gleaned from student, administrative, peer or self-evaluation.

Appendices E-52 and E-53 contain data comparing years teaching experience at other institutions to the faculty perceptions of the usage of the four types of evaluation at Belmont University and the disposition of evaluation information.

Total Years Experience

Chi-square analyses showed no significant differences in instructor responses based on total years teaching experience to the perceived usage of student, administrative, peer or self-evaluation for the improvement of teaching, the improvement of program/major, or for the making of administrative decisions (See Appendix E-58).

No significant differences in instructor responses based on total years teaching experience were found in the perceived disposition of information gleaned from student, administrative, peer or self-evaluation, based on chi-square analyses (See Appendix E-59).

Anecdotal Responses

Question 13, about how each type of evaluation practiced at Belmont University was used versus how faculty perceived it should be used yielded a variety of 16 anecdotal responses. Six responses tended to be general and vituperative, with responses like "our system is far from ideal" to "let's start over."

As for student evaluation, one respondent admitted to having reservations about giving too much weight to student evaluation in making administrative decisions. One faculty member was disgruntled over not being rehired because of student evaluations: "My dean has said: 'We might have been willing to renew your contract if your teaching had been better.' But the only data for my teaching competence was student evaluations. His judgment might have been accurate, but the data provided by student evaluations

alone is insufficient to provide grounds for that judgment."

One respondent's statement on student evaluation seemed particularly profound: "All student evaluation must be interpreted in light of the specific course, its role in the curriculum (e.g., major versus service course), the prevailing student attitudes toward the course, and must be tracked over several semesters to establish a base line." One faculty member suggested that administrative evaluation for decision making be separated from evaluation for teaching purposes, and one theme seemed to emerge regarding peer evaluation, and that was that peer evaluation currently practiced at Belmont was done informally, irregularly and voluntarily, and that peer evaluation may not include observation of teaching. One faculty member admitted having to request peer evaluation.

One faculty member stated that self-evaluation should never be done, and another comment on self-evaluation led to an overall anecdotal assessment of what was done with the evaluation information: "I have completed self-evaluations, requested by the dean of my school, in past years. While this information was presumably placed in another, more permanent file, its exact location was never shared, nor were the ramifications of such evaluations ever discussed by the dean or any other person. The whole evaluation process at Belmont is something of a mystery. It is purported to be a way of improving classroom effectiveness, but its real purpose is to intimidate faculty into silence."

One respondent stated that all types of evaluation were meant to assist instructors in becoming better teachers, but another faculty member wrote a statement which seemed to sum up how evaluation should be used: "All evaluation should be used carefully in making tenure and promotion decisions. Any questions about evaluation should be addressed in a meeting with the person evaluated. I try to use all evaluation to improve teaching."

There were 15 anecdotal comments on Question 14 concerning what is done and what should be done with the four types of evaluation information. Interestingly, there were more comments on self-evaluation (five) than on student evaluation (four), and there were equal numbers of comments on administrative and peer evaluations (three on each).

Comments on student evaluations showed a recurring pattern of favoring that the execution of student evaluations be at teacher discretion, but that, if done, then they should be returned to instructors quickly, so that not only could plans for improvement be made, if indicated by the evaluations, but that faculty be allowed to explain and react to those evaluations. There was one negative comment on the format used in student evaluations at Belmont.

As for administrative evaluations, there was some doubt as to whether they should be done. Comments about responses to administrative evaluations were rather consistent. Respondents thought, in the main, that if administrative evaluations were done, then faculty should be allowed to respond to those evaluations in writing, and in a meeting. There was also a rather clear indication that the opportunities for faculty to respond after being evaluated by administration are limited, since that response is not currently made in writing. One instructor phrased it this way: "There are times I feel there are things I would like to say, but I know I can't."

Comments on peer evaluations indicated that faculty who responded believed peer evaluations were done too rarely, too informally, and too randomly. One faculty member admitted being in favor of peer review, but only if there were no administrative evaluations done.

The comments on self-evaluation tended to support the idea that self-evaluation should be returned to instructors and that administrators of some kind should discuss

self-evaluations with instructors, perhaps at a meeting wherein those evaluations could be used to make future plans and goals.

Summary

The data answering Research Question Two seemed to suggest that, according to the perceptions of Belmont University faculty, administrative evaluation, above all, should be done, and that student and self-evaluation, also, should be done, but there was some doubt about whether peer evaluation should be done. The variables having the most significant impact on responses were school in which faculty taught and highest degree held.

Humanities/education faculty tended to react more strongly to administrative evaluation, supporting its use, but only mildly supporting the use of self-evaluation. Faculty with doctorates tended to be more supportive of student evaluation.

As for how the four types of evaluation should or should not be used, the consensus seemed to be that student, administrative and peer evaluation should be used for teaching improvement, that student and administrative evaluation could be used for improving program/major, and that administrative evaluation could be used to help in the making of administrative decisions. Gender influenced responses on the usage of the types of evaluation, as did age, school in which faculty taught, highest degree held and years teaching experience at Belmont University. Female faculty tended to favor self-evaluation to improve program/major; males supported the use of self-evaluation to make administrative decisions. Faculty in the 31-40 age category believed peer evaluation was used to make administrative decisions. Humanities/education faculty supported the efficacy of administrative evaluation to improve teaching, improve

program/major and make administrative decisions, and the use of self-evaluation only for improving teaching. More faculty with doctorates thought student evaluation was used to improving teaching. More faculty with 11-15 years teaching experience at Belmont thought self-evaluation was/should be used to make administrative decisions than faculty with 6-10 years Belmont teaching experience.

As for faculty perceptions about ideal versus actual disposition of evaluation information, most faculty agreed that student evaluation was being returned to instructors, and that was as it should be, but respondents believed that nothing should be done with this information. No other clear perceptions regarding disposition of evaluation emerged. Gender and age had an impact on responses in this regard, and so did school, highest degree held and years teaching experience at other institutions. More female faculty thought that self-evaluation information was and should be returned to instructors with no expectations for response than male faculty. More faculty in the 41-50 age group believed that nothing was done with administrative evaluation. Humanities/education faculty ran counter to business faculty in supporting the idea that administrative evaluation need not be returned to faculty. Faculty with terminal degrees also supported returning student evaluation information to instructors with no expectations for response. More faculty with 1-5 years teaching experience at other institutions believed that nothing was done with self-evaluation information than faculty with 6-10 years teaching experience at other institutions, but more faculty with 6-10 years teaching experience at other institutions thought that nothing should be done with self-evaluation information than faculty with 1-5 years teaching experience at other institutions.

This chapter has presented data that resulted from Belmont University faculty members' responses to the survey questionnaires. Chapter Five presents a discussion of

the findings of these data with the conclusions and recommendations that are a result of the findings.

CHAPTER V

SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Introduction

This chapter is organized into four sections. A summary of the study is presented in the first section. A summary of the findings is presented in the second section. Conclusions and discussion are presented in the third section, and recommendations for further research are presented in the final section.

Summary of the Study

A considerable body of literature has focused on the frequency, validity and credibility of faculty evaluation in higher education almost to the exclusion of research into the attitudes and perceptions of higher education faculty toward evaluation. Except for basically anecdotal findings, research on the attitudes and perceptions of faculty in higher education regarding the execution of evaluation, distribution of results, and purposes of evaluation was largely lacking. There was also virtually no research into faculty perceptions about how and what evaluation should be as compared to what it is.

The purpose of this study was to examine the attitudes and perceptions of faculty toward instructional evaluation by investigating the attitudes and perceptions of higher

education faculty at one institution, Belmont University, Nashville, Tennessee, which has an evaluation

system in place that encompasses the four most widely used types of evaluation--student, administrative, peer and self-evaluation. Thus it was possible to examine the faculty's attitudes and perceptions of "what is," as well as what "should be," in terms of the four types of evaluation. The following questions quided the study:

1. How do faculty at Belmont University perceive each of the four types of evaluation currently used to evaluate their teaching?

2. How do the faculty's perceptions of the approaches currently being used to evaluate their teaching compare with their perceptions of what should be used?

Data were collected by means of survey questionnaires that were mailed to the entire population of full-time undergraduate teaching faculty at Belmont University, Nashville, Tennessee. The population consisted of 131 full-time instructors, none of whom devoted more than one-third of their time to any administrative tasks. Of the 131 surveys mailed, 92 were returned and 89 were usable, which provided an overall return rate of 67.9 percent.

Data were coded, converted to perecentages and subjected to analysis, primarily through the use of the chi-square subprogram of the Mini-Tab package, version 5.0, computer program.

Summary of the Findings

Research Question 1--How do faculty at Belmont University perceive each of the four types of evaluation currently used to evaluate their teaching?

Belmont faculty generally perceived student evaluations as fair, but not objective and not reliable. Administrative evaluations were perceived to be fair, relatively objective and generally reliable. Peer evaluations were considered to be more fair than student evaluations, but less fair than administrative evaluations or self-evaluations, and peer evaluations were seen as marginally objective and marginally reliable.

Self-evaluations were considered to be the most fair and reliable type of evaluation practiced at Belmont University, but were rated only moderately objective; they also received the highest reliability rating of the four types of evaluation.

Perceptions on fairness of student evaluation declined with the age of the faculty. The oldest age group, 61-70, perceived student evalutions to be unfair, but faculty ranging in ages from 31 to 60 felt that student evaluations were fair.

Academic rank or tenure had little impact on the perception of fairness, but surprisingly, almost four times more faculty who taught exclusively believed administrative evaluation was fair than faculty who taught, plus had some administrative duties. By about four to one, faculty who taught exclusively rated self-evaluations fair over faculty who taught plus had administrative duties.

Faculty in humanities/education affirmed the fairness of student, administrative and self- evaluations more than their colleagues in nursing, religion, the sciences, business or music. Faculty in the sciences at Belmont tended to support the

claim of fairness of administrative evaluation more than faculty from other schools, and faculty with 1 to 5, 11 to 15 and 16 to 20 years Belmont teaching experience affirmed the fairness of administrative evaluation.

Belmont faculty perceived that student evaluation lacked objectivity, while administrative evaluation was perceived as the most objective of the four types. Twice as many male faculty felt student evaluation was non-objective compared to female faculty. And almost 10 times more faculty who had only teaching responsibilities felt administrative evaluations were objective than faculty who had added administrative duties; again, a departure from the expected.

More humanities/education faculty felt student evaluations were not objective than instructors in any other school, and regarded self-evaluation objective most often; business faculty led the other schools in numbers who supported the objectivity of administrative evaluation.

The fewer the years of Belmont teaching experience, the less objective the respondents perceived student evaluation to be. But the group with 20 plus years of Belmont teaching experience believed student evaluations to be objective. Almost the reverse was true of the perceived objectivity of administrative evaluations; the less Belmont teaching experience respondents had, the more likely they were to rate administrative evaluation objective, while the more years Belmont teaching experience the respondents had, the less likely they were to rate administrative evaluation as objective, particularly the respondents in the 20 plus years Belmont teaching experience group.

Belmont faculty generally agreed that the most reliable type of evaluation was self-evaluation, and more male faculty believed student evaluation was unreliable or very unreliable than their female counterparts. For administrative evaluations, almost

four times more faculty who taught exclusively thought that this type was reliable than respondents who taught and had some administrative responsibilities, again a departure from what might be expected.

Humanities/education faculty tended to rate student evaluations unreliable most often. For administrative evaluation, faculty in the sciences and business tended to rate this type of evaluation more reliable than faculty from other schools. The reliability of self-evaluation was affirmed by faculty from humanities/education.

The numbers and percentages seemed to show that the greater the number of years teaching experience faculty had at other institutions, the greater the confidence level faculty had with administrative evaluation.

There was inconsistency in the execution of three of the four types of evaluation practiced at Belmont University, specifically administrative, peer and self-evaluation. The only type of evaluation done consistently was student evaluation.

Belmont faculty seemed to indicate that student evaluation had been overemphasized at the university, almost to the exclusion of other types of evaluation, and they felt that current evaluation procedures offered no support in solving their instructional problems. Belmont faculty were slightly more dissatisfied than satisfied with the current evaluation system.

In general, Belmont University faculty expressed more dissatisfaction with student evaluation than with any other type. In the overall satisfaction rating, unsatisfactory and very unsatisfactory categories had roughly two percent more faculty responses than very satisfactory and satisfactory categories.

The majority of respondents who taught exclusively expressed moderate to extreme dissatisfaction with student evaluation; this was about four times the number of respondents who taught and who also had some administrative duties who agreed.

With administrative evaluation, the situation was almost the opposite. Over 40 percent of the faculty who taught exclusively were satisfied with this types of evaluation, as opposed to 12 percent of the faculty who taught plus had administrative duties. As for self-evaluation, almost half the respondents (47 percent) thought this type of evaluation was satisfactory, while 13 percent of their colleagues who had some administrative duties concurred.

Business faculty believed student evaluation was satisfactory most often. Administrative and peer evaluation found more advocates among humanities/education faculty, while the faculty from nursing expressed dissatisfaction with peer evaluation. Self-evaluation received the highest satisfaction rating among all the four types by almost two to one among the schools, with humanities/education leading in positive response percentages, and only in humanities/education did the most respondents express dissatisfaction with the overall Belmont evaluation procedures.

Faculty with the lowest number of years teaching experience (1 to 5; 6 to 10) and highest number of years teaching experience (16-20; 20 plus) tended to rate satisfaction with student evaluation low, while faculty with 11 to 15 years total teaching experience rated student evaluation satisfactory by a margin of two to one.

The majority of Belmont University faculty agreed that student, administrative and self-evaluation should be done, with the most responses being in favor of administrative evaluation being done, and the least responses being in favor of doing peer evaluation. Self-evaluation had the highest approval rating in terms of the perception that it should be done.

Student evaluation got the most support from faculty in the sciences and in business. With administrative evaluation, the most opposition came from faculty in humanities/education and religion. Administrative evaluation got the most support from

faculty who taught exclusively, while among faculty who taught and who had some administrative responsibility, student and self-evaluations were the most often listed as should be done, and three times more faculty who taught exclusively said peer evaluation should not be done than faculty who taught plus had some administrative duties.

Faculty at Belmont University were not certain improvement of instruction was being fostered by current evaluation practices. Most of the Belmont faculty said student, administrative and peer evaluation were ineffective in improving teaching. Faculty indicated overall that they believed self-evaluation to be more effective than ineffective in improving teaching.

Over five times more faculty who taught exclusively rated student evaluation effective in improving teaching than the instructors who taught and had some administrative duties. Over three times more faculty who taught exclusively thought administrative evaluation was effective for improving teaching than did those who had additional administrative duties, a surprising departure from what one might expect.

Rating self-evaluation ineffective in improving teaching were 20 times more faculty who taught exclusively than faculty who taught and had some administrative responsibilities, another surprising revelation.

More business faculty believed student evaluation was effective in improving teaching than any other schools, but more faculty from business rated administrative and self-evaluation ineffective in improving teaching.

Faculty with doctorates believed student evaluation was ineffective in improving teaching, and faculty with doctorates, master's degrees, bachelor's degrees, associate's degrees and who were ABD indicated that self-evaluation was effective in improving teaching.

Belmont faculty questioned the effectiveness of all four types of evaluation for

improving program/major, but said self-evaluation was the type most likely to be effective in that regard, followed by administrative evaluation, and then student evaluation. Significantly more male instructors said student evaluations were ineffective in improving program/major than did females.

Seven times more faculty thought administrative evaluation was ineffective or very ineffective in improving program/major than faculty who taught plus had some administrative duties.

Only in the school of business did more faculty perceive that student evaluation was ineffective in improving program/major, and only faculty from the sciences, business and music pronounced administrative evaluation ineffective in improving program/major. Faculty with doctorates rated student evaluation ineffective in improving program/major by almost two to one.

Most of the faculty indicated that administrative evaluation was the most effective type of evaluation to be used for making administrative decisions. More males at Belmont considered student evaluation ineffective in helping make administrative decisions than did female Belmont instructors by a margin of over two to one.

Almost seven times more respondents who taught exclusively believed administrative evaluation was ineffective in making administrative decisions than colleagues who taught plus were assigned some administrative tasks, more or less what one would expect. Only in the school of business did more faculty indicate they believed student evaluation was effective in making administrative decisions.

Research Question 2--How do the faculty's perceptions of approaches currently being used

to evaluate their teaching compare with their perceptions of what should be used?

Belmont faculty seemed to conclude that evaluation of teaching could be beneficial. As for how the four types of evaluation should or should not be used, the Belmont faculty, as a whole, seemed to agree that student evaluation should be used to improve teaching, that it should be used to improve program/major, and that it was already being used to make administrative decisions. They denounced the use of student evaluation in making administrative decisions.

For administrative evaluation, the response was similar, with most faculty believing administrative evaluation, while often based on too little data, should be used for improving teaching, improving program/major, and also for making administrative decisions. In fact, administrative evaluation was the only type acceptable for use in making administrative decisions.

Almost half the respondents said peer evaluation should be used to improve teaching, but seemed either skeptical or unsure of peer evaluation's place in improving program/major, since peer evaluation was not considered objective and was primarily done informally and irregularly and did not always include classroom observations. As a whole, Belmont faculty were skeptical of the use of peer evaluations for making administrative decisions.

Male faculty were more convinced that self-evaluation was being used to improve program/major and that it should be used in that regard. But male faculty

believed self-evaluation was used to make administrative decisions by a ratio of eight to one over female faculty.

Most faculty in the 31-40 age bracket thought that peer evaluation should be used to make administrative decisions--more faculty, in fact, than in any other age category.

The faculty who taught exclusively far outnumbered the faculty who also had some administrative duties in perceiving that student evaluation should be used to improve teaching, to improve program/major and to make administrative decisions. However, faculty as a whole were not as convinced that student evaluation ever should be used to make administrative decisions, even though they believed that student evaluation was being used to this end. Surprisingly, significantly more faculty who taught exclusively believed that administrative evaluation was and should be used to improve teaching than those who also had some administrative tasks assigned to them above their teaching loads. Far more faculty who taught exclusively believed that self-evaluation should be used to improve teaching than respondents who taught and had some additional administrative duties.

Religion, business and music faculty indicated that student evaluation wasn't being used as much to improve teaching as it should be; faculty in nursing and the sciences responded in the opposite extreme.

As for the use of student evaluation to improve program/major, humanities/education faculty, faculty in the sciences, and faculty in business thought that student evaluation should be used more to improve program/major than it was; nursing and religion faculty indicated student evaluation was being used to this end. And as for making administrative decisions, only the nursing faculty seemed to indicate that student evaluation should be used more for this purpose.

Administrative evaluation was perceived by nursing faculty as being used in improving teaching and in improving program/major, while humanities/education, business and music faculty believed administrative evaluation should be used more for these purposes.

In the school of nursing, more faculty believed peer evaluation was used to improve teaching and program/major, and only in business was there the perception that self-evaluation should be used more to improve teaching than it actually was being used.

Faculty with doctorates believed student evaluations were used to improve teaching, but faculty with master's degrees seemed to indicate that student evaluation needed to be done more at Belmont University to aid teaching improvement. Faculty with doctorates and master's degrees indicated that more use of student evaluation for improving program/major should be done than was currently practiced.

Faculty having the mid-range of years Belmont teaching experience (11-15) tended to believe self-evaluation was used to make administrative decisions.

For student evaluations, faculty endorsed, almost unanimously, the idea that nothing should be done with that information, yet they acknowledged that the information from student evaluations was being returned to instructors, and basically that was what should be done. Based on anecdotal responses, they also believed that faculty should be able to respond to that evaluation information. Faculty at Belmont were uncertain of the disposition of information derived from administrative, peer and self-evaluations.

Female faculty tended to believe that self-evaluation was being returned to faculty, and that was how it should be. Faculty in the middle-age group (41-50) perceived that nothing was done with the evaluation information gleaned from both administrative and self-evaluation.

Faculty who taught exclusively were more likely to believe that student evaluations were returned to instructors with no response expected than faculty who had additional administrative duties added to their teaching responsibilities, and faculty who taught exclusively believed that student evaluation information should be returned to faculty by a ratio of four to one over faculty who taught plus had some administrative duties. Faculty in the sciences indicated that student evaluation should be returned to instructors more often than they perceived that it was done. Tenured faculty favored not returning student evaluation to instructors more than non-tenured faculty. Faculty with doctorates were more likely to perceive that student evaluations were returned without the requirement of a response.

That administrative evaluation information be returned to faculty for response was not perceived to be done sufficiently by humanities/education faculty. Peer evaluation information, apparently, is more often not returned to faculty in the sciences and in business than in the other schools at Belmont, and only nursing faculty perceived that self-evaluation information was returned to instructors with no response requested.

The fewer the years experience at other institutions, the more faculty believed that nothing was done with self-evaluation information. Total years teaching experience did not seem to affect responses to this question.

Conclusions

The extant literature derived from extensive empirical research on student evaluation--its reliability, accuracy, etc. and less extensive research on administrative, peer and self-evaluation--revealed little information on faculty

attitudes toward evaluation, save that gleaned primarily from anecdotal responses. Some faculty favored student evaluation (Blank, 1978), and student evaluation was seen as potentially useful for improving teaching (Wheeler, 1972; Ory & Braskamp, 1984), yet some saw little potential for teaching improvement from student evaluation (Gross & Small, 1979). Most faculty questioned using student evaluation to make administrative decisions (Mahfous, 1979; Boyd & Shietinger, 1986), and were therefore guarded about favoring student evaluation (Millman, 1981; Aleamoni, 1987; Feldman, 1988).

Administrative evaluation caused some higher education faculty to have ambivalent reactions (Thomas, Scott & Beaird, 1976) because of its somewhat impressionistic (Genova, Madoff, Chin & Thomas, 1976), subjective nature (Hughes, 1975). Others viewed administrative evaluation as indirect, unreliable and of dubious value (Sheehan, 1975; Seldin, 1975), even potentially threatening or punitive (Whitman & Weiss, 1982). Yet some thought administrative evaluation could lead to improved teaching because administrative evaluators have experience and maturity (Hughes, 1975).

Peer evaluation was viewed as useful in potentially improving teaching (Aleamoni, 1987; Doyle, 1983; Cohen & McKeachie, 1980; Seldin, 1990), because it was perceived by faculty as less threatening than other types (DiNisi, Blencoe & Randolph, 1983), and because peers could be better judges of teaching than students or administrators (Batista, 1976). Yet some faculty felt peer evaluation's usefulness was a matter yet to be resolved (Whitman & Weiss, 1984) since some bias on the part of colleagues is possible (French-Lazovik, 1981).

Even though a few studies ranked self-evaluation high (Simpson, 1966; Young & Gwalamubisi, 1986) in terms of faculty acceptance and potential for formative purposes, most were skeptical because of its bias (Kulik & McKeachie, 1975; Bailey,

1981). Yet a few studies cited self-evaluation's potential for usefulness in improving teaching (Blackburn & Clark, 1975; Braskamp, Brandenburg & Ory, 1984).

This study was conducted in an attempt to determine the perceptions of faculty of one institution, Belmont University in Nashville, Tennessee, toward instructional evaluation as it was practiced at that institution, versus how the faculty believed evaluation should be practiced. Based on the analyses of the data and reasoned conclusions from the resulting findings, particularly the anecdotal responses, the following conclusions were reached:

1. Consistent with the literature, and supported primarily by anecdotal evidence, Belmont University faculty, when convinced that any of the four types of evaluation were for competency assessment, i.e. formative evaluation for diagnostic purposes, seemed supportive of the idea of evaluation.

2. In the actual execution of evaluation, the format of student evaluation, in particular, caused Belmont faculty to raise questions about the efficacy of student evaluation, which was again consistent with the literature, and again drawn primarily from anecdotal responses.

3. Although not supported statistically, there was a sense that most Belmont University faculty believed that honest diagnostic feedback from students concerning their teaching was of benefit both for instructional improvement and for the improvement of programs and majors--somewhat inconsistent with the literature, which indicated a lack of consistency in support of the benefits of student evaluation as a measure of teaching effectiveness. Again, entirely consistent with the literature, Belmont University faculty were guarded about favoring student evaluation, indicating its possible acceptance for use in improving teaching, but not for making administrative

decisions.

4. Consistent with the literature, and based as much on anecdotal evidence as statistical inference, Belmont faculty questioned the objectivity and reliability of student evaluation; contrary to the literature, Belmont faculty affirmed the fairness of student evaluation.

5. Belmont University faculty seemed suspicious of a type of "hidden agenda" in student evaluation in particular. The suspicion was that for all the indications that student evaluation was used formatively for improving teaching, it was in reality also used summatively, i.e. for decisions on salaries, tenure, etc. Such suspicions were consistent with research from the literature, and were gleaned from responses to open-ended questions.

6. Consistent with the literature, Belmont faculty confirmed the use of administrative evaluation for making administrative decisions, but departed sharply from the literature which indicates that faculty feel administrative evaluation is questionable for use in improving teaching. Belmont faculty said administrative evaluation should be used to improve teaching. Belmont faculty, again consistent with the literature, considered administrative evaluation political, even threatening or potentially punitive. Evidence to support this conclusion was both statistical and anecdotal.

7. Also consistent with the literature was the perception of Belmont University faculty, drawn more from anecdotal evidence, that administrative evaluation relied more on inferences than on observations.

8. According to statistical and anecdotal information, Belmont faculty, again consistent with the literature, supported the use of peer evaluation for improving teaching.

9. Where Belmont faculty clearly departed from research in the literature was in the perception that administrative evaluation was objective, a conclusion supported statistically. Most information from the literature indicates that faculty perceive administrative evaluation to be non-objective.

10. Based on reasoned conclusions from anecdotal evidence, Belmont faculty seemed to exhibit more faith and optimism for administrative evaluation than in other types, and they observed that administrative evaluation could lead to significant instructional improvement, yet they believed such improvement was unlikely, an observation consistent with most of what appears in the literature.

11. Belmont University faculty perceptions disagreed with much of the extant literature that peer evaluation was useful for formative purposes; statistically, the Belmont faculty ranked peer evaluation lowest in fairness, objectivity and reliability, and in effectiveness in improving teaching, improving programs/majors and in making administrative decisions.

12. Belmont faculty agreed with the literature which indicated that self-evaluation was too subjective. Statistically, Belmont faculty rated self-evaluation the least objective type, but said it was the most fair, most reliable, and most effective types for improving teaching, a clear deviation from the conclusions in the literature.

13. As for the impact that the gender of the respondents had on survey responses, based on anecdotal responses, female faculty members were more willing to give all evaluation methods the benefit of the doubt for effectiveness, fairness, objectivity and reliability than their male counterparts, who seemed more skeptical about all evaluation types, and more critical of evaluation in general.

14. The impact of age of respondents was negligible, in most venues of the survey; the only exception was student evaluations, wherein both younger and older

faculty seemed more leery of its effectiveness and of its fairness, objectivity and reliability than their counterparts who were in more median age categories; a conclusion supported more by anecdotal evidence than statistical inference.

15. Although not supported statistically, there was a sense that tenured faculty seemed less likely to question the effectiveness of all types of evaluation, and were more likely to believe that evaluation was objective, fair and reliable. Non-tenured faculty were particularly leery of student evaluation and its potential for misuse.

16. Considering all four types of evaluation done at Belmont University--had the faculty been asked to "assign grades" for student, administrative, peer and self-evaluations, a reasoned conclusion, based on statistics and anecdotal evidence, would be that the faculty would assign the grade of "A-" to administrative and self-evaluation; a "C-" to student evaluation, and a "D" to peer evaluation.

Discussion

Based on the findings and conclusions, it appeared that Belmont University faculty, if not positively disposed toward all four types of evaluation, could perceive that there was some merit in each type, if properly used. Faculty respondents seemed resigned to the fact that evaluation was not going away. In other words, evaluation was to be endured . Instructor attitudes such as these are not surprising, in that some of the literature has shown instructor attitudes ranging from apathy to mistrust regarding evaluation of instruction.

Belmont faculty appeared to have adopted an "us versus them" mentality toward evaluation, with the university administration being "them." While this almost adversarial approach may be somewhat typical, the Belmont faculty also seemed open to

changing this disposition, given the opportunity through feedback into how evaluation is done at the university. Also, Belmont faculty seemed to be sending a message that they be granted reciprocity, i.e. be allowed to evaluate their administrators.

Lack of responses to a number of questions seemed to be a clear indication that two types of evaluation--peer and self, were not practiced with regularity in most of the schools. The exceptions were nursing and business, where, apparently, peer and self-evaluations were executed on somewhat of a regular basis. To a limited extent, humanities/education faculty seemed used to these types of evaluation.

Judging from the percentages of responses to closed questions and anecdotal responses, Belmont faculty seemed to consider student evaluation banal at best. While the university faculty seemed to accept, philosophically, the idea of students evaluating their instructors, the actual practices, e.g. the timing of student evaluations, the phrasing of evaluation questions, and the capabilities of students as evaluators seemed less than desirable.

Belmont faculty seemed surprisingly trusting of administrative evaluation, and not for formative purposes like improving teaching or improving programs/majors, but for summative purposes, such as making administrative decisions on matters such as tenure, salaries, etc.--quite the opposite of what one might reasonably expect. And Belmont faculty were more skeptical of peer evaluation than administrative evaluation, another surprise.

Self evaluation, declared the fairest and most reliable type by Belmont faculty, was the over-whelming choice of evaluation types to be used, ideally, to improve teaching, improve program/major or to make administrative decisions. This was a surprising revelation, given that a great deal of the literature on self-evaluation seemed to question its efficacy because of its obvious bias. But faculty seemed eager to have

self-evaluations done and eager to respond to them after the initial self-evaluations were returned to faculty members themselves. This last procedure seemed, judging from anecdotal responses, to be a way for instructors to validate for themselves, the kind of teaching they were doing.

As for the impact of some of the demographic variables on the responses of Belmont University faculty, there were no real surprises. But some inferences may be drawn from the findings of the study regarding demographic variable influences. For example, the impact gender might have had on responses of faculty may have been somewhat neutralized by the religious atmosphere at Belmont. Because older faculty are generally more secure in their positions than younger colleagues, the older faculty may have been more willing to speak of their distrust of student evaluation.

Also, because of the religious foundation of Belmont, faculty who taught exclusively may have been less critical of the objectivity of administrative evaluation. In a rather surprising twist, faculty in religion were not the most tolerant of all forms of evaluation--nursing faculty were.

Also, because Belmont faculty had attained doctoral status or were ABD, they tended to look with disdain more, perhaps, on all types of evaluation than faculty with less education. Number of years teaching experience at Belmont University did not seem to cause faculty to be any more tolerant of evaluation than teaching experience anywhere else, surprisingly.

As another possible consequence of religious affiliation, Belmont University's faculty may have been less likely to be openly critical of evaluation in general. Perhaps that is why there was a more covert undercurrent of a perception that evaluation at Belmont was just so much paperwork, and an exercise of doing something not out of the administration's desire to improve the teaching as much as to make the university look

good.

Faculty at Belmont did appear to be appalled by some of the ways evaluation was done at the school, including the ideas that students could be coached into giving better evaluations or that student evaluations were ratings more of instructor personalities than teaching competencies. But underlying this attitude seemed to be a genuine desire on the part of faculty to improve from evaluation, if evaluations were fair, well-planned and well-executed. Belmont university faculty seemed almost anxious to wade in and change the evaluation procedures, particularly student evaluation.

While this study has provided data, findings and conclusions that indicated both congruence and discrepancies in the perceptions of faculty at Belmont University in Nashville, Tennessee about the evaluation of their teaching and about how evaluation is executed at Belmont versus how faculty think evaluation should be handled in the ideal, there is still evidence gleaned from this research that could have application in situations beyond this venue. Furthermore, this study has shown some differences in faculty responses which were a departure from what might have been expected from the literature search, and these differences, while due, perhaps, to the unique population of this study, can at least stimulate interest in further research on how evaluation is perceived and executed in higher education.

As a result, several recommendations can be made with regard to the way evaluation may be promoted and practiced.

1. If it can be agreed that instructional improvement is the central purpose of evaluation of faculty in higher education, then college and university administrators need to become more aware of the institutional needs of instructors and find methods by which they can help meet those needs.

2. Training/orientation of all evaluators from students to peers should be

undertaken by colleges and universities to ensure the most accurate appraisal of teaching and the cooperation and understanding of evaluators and evaluees.

3. Instructors need to be made aware of the intent of, purposes of, procedures in, disposition of, and ultimate usage of all types of evaluation at colleges and universities, with the intent being to increase awareness and understanding of the objectives of evaluation, which could lead to wider acceptance of evaluation procedures.

4. Instructors and administrators should strive for full utilization of information gleaned from all types of evaluation for the ultimate enhancement of teaching methods and procedures.

5. Universities and colleges should consider forming a committee assigned the task of developing, monitoring and even executing the fairest, most objective, reliable and consistent series of student, administrative, peer and self-evaluations within their capabilities.

6. Faculty should be given every oppportunity to provide input into the evaluation process, from framing of questions to responding to evaluation information--even to the arbitration of grievances between faculty and administration.

7. Evaluation procedures should be subjected, often, to review and scrutiny, the intent being to refine and update those procedures so that they can meet their goals.

Recommendations for Further Research

This study has attempted to determine the perceptions of higher education faculty of one institution, Belmont University, Nashville, Tennessee, regarding the current practice of four types of evaluation at the university versus the ideal practice of evaluation. What has been shown is that differences do exist between what faculty believe

is done in the evaluation of their teaching by students, administrators, peers and by themselves, and what faculty believe should be done in ideal evaluative circumstances. What has not been shown is why these perceptual differences exist. Because this study was limited to one institution, the results are not generalizable to other institutions. It is suggested that further research be done to determine if these perceptions are typical or atypical of higher education faculty elsewhere. Such research could be as follows:

1. A future study could replicate the examination of perceptions of faculty at other private, liberal arts colleges and universities concerning evaluation to check for agreement or disagreement in faculty perceptions of evaluation, and to provide longitudinal evidence from which to draw more exact conclusions about perceptions concerning evaluation.

2. A future study could also examine the perceptions toward evaluation of faculty at public colleges and universities in the United States to compare the findings with those at private colleges and universities.

3. A study involving more in-depth examination of higher education faculty perceptions through an interviewing process could be done that focuses on determining why faculty perceptions are the way they are.

4. Research should be done on what are perceived to be the ideal practices of evaluation by faculty, students and administrators, to see if gaps in perceptions do exist and what steps could be taken to ameloriate the differences. If gaps are closed, then faculty, administration and students should be better able to work together toward the best possible higher education instructional program.

5. Research needs to be done to see if students are "coached" in how to evaluate effectively. Do college and university students know what is expected of their efforts toward evaluation? Students may be unaware of how properly to use the evaluation

instrument, and what the outcomes of instructor evaluation are.

6. Research needs to be done on faculty understanding and involvement in the evaluation process. Do faculty have input, for example, into the extent, timing, instrumentation and results of evaluation of teaching in higher education? The involvement of faculty in the evaluation process from start to finish can be crucial if instructional improvement is the ultimate goal. Involvement can clear up misconceptions about intent and use of evaluation and lead to a more positive reception by faculty toward evaluation information.

LIST OF REFERENCES

REFERENCES

Abrami, P. C., Perry, R. P. & Leventhal, L. (1982). The relationship between student personality characteristics, teacher ratings, and student achievement. Journal of Educational Psychology, 74, 111-125.

Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), Handbook for Teacher Evaluation. Beverly Hills: Sage, Inc., 110-145.

Aleamoni, L. M. (1985). Peer evaluation of instructors and instruction. Instructional Evaluation, 8, 111-125.

Aleamoni, L. M. (1987). Techniques for Evaluating and Improving Instruction. San Francisco: Jossey-Bass, Inc.

Aleamoni, L. M. & Spenser, R. E. (1973). The Illinois course evaluation questionnaire: a description of its development and a report of some of its results. Educational and Psychological Measurement, 33(3), 669-684.

Alkin, M. C. (1969) Evaluation theory development. Educational Comment, 2, 2-7.

American Association for Higher Education (1982). Improving Instruction: Issues and Alternatives for Higher Education (AAHE Higher Education Report No. 4), Washington, D. C.

American Association of University Professors (1938). Guidelines for Association Operation (AAUP Report No. 4), Washington, D. C.

Arreola, R. A. (1987) Evaluating the dimensions of teaching. Instructional Evaluation, 8, 4-12.

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

Bailey, G. D. (1981). Teacher Self-Assessment: A Means for Improving Classroom Instruction. Washington, D. C. : National Education Association.

Barr, A. S. (1948). The measurement and prediction of teaching efficiency: A summary of investigations. Journal of Experimental Education, 16, 203-283.

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.

Bayley, D. H. (1967). Making college teaching a profession. Improving College and University Teaching, 15, 115-119.

Bendig, A. W. (1954). A factor analysis of student ratings of psychology instructors on the Purdue scale. Journal of Educational Psychology, 45, 385-393.

Benton, S. E. (1982). Rating College Teaching. Washington, D. C. : American Association for Higher Education.

Blackburn, R. T. & Clark, M. J. (1975) An assessment of faculty performance: Some correlates between administrator, colleague, student and self-ratings. Sociology of Education, 48, 242-256.

Blank, R. (1978). Faculty support for evaluation of teaching. Journal of Higher Education, 49(2), 164-176.

Bolton, D. L. (1973). Selection and Evaluation of Teachers. Berkeley: McCutchan Publishing.

Boyd, J. E. & Schietinger, E.F. (1976). Faculty Evaluation Procedures in Southern Colleges and Universities. Atlanta: Southern Regional Education Board. (ERIC Document Reproduction Services No. ED 121 153).

Boyer, E. L. (1987). College: The Undergraduate Experience in America. New York: Harper & Row.

Boyer, C.M.; Ewell, P.T.; Finney J.E. & Mingle, J. (1987), Assessment and outcomes measurement: A view from the states. AAHE Bulletin, 39, 8-12.

Brandenburg, D. C. & Ory, J. C. (1981). Considerations for an evaluation program of instructional quality. CEDR, 13, 8-12.

Braskamp, L. A., Brandenburg, D. C. & Ory, J. C. (1984). Evaluating Teaching Effectiveness. Beverly Hills: Sage, Inc..

Butler, J. R. & Tipton, R. M. (1976). Rating stability shown after feedback of prior ratings of instruction. Improving College and University Teaching, 24 (2), 111-115.

Castetter, W. B. (1981). The Personnel Function in Educational Administration. (3rd ed.) New York: McMillan.

Centra,. J. A. (1972). Two Studies on the Utility of Student Ratings for Instructional Improvement. Princeton: Educational Testing Service.

Centra. J. A. (1973). Self-ratings of college teachers: A comparison with student ratings. Journal of Educational Measurement, 10 (4), 287-295.

Centra, J. A. (1975). Colleagues as raters of classroom instruction. Journal of Higher Education, 46 (1), 327-337.

Centra, J. A. (1977). How Universities Evaluate Faculty Performance. Princeton: Educational Testing Service.

Centra, J.A. (1979). Determining Faculty Effectiveness. San Francisco: Jossey-Bass.

Centra, J. A. (1981). Research Report: Research Productivity and Teaching Effectiveness. Princeton: Educational Testing Service.

Clark, K. (1961). Studies of faculty evaluation. Studies of College Faculty. Berkeley: Center for the Study of Higher Education.

Coffman, W. E. (1954). Determining students' concepts of effective teaching from their ratings of instruction. Journal of Educational Psychology, 45, 385-393.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multi-section validity studies. Review of Educational Research, 6, 281-309.

Cohen, P.A. & McKeachie, W. J. (1980). The role of colleagues in the evaluation of college teaching. Improving College and University Teaching, 28, 147-153.

Cosgrove, D. S. (1959). Diagnostic rating of teacher performances. Journal of Educational Psychology, 50, 200-204.

Costin, F. Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity and usefulness. Review of Educational Research, 41 (5), 511-535.

Crannell, C.W. (1953). A preliminary attempt to identify the factors in student-instructor evaluation. Journal of Psychology, 36, 417-422.

Cronbach, L. J. (1963). Course improvement through evaluation. Teachers College Record, 64, 672-683.

Cross, K. P. (1986, March). The adventures of education in Wonderland: Keynote address at the Lilly Conference on College Teaching, Miami University, Miami, OH.

DeNisi, A. S., Randolph, W.A. & Blencoe, A. G. (1983). Potential problems with peer ratings. Academy of Management Journal, 26(3), 457-464.

Deshpende, A.S., Webb, S.C. & Marks, E. (1970). Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. American Educational Research Journal, 7, 289-305.

DeWolf, W.A. (1974). Student ratings of instruction in post secondary institutions: A comprehensive examination of research reported since 1968. Seattle: University of Washington Educational Assessment Center.

Doyle, K. O. Jr.. (1975). Evaluating Teaching. Lexington: D.C. Heath and Company.

Doyle, K. O. Jr. (1983). Evaluating Teaching (2nd ed.). Lexington: D.C. Heath and Company.

Doyle, K. O. Jr. (1985). Evaluating Teaching (rev. ed.). Lexington: Lexington Books.

Doyle, K. O. Jr., & Crichton, L. I. (1978). Student, peer and self-evaluation of college instruction. Journal of Educational Psychology, 70(5), 815-826.

Dressel, P. L. & Marcus, D. (1982). On Teaching and Learning in College. San Francisco: Jossey-Bass.

Drucker, A. J. & Remmers, H. H. (1951). Do Alumni and students differ in their attitudes toward instructors? In H. H. Remmers (Ed.), Studies in Higher Education: Studies in College and University Staff Evaluation. Lafayette: Purdue University Press.

Eble, K. (1971). Th Recognition and Evaluation of Teaching. Salt Lake City: Project to Improve College Teaching.

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

Elliott, D. H. (1950). Characteristics and relationships of various criteria of college and university teaching. Purdue Studies in Higher Education, 70, 5-61.

Erickson, S. C. (1984). The Essence of Good Teaching. San Francisco: Jossey-Bass.

Feldman, K. A. (1976). The superior teacher from the student view. Research in Higher Education, 5, 243-288.

Feldman, K.A. (1977). Consistency and variability among college students in rating their teachers and courses. Research in Higher Education, 6, 223-274.

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. Research in Higher Education, 10, 149-172.

Ferren, A. & Geller, W. (1983). Classroom consultants: Colleagues helping colleagues. Improving College Teaching, 31(2), 82-86.

Fitzgerald, M. J. (1979). A comparative study of faculty evaluation by peers and students within a community setting. Dissertation Abstracts International, 39, 3999-099A.

French-Lazovich, G. (1981). Peer review: Documenting evidence in the evaluation of teaching. In J. Millman (Ed.) Handbook of Teacher Evaluation. Beverly Hills: Sage, Inc., 73-89.

Frey, P.W. (1978). A two dimensional analysis of student ratings of instruction. Research in Higher Education, 9, 69-71.

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

Geis, G.L. (1984). The context of evaluation. In P. Seldin (Ed.), Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass.

Genova, W. J., Madoff, M. K., Chin, R. & Thomas, G.B. (1976). Mutual Benefit Evaluation of Faculty and Administration in Higher Education. Cambridge: Ballinger Publishing Co.

Gilmore, G. M., Kane, M.T., & Naccarato, R.W. (1978). The generalizability of student ratings of instruction: Estimates of teacher and course components. Journal of Educational Measurement, 15, 1-13.

Goode, D.M. (Ed). (1968). Encouraging and appraising teaching quality. Improving College and University Teaching, 16: 191-209.

Goodlad, J. I. (1983). A Place Called School: Prospects for the Future. St. Louis: McGraw-Hill.

Goodwin, H. I. & Smith, E.R. (1983). Faculty and administrator evaluation: Constructing the instruments. Wheeling: West Virginia University Press.

Gross, R.F. & Small, C.E. (1979). Faculty growth contracts. Faculty Development and Evaluation in Higher Education, 5, 9-14.

Gustad, J. W. (1961). Policies and practices in faculty evaluation. Educational Record, 42, 194-211.

Gustad, J.W. (1967). Evaluation of teaching performance: Issues and possibilities. In C.B.T. Lee (Ed.) Improving College Teaching. Washington, D.C.: American Council on Education.

Guthrie, E. R. (1954). The Evaluation of Teaching: A Progress Report. Seattle: University of Washington.

Hildebrand, M. & Wilson, R.C. (1970). Effective University Teaching. Davis: University of California.

Hildebrand, M., Wilson, R.C., & Dienst, E.R. (1971). Evaluating University Teaching. Berkeley: University of California Press.

Holmes, D. S. (1971) The teaching assignment blank: A form for the student assessment of college instructors. The Journal of Experimental Education, 39, 34-38.

Holzbach, R. L. (1978). Rater bias in performance ratings: Supervisor, self-, and peer ratings. Journal of Applied Psychology, 63(5), 579-588.

Hughes, C. R. (1975). Faculty attitudes toward evaluation and teaching improvement. Improving College and University Teaching Yearbook, 74-78.

Isaacson, R. L.; McKeachie, W. J.; Milholland, J.E.; Lin, Y.G.; Hofeller, M.; Bearwaldt, J.W. & Zinn, K. L. (1964). Dimensions of student evaluations in teaching. Journal of Educational Psychology, 55(6), 344-351.

Kerlinger, F. (1971). Student evaluation of university professors. School and Society, 99, 353-356.

Kirschling, W.R. (Ed.). (1978). Evaluating Faculty Performance and Vitality. San Francisco: Jossey-Bass, Inc.

Kronk. A. K. & Shipka, T.A. (1980). Evaluation of Faculty in Higher Education. Washington, D.C. : National Education Association.

Kulik, J.A. & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), Review of Research in Education, 3, Itaska: F.E. Peacock Publishers, Inc.

Ladd, E.C. Jr. & Lipset, S.M. (1973) Professors, Unions and Higher Education. Berkeley: Carnegie Foundation for the Advancement of Teaching.

Levinson-Rose, J. & Menges, R. J. Improving college teaching: A critical review of research. Review of Educational Research, 51(3), 403-434.

Licata, C. (1986). Post Tenure Faculty Evaluations: Threat or Opportunity? Washington, D.C. : Association for Study of Higher Education.

Mahfous, A. F. M. (1979). Faculty evaluation: An analysis of the attitudes of faculty members and administrators toward student evaluation of teaching. Dissertation Abstracts International, 39, 39010A.

Marsh, J.E.; Burgess, G. & Smith, P.N. (1956). Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 47, 79-88.

Marsh, H.W. (1980). Research on students' evaluation of teaching effectiveness. Instructional Evaluation, 4, 5-13.

Marsh, H.W., Overall, J.U. & Kessler, T. (1982). Student evaluations of teaching: An update. American Association for Higher Education Bulletin, 35(4), 9-13.

Marsh, H.W. & Hocevar, D. (1984). The factorial invariance of student evaluation of college teaching. American Educational Research Journal, 21, 341-366.

Marsh, H.W. (1987). Student evaluation in higher education. International Journal of Educational Research, 11(4), 257-378.

Marsh, H.W. (1989). Students' evaluations: History, dimensionality and directions for future research. International Journal of Educational Research, 13 (4), 248-379.

Maslow, A.H. & Zimmerman, W. (1956). College teaching ability, scholarly activity personality. Journal of Educational Psychology, 47, 185-187.

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

McKeachie, W. J. & Solomon, D. (1958). Student ratings of instructors: A validity study. Journal of Educational Research, 51, 379-382.

McMartin, J.A. & Rich, H.E. (1979). Faculty opinion at California State University, Northridge, toward student evaluation of teaching effectiveness. Unpublished research paper.

Meyer, H. H. (1980). Self-appraisal of job performance. Personnel Psychology, 33(2), 291-295.

Miller, M. T. (1971). Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 62, 235-239.

Millman, J. (1981). Handbook of Teacher Evaluation. Beverly Hills: Sage, Inc.

Murray, H.G. (1980, March). Student evaluation of university teaching: Uses and abuses. Colluquium presented at the University of British Columbia, Vancouver.

Naftulin, D. H., Ware, J.E. & Donnelly, F.A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. Journal of Medical Education, 48, 630-635.

National Center for Research in Post Secondary Teaching and Learning (1987). Performance Appraisal: Its Implications in Higher Education (NCRIPTL Publication No. PAI 87-1195). Washington, D. C.: Author.

Neely, J. (1973). A Review of Research on Teacher Evaluation. Columbus: ERIC/SMEAC.

Office of Educational Research and Improvement (1986). Assessment in American Higher Education: Issues and Contexts. Washington, D.C.: U.S. Government Printing Office.

Ormrod, J.E. (1986). Prediction of faculty dissatisfaction with an annual performance evaluation. Journal of Higher Education, 37(3), 13-17.

Ory, J.C. & Braskamp, L.A. (1981). Faculty perceptions of the quality and usefulness of three types of evaluative information. Research in Higher Education, 15(3),271-278.

Ory, J.C. (1991). Changes in evaluating teaching in higher education. Theory into Practice, 30(1), 30-36.

Ory, J.E. & Parker, S. (1989). A survey of assessment activities at large research universities. Research in Higher Education, 30, 373-383.

Overall, J. U. & Marsh, H.W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 71, 856-865.

Overall, J.U. & Marsh, H.W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. Journal of Educational Psychology, 72, 321-325.

Overholt, G.E., & Urbana, W.J. (1976). Accountability in American Education: A Critique. Princeton: Princeton Book Company.

Ozmon, H. (1967). Publications and teaching. Improving College and University Teaching, 15, 106-107.

Page, C.F. (1975). Student Evaluation of Teaching: The American Experience. London: Society for Research into Higher Education.

Payne, D. A. & Hobbs, A.M. (1979). The effect of college course evalution feedback on instructor perceptions of instructional climate and effectiveness. Higher Education, 8, 525-533.

Phillipi, H.A. (1979). A Dean's Response to a Review of Procedures and Policies Governing Appointment, Promotion and Tenure in New England Institutions of Higher Education.Boston: Massachusetts Department of Post Secondary Education. (ERIC Document Reproduction Service No. ED 174 132).

Pollack, D.J. (1976). The Development and Testing of a Criterion Referenced Evaluation System for Faculty and Administrators in Technical Institutions/Community Colleges. Raleigh, N.C.: State Department of Public Instruction. (ERIC Document Reproduction Service No. ED 133 578).

Remmers, H.H. (1929). Student ratings of college teaching: A reply. School and Society, 30, 232-234.

Rodin, M. & Rodin, B. (1972). Student evaluations of teachers. Science, 177, 1164-1166.

Root, A.R. (1931). Student ratings of teachers. Journal of Higher Education, 2, 311-315.

Seldin, P. (1973). A study to determine the current policies and practices used in liberal arts colleges to evaluate classroom teaching performance of members of the faculty. Fordham University; part published: How deans evaluate teachers. Change, 6 (1974), 48-49.

Seldin, P. (1975). Evaluating professors in the ivory tower. Improving College Teaching, 5, 20-25.

Seldin, P. (1978). How Colleges Evaluate Professors: Current Policies and Practices in Evaluating Classroom Teaching Performance in Liberal Arts Colleges. Cronton-on-Hudson: Blythe-Pennington.

Seldin, P. (1980). Successful Faculty Evaluation Programs. Crugers: Coventry Press.

Seldin, P. (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass.

Seldin, P. (1990). How Administrators Can Improve Teaching: Moving from Talk to Action in Higher Education. San Francisco: Jossey-Bass.

Sheehan, D.S. (1975). On the invalidity of student ratings for administrative personnel decisions. Journal of Higher Education, 46(6), 687-699.

Simpson, R. H. (1966). Teacher Self-Evaluation. London: The Macmillan Company.

Skipper, C.E. (1975). A technique for teacher self-evaluation. Improving College Teaching, 3, 11-13.

Sorey, K.E. (1968). A study of the distinguishing personality characteristics of college faculty who are superior in regard to the teaching function. Dissertation Abstracts, 28 (12-A), 4916.

Stufflebeam. D.L. (1969). Evaluation as enlightenment for decision making. In W.H. Beatty (Ed.), Improving Educational Assessment and an Inventory for Measures of Affective Behavior. Washington, D.C.: National Education Association.

Stuit, D. B. & Ebel, R.L. (1951). General education at the State University of Iowa. In W. H. Stickler (Ed.), Organization and Administration of General Education, (pp. 212- 243). Dubuque: Wm. C. Brown & Co.

Stuit, D.B. & Ebel, R.L. (1952). Instructor rating at a large state university. College and University, 27, 247-254.

Thorne, G.L., Scott, C.S. & Beaird, J.H. (1976). Assessing Faculty Performance. Monmouth: Oregon State System of Higher Education, Teaching Research Division.

Tipton, R.M. (1976). Rating stability shown after feedback of prior ratings of instruction. Improving College and University Teaching, 24(2),111-115.

Treffinger, D. J. & Feldhusen, J.F. (1970). Predicting students' ratings of instruction. Proceedings of the 78th Annual Convention of the American Psychological Association, 5, 621-622.

Tyler, R.W. (1950). The function of measurement in improving instruction. In E. F. Lindquist (Ed.), Educational Measurement (pp. 239-252). Washington, D.C.: American Council on Education.

Webb, W.B. & Nolan, C.Y. (1955). Student, supervisor and self-ratings of instructional proficiency. Journal of Educational Psychology, 46, 42-46.

Weimer, M.G. (1987). Translating evaluation results into teaching improvements. AAHE Bulletin, 5, 8-11.

Wheeler, B. (1972). Students' evaluations of university instructors: The application of some instruments. Teaching and Teacher Education: An International Journal of Research and Studies, 1, 123-138.

Whitman, N. & Weiss, E. (1982). Faculty Evaluation: The Use of Explicit Criteria for Promotion, Retention and Tenure. Washington, D.C.: American Association for Higher Education.

Wilson, L. (1942). The Academic Man. London: Oxford University Press.

Wilson, R.C. & Gaff, J.G. (1975). College Professors and Their Impact on Students. New York: John Wiley & Sons.

Young, A.J. & Gwalamubisi, Y. (1986). Perceptions about current and ideal methods and purposes of faculty evaluation. Community College Review, 13(4), 27-32.

APPENDICES

Appendix A

COVER LETTERS

OFFICE OF
ACADEMIC AFFAIRS

BELMONT
UNIVERSITY

MEMO TO: Belmont Faculty

FROM: Jerry L. Warren,
Acting Vice-President for Academic Affairs

RE: Survey on teacher evaluation

DATE: September 27, 1991

Enclosed is a survey form from Professor Sheridan Barker of Carson-Newman College. Mr. Barker is a doctoral student in educational leadership at the University of Tennessee in Knoxville. Mr. Barker is using the Belmont faculty in research for his dissertation, exploring the issue of teacher evaluation. I encourage your timely response to this survey. Mr. Barker has offered to share the results of his research with us. This data will be of assistance to your ad hoc committee on evaluation as they continue to study this issue.

Dr. Robert Byrd, representing the Faculty Affairs Committee, and Dr. George Sims, representing the ad hoc committee on evaluation, have both reviewed the survey , and they agree that this is a valid tool for our participation.

Thank you for your assistance in this project.

/jb

1900 Belmont Boulevard Phone 615-385-6400
Nashville, TN 37212-3757 Fax 615-385-6446

March 20, 1992

Faculty of Belmont University
Nashville, TN

Dear Faculty Member:

I am doing research for a dissertation in Educational Leaderhsip which will explore the perceptions of higher education teaching faculty toward evaluation. My research calls for a survey of the entire population of teaching faculty at Belmont as a way of seeing how to tap into such perceptions in higher education.

Attached is a questionnaire which asks in a straight forward manner for your perceptions of evaluation at Belmont, and what you believe evaluation should be, generally. Please fill out the questionnaire and return it in the envelope provided. The multiple choice questions ask for the responses which most closely reflect your view. The open-ended questions ask for information in your own words that amplify and clarify your multiple choice responses. YOUR INPUT IS VITAL TO THIS RESEARCH. Participation is, of course, voluntary. Filling out the questionnaire should take you no more than fifteen or twenty minutes.

The questionnaire contains no questions or markings which identify you as a respondent. The data obtained from the survey will be kept in a strong box and only the researcher and the chair of his dissertation committee, Dr. Norma Mertz of the Department of Educational Leadership, the University of Tennessee, Knoxville will see the data. The results will be analyzed and reported only in aggregate form to assure anonymity. The aggregated results will be made available to the Belmont ad hoc committee on evaluation, and could help to improve the evaluation procedures currently in use at Belmont.

If you have any questions about this project, please contact me at Box 71869, Carson-Newman College, Jefferson City, TN 37760; business phone: 471-3267; home phone: 581-7731.

Your return of the questionnaire will constitute your informed consent to participate in the study. PLEASE RETURN THE COMPLETED FORMS IN THE POSTAGE-PAID, PRE-ADDRESSED ENVELOPE BY April 3, 1992. THANK YOU FOR YOUR TIME AND COOPERATION!

Sincerely,

Sheridan Barker

Appendix B

SURVEY FORM

SURVEY OF TEACHING FACULTY PERCEPTIONS OF EVALUATION
BELMONT UNIVERSITY

ALL SURVEY INFORMATION WILL BE KEPT STRICTLY CONFIDENTIAL.
Please do not sign your name. Thank you for your time.

I. DEMOGRAPHIC INFORMATION

Circle the number of your response except where otherwise indicated, or fill in the blank

1. **Sex**
   1.1 Female
   1.2 Male

2. **Age** (Check the appropriate category)

2.1___21-25	2.4___36-40	2.7___51-55
2.2___26-30	2.5___41-45	2.8___56-60
2.3___31-35	2.6___46-50	2.9___61-65
		2.10__66-70

3. **Academic Rank**
   3.1 Instructor
   3.2 Assistant Professor
   3.3 Associate Professor
   3.4 Full Professor

4. **Tenured**
   4.1___Yes
   4.2___No

5. **Status**
   5.1 Full-time teaching faculty
   5.2 Teach/have other administrative duties
   5.3 Other (Please specify):_____

6. **School**
   6.1 Nursing
   6.2 Religion
   6.3 Humanities/Education
   6.4 Sciences
   6.5 Business
   6.6 Music

7. **Highest Degree Earned**
   7.1 Doctorate
   7.2 Master's
   7.3 Bachelor's
   7.4 Other (Please Specify):_____

8. **Years Teaching Experience in Higher Education:**

   8.1 At Belmont (Include current year) ___
   8.2 At Other Institutions___
   8.3 Total___

II. PERCEPTIONS OF TYPES OF FACULTY EVALUATION CURRENTLY IN USE AT BELMONT UNIVERSITY:

DIRECTIONS: Each of the questions in this section relates to the four types of teaching evaluation currently in use at Belmont. Please answer the questions asked by checking your responses and writing in any comments you wish to make under "COMMENTS," using extra sheets if needed

9. How do you perceive each type of evaluation currently in use at Belmont? Place an "X" in boxes that apply under **Fair, Objective, and Reliable** (One response per category).

TYPE OF EVALUATION	FAIR				OBJECTIVE				RELIABLE			
	VERY FAIR	FAIR	UNFAIR	VERY UNFAIR	VERY OBJECTIVE	OBJECTIVE	NON-OBJECTIVE	VERY NON-OBJECTIVE	VERY RELIABLE	RELIABLE	UNRELIABLE	VERY UNRELIABLE
9.1 Student												
9.2 Administrator												
9.3 Peer												
9.4 Self												

COMMENTS: _____

_____

_____

_____

10. Indicate your perceptions of how **satisfied** you are with each type of evaluation in use at Belmont is by placing an "X" in the appropriate box below:

TYPE OF EVALUATION	VERY SATISIED	SATISFIED	UNSATISFIED	VERY UNSATISFIED
10.1 Student				
10.2 Administrator				
10.3 Peer				
10.4 Self				

COMMENTS: _____

_____

11. Please indicate your perceptions about whether the four types of evaluation done at Belmont **should be done** at all by placing an "X" on the appropriate line below right.

TYPE	SHOULD BE DONE	SHOULD NOT BE DONE
11.1 Student	___	___
11.2 Administrator	___	___
11.3 Peer	___	___
11.4 Self	___	___

COMMENTS: _____
_____
_____
_____

12. Indicate your perceptions of how **satisfied** you are. OVERALL, with the evaluation procedures currently in use at Belmont by placing an "X" on the appropriate line below.

WITH BELMONT'S EVALUATION PROCEDURES, I AM, OVERALL:

___12.1 Very Satisfied
___12.2 Satisfied
___12.3 Unsatisfied
___12.4 Very Unsatisfied

COMMENTS_____
_____
_____
_____

III. PERCEPTIONS OF HOW FACULTY EVALUATION IS CONDUCTED AT BELMONT VS. HOW FACULTY EVALUATION SHOULD BE CONDUCTED

13. What best describes your perceptions about how each type of evaluation **is used** at Belmont and how you think each type of evaluation **should be used.** Use the left-hand column to indicate your perceptions about how evaluation **IS USED** at Belmont. Use the right-hand column to indicate your perceptions about how this type of evaluation **SHOULD BE USED**. Place an "X" on all lines that apply.

13.1 STUDENT EVALUATION

IS USED		SHOULD BE USED
___	TO IMPROVE TEACHING	___
___	TO IMPROVE PROGRAM, MAJOR	___
___	TO MAKE ADMINISTRATIVE DECISIONS, e.g. TENURE, PROMOTIONS, ETC.	___
___	DON'T KNOW	___

13.2 ADMINISTRATOR EVALUATION

IS USED		SHOULD BE USED
___	TO IMPROVE TEACHING	___
___	TO IMPROVE PROGRAM, MAJOR	___
___	TO MAKE ADMINISTRATIVE DECISIONS, e.g. TENURE, PROMOTIONS, ETC.	___
___	DON'T KNOW	___

13.3 PEER EVALUATION

IS USED		SHOULD BE USED
___	TO IMPROVE TEACHING	___
___	TO IMPROVE PROGRAM, MAJOR	___
___	TO MAKE ADMINISTRATIVE DECISIONS, e.g. TENURE, PROMOTIONS, ETC.	___
___	DON'T KNOW	___

13.4 SELF-EVALUATION

IS USED		SHOULD BE USED
___	TO IMPROVE TEACHING	___
___	TO IMPROVE PROGRAM, MAJOR	___
___	TO MAKE ADMINISTRATIVE DECISIONS, e.g. TENURE, PROMOTIONS, ETC.	___
___	DON'T KNOW	___

COMMENTS:

14. What is done with information collected from the 4 types of evaluation done at Belmont, and what **should be done** with that information ideally? Place an "X" on all lines that apply.

14.1 STUDENT EVALUATIONS

IS DONE		SHOULD BE DONE
___	RETURNED TO INSTRUCTOR; NO RESPONSE REQUESTED	___
___	RETURNED TO INSTRUCTOR FOR RESPONSE-E.G. COMMENTS, SETTING GOALS, WAYS TO IMPROVE, ETC.	___
___	NOT RETURNED TO INSTRUCTOR	___
___	NOTHING	___

14.2 ADMINISTRATIVE EVALUATIONS

IS DONE		SHOULD BE DONE
___	RETURNED TO INSTRUCTOR; NO RESPONSE REQUESTED	___
___	RETURNED TO INSTRUCTOR FOR RESPONSE-E.G. COMMENTS, SETTING GOALS, WAYS TO IMPROVE, ETC.	___
___	NOT RETURNED TO INSTRUCTOR	___
___	NOTHING	___

14.3 PEER EVALUATIONS

IS DONE		SHOULD BE DONE
___	RETURNED TO INSTRUCTOR; NO RESPONSE REQUESTED	___
___	RETURNED TO INSTRUCTOR FOR RESPONSE-E.G. COMMENTS, SETTING GOALS, WAYS TO IMPROVE, ETC.	___
___	NOT RETURNED TO INSTRUCTOR	___
___	NOTHING	___

14.4 SELF-EVALUATIONS

IS DONE		SHOULD BE DONE
___	RETURNED TO INSTRUCTOR; NO RESPONSE REQUESTED	___
___	RETURNED TO INSTRUCTOR FOR RESPONSE-E.G. COMMENTS, SETTING GOALS, WAYS TO IMPROVE, ETC.	___
___	NOT RETURNED TO INSTRUCTOR	___
___	NOTHING	___

COMMENTS_____
_____
_____

15. How effective is each of the four types of evaluation in use at Belmont in accomplishing the purposes listed below? Place an "X" in the boxes that apply.

TO IMPROVE TEACHING

TYPE OF EVALUATION	VERY EFFECTIVE	EFFECTIVE	INEFFECTIVE	VERY INEFFECTIVE
15.1 Student				
15.2 Administrator				
15.3 Peer				
15.4 Self				

TO IMPROVE PROGRAM, MAJOR

	VERY EFFECTIVE	EFFECTIVE	INEFFECTIVE	VERY INEFFECTIVE
15.5 Student				
15.6 Administrator				
15.7 Peer				
15.8 Self				

TO MAKE ADMINISTRATIVE DECISIONS

	VERY EFFECTIVE	EFFECTIVE	INEFFECTIVE	VERY INEFFECTIVE
15.9 Student				
15.10 Administrator				
15.11 Peer				
15.12 Self				

COMMENTS_____

_____

Appendix C

FOLLOW-UP POST CARD

April 15, 1992

Dear Belmont Faculty Member,

Recently, I sent you a survey form in an effort to learn your perceptions regarding the present and ideal practice of evaluation at Belmont University.

I realize how easy it is to put surveys aside, especially if you feel your response will not be missed, but believe me, YOUR response is vitally important to the success of this project.

Please take just a few minutes to complete the form now and mail it in the stamped envelope provided.

Thank You Very Much,

Sheridan C. Barker

Appendix D

TABLES

Table 1: Instructor Perceptions of Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Type	No./%	No./%	No./%	No./%	No./%
			Fairness of Evaluation		
	Very Fair	Fair	Unfair	Very Unfair	NR
Student	3(3.37)	53(59.55)	25(29.08)	3(3.37)	5(5.62)
Admin.	9(10.11)	52(58.43)	12(13.48)	2(2.25)	14(15.73)
Peer	6(6.74)	34(38.20)	6(6.74)	1(1.12)	42(47.19)
Self	9(10.11)	55(61.80)	8(8.99)	4(4.49)	13(14.61)
			Objectivity of Evaluation		
	Very Objective	Objective	Non Objective	Very Non Objective	NR
Student	2(2.25)	30(33.71)	43(48.31)	9(10.11)	5(5.62)
Admin.	5(5.62)	43(48.31)	23(25.84)	4(4.49)	14(15.73)
Peer	4(4.49)	24(26.97)	18(20.22)	1(1.12)	42(47.19)
Self	4(4.49)	38(42.70)	23(25.84)	6(6.74)	18(20.22)
			Reliability of Evaluation		
	Very Reliable	Reliable	Unreliable	Very Unreliable	NR
Student	0(0.00)	33(37.08)	39(43.82)	11(12.36)	6(6.74)
Admin.	5(5.62)	47(52.81)	18(20.22)	5(5.62)	14(15.73)
Peer	4(4.49)	31(34.83)	8(8.99)	5(5.62)	41(46.07)
Self	5(5.62)	52(58.43)	12(13.48)	3(3.37)	17(19.10)

Admin.=Administrative; NR=No Response

Table 2. Instructor Perceptions of Satisfaction with Belmont University Evaluation System

Type	No./%	No./%	No./%	No./%	No./%
Satisfaction Level of Belmont Faculty with the Four Types of Evaluation Practiced at Belmont					
	Very Satisfied	Satisfied	Unsatisfied	Very Unsatisfied	NR
Student	1(1.12)	33(37.08)	34(38.20)	16(17.98)	5(5.62)
Admin.	5(5.62)	43(48.31)	21(23.60)	10(11.24)	10(11.24)
Peer	3(3.37)	29(32.58)	11(12.46)	10(11.24)	36(40.45)
Self	9(10.11)	45(50.56)	15(16.85)	5(5.62)	15(16.85)

Overall Satisfaction Rating of Belmont Faculty on Evaluation Procedures (Percent)

Very Satisfied	Satisfied	Unsatisfied	Very Unsatisfied	NR
1(1.12)	41(46.07)	35(39.23)	9(10.11)	3(3.37)

Admin.=Administrator; NR=No Response

Table 3. Instructor Perceptions About Whether the Four Types of Evaluation Should/Should Not Be Done at Belmont University

Type	No./%	No./%	No./%
	Should Be Done	Should Not Be Done	NR
Student	74(83.15)	10(11.24)	5(5.62)
Administrator	75(84.27)	8(8.99)	6(6.74)
Peer	62(69.66)	19(21.35)	8(8.99)
Self	72(80.90)	9(10.11)	8(8.99)

Table 4. Instructor Perceptions Regarding How the Four Evaluation Types Are/Should Be Used

Use	No./%	No./%	No./%	No./%
		Student Evaluations		
	Is Used	Should Be Used	Is/Should Be Used	NR
Improve Teach.	6(6.74)	41(46.07)	35(39.33)	7(7.87)
Improve Prog./Maj.	4(4.49)	44(49.44)	19(21.35)	22(24.72)
Make Adm. Dec.	28(31.46)	11(12.36)	27(30.34)	23(25.84)
DN	- - - -	- - - -	- - - -	- - - -
		Administrative Evaluations		
	Is Used	Should Be Used	Is/Should Be Used	NR
Improve Teach.	3(3.37)	38(42.70)	19(21.35)	29(32.58)
Improve Prog./Maj.	3(3.37)	42(47.19)	13(14.61)	31(34.83)
Make Adm. Dec.	16(17.98)	10(11.24)	32(35.96)	31(34.83)
DN	17(19.10)	0(0.60)	3(3.37)	69(77.53)
		Peer Evaluations		
	Is Used	Should Be Used	Is/Should Be Used	NR
Improve Teach.	2(2.25)	44(49.44)	7(7.87)	36(40.45)
Improve Prog./Maj.	0(.60)	22(24.72)	1(1.12)	30(33.71)
Make Adm. Dec.	6(6.74)	19(21.35)	8(8.99)	56(62.92)
DN	27(30.34)	2(2.25)	5(5.62)	55(61.80)
		Self-Evaluation		
	Is Used	Should Be Used	Is/Should Be Used	NR
Improve Teach.	3(3.37)	7(7.87)	19(21.35)	36(40.65)
Improve Prog./Maj.	7(7.87)	26(29.21)	22(24.72)	34(38.20)
Make Adm. Dec.	9(10.11)	12(13.48)	18(20.22)	50(56.18)
DN	15(16.85)	0(.60)	2(2.25)	72(80.90)

Teach.=Teaching; Prog./Maj.=Program/Major; Adm. Dec.=Administrative Decisions; DN= Don't Know; NR=No Response; --= Insufficient responses in the DN category of Student Evaluations caused no data to appear on computer printout.

Table 5. Instructor Perceptions of What Is Done/Should Be Done with Evaluation Information

Disposition	No./%	No./%	No./%	No./%
		Student Evaluations		
	Is Done	Should Be Done	Is & Should Be Done	NR
Return to Instructor/NRR	51(57.30)	4(4.49)	25(28.09)	9(10.11)
Return to Instructor/Res.	2(2.25)	34(38.20)	4(4.49)	49(55.06)
Not Returned	5(5.62)	2(2.25)	0	82(92.13)
Nothing	0	4(4.49)	85.(95.51)	0
		Administrator Evaluations		
	Is Done	Should Be Done	Is & Should Be Done	NR
Return to Instructor/NRR	18(20.22)	5(5.62)	11(12.36)	55(61.80)
Return to Instructor/Res.	2(2.25)	32(35.96)	10(11.24)	45(50.56)
Not Returned	16(17.98)	2(2.25)	2(2.25)	69(77.53)
Nothing	11(12.36)	1(1.12)	1(1.12)	76(85.39)
		Peer Evaluations		
	Is Done	Should Be Done	Is & Should Be Done	NR
Return to Instructor/NRR	5(5.62)	10(11.24)	5(5.62)	69(77.53)
Return to Instructor/Res.	1(1.12)	23(25.84)	0	65(73.03)
Not Returned	11(12.36)	1(1.12)	1(1.12)	76(85.39)
Nothing	27(30.34)	5(5.62)	0	57(64.04)
		Self-Evaluations		
	Is Done	Should Be Done	Is & Should Be Done	NR
Return to Instructor/NRR	14(15.73)	2(2.25)	6(6.74)	67(75.28)
Return to Instructor/Res.	1(1.12)	23(25.84)	9(10.11)	56(62.92)
Not Returned	13(14.61)	1(1.12)	4(4.49)	71(79.78)
Nothing	13(14.61)	0(.60)	4(4.49)	36(40.45)

NR=No Response; NRR=No Response Required; Res.=Respond

Table 6. Instructor Perceptions of Effectiveness of Four Types of Evaluation at Belmont University in Improving Teaching, Improving Program/Major and Making Administrative Decisions

Type	No./%	No./%	No./%	No./%	No./%
Perceived Effectiveness of Four Types of Evaluation at Belmont in Improving Teaching					
	Very Effective	Effective	Ineffective	Very Ineffective	NR
Student	5(5.62)	35(39.33)	36(40.45)	8(8.99)	5(5.64)
Admin.	1(1.12)	34(38.20)	25(28.09)	15(16.85)	14(15.73)
Peer	4(4.49)	14(15.73)	18(20.22)	17(19.10)	36(40.45)
Self	7(7.87)	41(46.07)	21(23.60)	7(7.87)	13(14.61)
Perceived Effectiveness of Four Types of Evaluation at Belmont in Improving Program/Major					
	Very Effective	Effective	Ineffective	Very Ineffective	NR
Student	2(2.25)	27(30.34)	40(44.94)	14(15.73)	6(6.74)
Admin.	1(1.12)	30(33.71)	24(26.97)	21(23.60)	13(39.33)
Peer	1(1.12)	13(14.61)	21(23.60)	19(21.35)	35(39.33)
Self	4(4.49)	31(34.83)	26(29.21)	14(15.73)	14(15.73)
Perceived Effectiveness of Four Types of Evaluation at Belmont in Making Administrative Decisions					
	Very Effective	Effective	Ineffective	Very Ineffective	NR
Student	0(0.00)	28(31.46)	32(35.96)	16(17.98)	13(14.61)
Admin.	1(1.12)	40(44.94)	18(20.22)	11(12.36)	19(21.35)
Peer	1(1.12)	20(22.47)	16(17.98)	14(15.73)	38(42.70)
Self	1(1.12)	28(31.46)	22(24.72)	16(17.98)	22(24.72)

Admin.=Administrative; NR=No Response

Appendix E

VARIABLES/CHI SQUARE STATISTICAL INFORMATION

Appendix E1. A Comparison of Instructor Responses, by Sex, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fairness (No./Percent)

Type	Males					Females					df	X2
	Very Fair	Fair	Unfair	Very Unfair	NR	Very Fair	Fair	Unfair	Very Unfair	NR		
Stu.	0	33(37.08)	17(19.10)	2(2.25)	1(1.12)	3(3.37)	20(22.47)	8(8.99)	1(1.12)	4(4.49)	4	8.634
Adm.	5(5.62)	27(30.34)	11(12.36)	2(2.25)	8(8.99)	4(4.49)	25(28.09)	1(1.12)	0	6(6.74)	4	7.846
Peer	2(2.25)	20(22.47)	5(5.62)	1(1.12)	25(28.09)	4(4.49)	14(15.73)	1(1.12)	0	17(19.10)	4	3.808
Self	2(2.25)	36(40.45)	6(6.74)	2(2.25)	7(7.87)	7(7.87)	19(21.35)	2(2.25)	2(2.25)	6(6.74)	4	7.122

Objectivity (No./Percent)

Type	Males					Females					df	X2
	Very Obj.	Obj.	Non Obj.	Very Non Obj.	NR	Very Obj.	Obj.	Non Obj.	Very Non Obj.	NR		
Stu.	0	17(19.10)	31(34.83)	4(4.49)	1(1.12)	2(2.25)	13(14.61)	12(13.48)	5(5.62)	4(4.49)	4	9.9.56
Adm.	3(3.37)	23(25.84)	16(17.98)	3(3.37)	8(8.99)	2(2.25)	20(22.47)	7(7.87)	1(1.12)	6(6.74)	4	2.0444
Peer	1(1.12)	13(14.61)	13(14.61)	1(1.12)	25(28.09)	3(3.37)	11(12.36)	5(5.62)	0	17(19.10)	4	4.150
Self	1(1.12)	24(26.97)	14(15.73)	3(3.37)	11(12.36)	3(3.37)	14(15.73)	9(10.11)	3(3.37)	7(7.87)	4	2.450

Reliability (No./Percent)

Type	Males					Females					df	X2
	Very Rel.	Rel.	Unrel.	Very Unrel.	NR	Very Rel.	Rel.	Unrel.	Very Unrel.	NR		
Stu.	–	15(16.85)	29(32.58)	7(7.87)	2(2.25)	–	18(20.22)	10(11.24)	4(4.49)	4(4.49)	3	8.061
Adm.	2(2.25)	25(28.09)	14(15.73)	4(4.49)	8(8.99)	3(3.37)	22(24.72)	4(4.49)	1(1.12)	6(6.74)	4	4.967
Peer	1(1.12)	19(21.35)	4(4.49)	5(5.62)	24(26.97)	3(3.37)	12(13.48)	4(4.49)	0	17(19.10)	4	5.738
Self	2(2.25)	31(34.83)	7(7.87)	2(2.25)	11(12.36)	3(3.37)	21(23.60	5(5.62)	1(1.12)	6(6.74)	4	1.052

Note. Stu=Student; Adm.=Administrative; NR=No Response
Obj.=Objective; Rel.=Reliable; Unrel.=Unreliable

Appendix E2. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by Sex, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University

	Males (No./Percent)					Females (No./Percent)						
Type	Very Sat.	Satis.	Unsat.	Very Unsat.	NR	Very Sat.	Satis.	Unsat.	Very Unsat.	NR	df	X2
Stu.	0	18(20.22)	23(25.84)	11(12.36)	1(1.12)	1(1.12)	15(16.85)	11(12.36)	5(5.62)	4(4.49)	4	6.550
Adm.	4(4.49)	22(24.72)	14(15.73)	8(8.99)	5(5.62)	1(1.12)	21(23.60)	7(7.87)	2(2.25)	5(5.62)	4	4.680
Peer	2(2.25)	14(15.73)	7(7.87)	7(7.87)	23(25.84)	1(1.12)	15(16.85)	4(4.49)	3(3.37)	13(14.61)	4	2.404
Self	4(4.49)	25(28.09)	11(12.36)	3(3.37)	10(11.24)	5(5.62)	20(22.47)	4(4.49)	2(2.25)	5(5.62)	4	2.649

A Comparison of Instructor Responses, by Sex, to the Overall Satisfaction Level of the Evaluation Practiced at Belmont

Males (No./Percent)					Females (No./Percent)					
Very Sat.	Satis.	Unsat.	Very Unsat.	NR	Very Sat.	Satis.	Unsat.	Very Unsat.	NR	df	X2
0	22(24.72)	24(26.97)	6(6.74)	1(1.12)	1(1.12)	19(21.35)	11(12.36)	3(3.37)	2(2.25)	4	4.291

Note: Stu.=Student; Adm.=Administrator; Very Sat.=Very Satisfied; Satis.=Satisfied; Unsat.=Unsatisfied;NR=No Response

Appendix E3. A Comparison of Instructor Responses, by Sex, About Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

	Male (No./Percent)			Female (No./Percent)				
Type	Should Be Done	Should Not Be Done	NR	Should Be Done	Should Not Be Done	NR	df	X2
Stu.	42(47.19)	9(10.11)	2(2.25)	32(35.96)	1(1.12)	3(3.37)	2	4.882
Adm.	44(49.44)	6(6.740)	3(3.37)	31(34.83)	2(2.25)	3(3.37)	2	1.044
Peer	36(40.07)	12(13.48)	5(5.62)	26(29.21)	7(7.87)	3(3.37)	2	0.188
Self	41(46.07)	8(8.99)	4(4.49)	31(34.83)	1(1.12)	4(4.49)	2	3.722

Note. Stu=Student; Adm=Administrative; NR=No Response

Appendix E4. A Comparison of Instructor Responses, by Sex, to the Perceptions of the Usage of the Four Types of Evaluation Practiced at Belmont University

Student
(No./Percent)

Male					Female					
IU	SBU	I/SBU	NR	USE	IU	SBU	I/SBU	NR	df	X2
5(5.62)	26(29.21)	20(22.47)	2(2.25)	Imp. Tch.	1(1.12)	15(16.85)	15(16.85)	5(5.62)	3	4.536
3(3.37)	27(30.34)	9(10.11)	14(15.73)	Imp.P/M	1(1.12)	17(19.10)	10(11.24)	8(8.99)	3	1.799
20(22.47)	6(6.74)	17(19.10)	10(11.24)	MAD	8(8.99)	5(5.62)	10(11.24)	13(14.61)	3	4.351
2(2.25)	23(25.84)	11(12.36)	17(19.10)	DK	1(1.12)	15(16.85)	8(8.99)	12(13.48)	3	0.110

Administrative
(No./Percent)

Male					Female					
IU	SBU	I/SBU	NR	USE	IU	SBU	I/SBU	NR	df	X2
2(2.25)	23(25.84)	11(12.36)	17(19.10)	Imp. Tch.	1(1.12)	15(16.85)	8(8.99)	12(13.48)	3	0.110
1(1.12)	25(28.09)	5(5.62)	22(24.72)	Imp. P/M	2(2.25)	17(19.10)	8(8.99)	9(10.11)	3	4.934
10(11.24)	5(6.62)	17(19.10)	21(23.60)	MAD	6(6.74)	5(5.62)	15(16.85)	10(11.24)	3	1.848
11(12.36)	0	3(3.37)	39(43.82)	DK	6(6.74)	0	0	30(33.71)	2	2.488

Peer
(No./Percent)

Male					Female					
IU	SBU	I/SBU	NR	USE	IU	SBU	I/SBU	NR	df	X2
0	29(32.58)	2(2.25)	22(24.72)	Imp. Tch.	2(2.25)	15(16.85)	5(5.62)	14(15.73)	3	6.508
0	22(24.72)	1(1.12)	30(33.71)	Imp. P/M	1(1.12)	12(13.48)	6(6.74)	17(19.10)	3	8.159
2(2.25)	10(11.24)	4(4.49)	37(41.57)	MAD	4(4.49)	9(10.11)	4(4.49)	19(21.35)	3	3.381
22(24.72)	0	4(4.49)	27(30.34)	DK	5(5.62)	2(2.25)	1(1.12)	28(31.46)	3	11.702 *

Self
(No./Percent)

Male					Female					
IU	SBU	I/SBU	NR	USE	IU	SBU	I/SBU	NR	df	X2
6(6.74)	18(20.22)	20(22.47)	9(10.11)	Imp. Tch.	3(3.37)	7(7.87)	19(21.35)	7(7.87)	3	2.977
5(5.62)	16(17.98)	7(7.87)	25(28.09)	Imp. P/M	2(2.25)	10(11.24)	15(16.85)	9(10.11)	3	10.235
8(8.99)	7(7.87)	5(5.62)	33(37.08)	MAD	1(1.12)	5(5.62)	13(14.61)	17(19.10)	3	11.630
11(12.36)	0	0	42(47.19)	DK	4(4.49)	0	2(2.25)	30(33.71)	2	4.172

Note . IU=Is Used; SBU=Should Be Used; I/SBU=Is & Should Be Used; NR=No Response; Imp. Tch.=Improve Teaching; Imp. P/M= Improve Program/Major; MAD=Made Administrative Decisions; DK=Don't Know

Appendix E5. A Comparison of Instructor Responses, by Sex, to the Disposition of the Four Types of Evaluation Practiced at Belmont University

					Student (No./Percent)					
	Male						Female			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
33(37.08)	3(3.37)	15(16.85)	2(2.25)	RINR	18(20.62)	1(1.12)	10(11.24)	7(7.87	3	6.167
2(2.25)	17(19.10)	2(2.25)	32(35.96)	RIFR	0	17(19.10)	2(2.25)	17(19.10)	3	3.471
3(3.37)	2(2.25)	0	48(53.93)	NOR	2(2.25)	0	0	34(38.20)	2	1.394
0	4(4.49)	0	53(59.55)	NOTH.	0	0	0	36(40.45)	1	2.845
					Administrator (No./Percent)					
	Male						Female			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
8(8.99)	4(4.49)	6(6.74)	35(39.33)	RINR	10(11.24)	1(1.12)	5(5.62)	20(22.47)	3	3.069
1(1.12)	15(16.85)	6(6.64)	31(34.83)	RIFR	1(1.12)	17(19.10)	4(4.49)	14(15.73)	3	3.846
13(14.61)	1(1.12)	2(2.25)	37(41.57)	NOR	3(3.37)	1(1.12)	0	32(35.96)	3	5.568
7(7.87)	1(1.12)	1(1.12)	44(49.44)	NOTH.	4(4.49)	0	0	32(35.96)	3	1.521
					Peer (No./Percent)					
	Male						Female			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
3(3.37)	8(8.99)	2(2.25)	40(44.94)	RINR	2(2.25)	2(2.25)	3(3.37)	29(32.58)	3	2.601
1(1.12)	10(11.24)	0	42(47.19)	RIFR	0	13(14.61)	0	23(25.84)	2	3.838
8(8.99)	1(1.12)	1(1.12)	43(48.31)	NOR	3(3.37)	0	0	33((37.08)	3	2.430
18(20.22)	5(5.62)	0	30(33.71)	NOTH	9(10.11)	0	0	36(40.45)	2	5.097
					Self (No./Percent)					
	Male						Female			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
3(3.37)	1(1.12)	2(2.25)	47(52.81)	RINR	11(12.36)	1(1.12)	4(4.49)	20(22.47)	3	13.359
1(1.12)	12(13.48)	5(5.62)	35(39.33)	RIFR	0	11(12.36)	4(4.49)	21(23.60)	3	1.461
10(11.24)	1(1.12)	2(2.25)	40(44.94)	NOR	3(3.37)	0	2(2.25)	31(34.83)	3	2.764
13(14.61)	0	4(4.49)	36(40.45)	NOTH.	6(6.74	1(1.12)	1(1.12)	28(31.46)	3	3.256

Note. ID=Is Done; SBD=Should Be Done; I/SBD= Is & Should be Done; NR=No Response; Disp.=Disposition; RINR=Return to Instructor/No Response Requested; RIFR=Return to Instructor for Response; NOR=Not Returned; Noth.=Nothing

Appendix E6. A Comparison of Instructor Responses, by Sex, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University to Improve Teaching, Improve Program/Major, and to Make Administrative Decisions

Improve Teaching
(No./Percent)

	Male					Female						
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu.	0	21(23.60)	25(28.09)	6(6.74)	1(1.12)	5(5.62)	14(15.73)	11(12.36)	2(2.25)	4(4.49)	4	12.867
Adm.	0	20(22.47)	17(19.10)	10(11.24)	6(6.74)	1(1.12)	14(15.73)	8(8.99)	5(5.62)	8(8.99)	4	4.156
Peer	1(1.12)	9(10.11)	10(11.24)	12(13.48)	21(23.60)	3(3.37)	5(5.62)	8(8.99)	5(5.62)	15(16.85)	4	3.114
Self	0	26(29.21)	14(15.73)	5(5.62)	8(8.99)	7(7.87)	15(16.85)	7(7.87)	2(2.25)	5(5.62)	4	11.433*

Improve Program/Major
(No./Percent)

	Male					Female						
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu.	0	14(15.73)	28(31.46)	10(11.24)	1(1.12)	2(2.25)	13(14.61)	12(13.48)	4(4.49)	5(5.62)	4	10.823
Adm.	1(1.12)	16(17.88)	15(16.85)	15(16.85)	6(6.74)	0	14(15.73)	9(10.11)	6(6.74)	7(7.87)	4	3.346
Peer	--	--	--	--	--	--	--	--	--	--	-	--
Self	1(1.12)	19(21.35)	14(15.73)	19(11.24)	9(10.11)	3(3.37)	12(13.48)	12(13.48)	4(4.49)	5(5.62)	4	3.323

Make Administrative Decisions
(No./Percent)

	Male					Female						
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu.	0	16(17.98)	22(24.72)	12(13.48)	3(3.37)	0	12(13.48)	10(11.24)	4(4.49)	19(11.24)	3	9.957*
Adm.	0	24(26.97)	12(13.48)	9(10.11)	8(8.99)	1(1.12)	16(17.98)	6(6.74)	2(2.25)	11(12.36)	4	6.519
Peer	0	13(14.61)	11(12.36)	8(8.99)	21(23.60)	1(1.12)	7(7.87)	5(5.62)	6(6.74)	17(19.10)	4	2.605
Self	0	18(20.22)	13(14.61)	11(12.36)	11(12.36)	1(1.12)	10(11.24)	9(10.11)	5(5.62)	11(12.36)	4	3.130

Note. Stu.=Student; Adm.=Administrative; VE=Very Effective; E=Effective; INE=Ineffective; VIE=Very Ineffective; NR= No Response

Appendix E7. A Comparison of Instructor Responses, by Age, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation at Belmont University

Fairness (Percent)

	21-30					31-40					41-50					51-60						61-70						
Type	VF	F	LF	VUF	NR	VF	F	LF	VUF	NR	VF	F	LF	VUF	NR	VF	F	LF	VUF	NR	VR	F	LF	VUF	NR	df	χ2
Stu	1(1.12)	1(1.12)	0	0	0	0	18(20.22)	8(8.99)	2(2.25)	0	1(1.12)	18(20.22)	8(8.99)	1(1.12)	2(2.25)	1(1.12)	13(14.61)	4(4.49)	0	3(3.37)	0	3(3.37)	5(5.62)	0	0	16	26.843
Adm	0	2(2.25)	0	0	0	2(2.25)	20(22.47)	5(5.62)	0	1(1.12)	3(3.37)	16(17.98)	4(4.49)	1(1.12)	6(6.74)	3(3.37)	10(11.24)	1(1.12)	1(1.12)	6(6.74)	1(1.12)	4(4.49)	2(2.25)	0	1(1.12)	16	12.695
Peer	0	2(2.25)	0	0	0	3(3.37)	13(14.61)	2(2.25)	0	10(11.24)	2(2.25)	10(11.24)	3(3.37)	0	15(16.85	0	7(7.87)	1(1.12)	0	13(14.61)	1(1.12)	2(2.25)	0	1(1.12)	4(4.49)	16	19.929
Self	0	2(2.25)	0	0	0	3(3.37)	19(21.35)	2(2.25)	3(3.37)	1(1.12)	4(4.49)	18(20.22)	3(3.37)	0	5(5.62)	2(2.25)	11(12.36)	2(2.25)	0	6(6.74)	0	5(5.62)	1(1.12)	1(1.12)	1(1.12)	16	14.204

Objectivity (Percent)

	21-30					31-40					41-50					51-60						61-70						
Type	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	χ2
Stu	0	2(2.25)	0	0	0	0	7(7.87)	16(17.98)	5(5.62)	0	1(1.12)	12(13.48)	12(13.48)	3(3.37)	2(2.25)	1(1.12)	6(6.74)	11(12.36)	0	3(3.37)	0	3(3.37)	4(4.49)	1(1.12)	0	16	16.348
Adm.	0	2(2.25)	0	0	0	0	16(17.98)	10(11.24)	1(1.12)	1(1.12)	2(2.25)	11(12.36)	10(11.24)	1(1.12)	6(6.74)	2(2.25)	9(10.11)	2(2.25)	2(2.25)	6(6.74)	1(1.12)	5(5.62)	1(1.12)	0	1(1.12)	16	18.133
Peer	0	2(2.25)	0	0	0	2(2.25)	9(10.11)	7(7.87)	0	10(11.24)	1(1.12)	5(5.62)	9(10.11)	0	15(16.85	0	6(6.74)	2(2.25)	0	13(14.61)	1(1.12)	2(2.25)	0	1(1.12)	4(4.49)	16	25.969
Self	0	1(1.12)	1(1.12)	0	0	1(1.12)	12(13.48)	12(13.48)	2(2.25)	1(1.12)	2(2.25)	9(10.11)	8(8.99)	2(2.25)	9(10.11	1(1.12)	12(13.48)	1(1.12)	1(1.12)	6(6.74)	0	4(4.49)	1(1.12)	1(1.12)	292.25)	16	17.996

Reliability (Percent)

	21-30					31-40					41-50					51-60						61-70						
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	χ2
Stu	0	2(2.25)	0	0	0	0	11(12.36)	13(14.61)	4(4.49)	0	0	12(13.48)	10(11.24)	5(5.62)	3(3.37)	0	6(6.74)	11(12.36)	1(1.12)	3(3.37)	0	2(2.25)	5(5.62)	1(1.12)	0	12	12.197
Adm	0	2(2.25)	0	0	0	1(1.12)	14(15.73)	9(10.11)	3(3.37)	1(1.12)	2(2.25)	16(17.98)	5(5.62)	1(1.12)	6(6.74)	1(1.12)	11(12.36)	2(2.25)	1(1.12)	6(6.74)	1(1.12)	494.48)	2(2.25)	0	1(1.12)	16	13.530
Peer	0	2(2.25)	0	0	0	2(2.25)	13(14.61)	3(3.37)	0	10(11.24)	1(1.12)	8(8.99)	4(4.49)	3(3.37)	14(15.73	0	6(6.74)	1(1.12)	1(1.12)	13(14.61)	1(1.12)	2(2.25)	0	1(1.12)	4(4.49)	16	15.451
Self	0	2(2.25)	0	0	0	2(2.25)	18(20.22)	6(6.74)	1(1.12)	1(1.12)	1(1.12)	16(17.98)	4(4.49)	1(1.12)	8(8.99)	2(2.25)	12(13.48)	1(1.12)	0	6(6.74)	0	4(4.49)	1(1.12)	1(1.12)	2(2.25)	16	13.995

Note: Stu=Students; Adm=Administrators; VF=Very Fair; F=Fair; LF=Unfair; VUF=Very Unfair; NR=No Response; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR= Unreliable; VUR=Very Unreliable

Appendix E8. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by Age, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont (Fe' University cont.)

	21-30					31-40					41-50					51-60					61-70						
Type	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	L	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
Stu	0	2(2.25)	0	0	0	0	10(11.24)	13(14.61)	5(5.62)	0	1(1.12)	13(14.61)	8(8.99)	6(6.74)	2(2.25)	0	6(6.74)	9(10.11)	3(3.37)	3(3.37)	0	2(2.25)	4(4.49)	2(2.25)	0	16	13.682
Adm	0	2(2.25)	0	0	0	1(1.12)	13(14.61)	9(10.11)	4(4.49)	1(1.12)	1(1.12)	13(14.61)	7(7.87)	4(4.49)	5(5.62)	2(2.25)	11(12.36	3(3.37)	1(1.12)	4(4.49)	1(1.12)	4(4.49)	2(2.25)	1(1.12)	0	16	11.204
Peer	0	2(2.25)	0	0	0	4(4.49)	15(16.85)	8(8.99)	0	1(1.12)	3(3.37)	14(15.73)	3(3.37)	3(3.37)	7(7.87)	0	6(6.74)	1(1.12)	3(3.37)	11(12.36	1(1.12)	1(1.12)	1(1.12)	1(1.12)	4(4.49)	16	12.857
Self	-	- -	-	-	-	- -	- - -	- - -	-	-	- -	- - -	- -	- -	- - -	2(2.25)	10(11.24	3(3.37)	1(1.12)	5(5.62)	0	4(4.49)	1(1.12)	1(1.12)	2(2.25)	16	14.617

A Comparison of Instructor Responses, by Age, to the Perception of Overall Satisfaction Level of the Evaluation Practices at Belmont University

	21-30					31-40					41-50					51-60					61-70						
VS	S	L	VUS	NR	VS	S	U	VUS	NR	VS	S	L	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
0	2(2.25)	0	0	0	0	13(14.61	12(13.48	3(3.37)	0	1(1.12)	15(16.85	11(12.36	2(2.25)	1(1.12)	0	7(7.87)	10(11.24	2(2.25)	2(2.25)	0	4(4.49)	2(2.25)	2(2.25)	0	16	11.776

Note: Stu=Student; Adm.=Administrative; VS=Very Satisfactory; S=Satisfactory; U=Unsatisfactory; VUS=Very Unsatisfactory; NR= No Response
--=Insufficient responses to perform chi-square calculations

Appendix E9. A Comparison of Instructor Responses, by Age, to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

	21-30			31-40			41-50			51-60			61-70				
Type	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	df	X2
Stu	2(2.25)	0	0	25(28.09)	3(3.37)	0	26(29.21)	2(2.25)	2(2.25)	17(19.10)	2(2.25)	2(2.25)	4(4.49)	3(3.37)	1(1.12)	8	10.053
Adm.	2(2.25)	0	0	24(26.97)	3(3.37)	1(1.12)	25(28.09)	2(2.25)	3(3.37)	17(19.10)	2(2.25)	2(2.25)	7(7.87)	1(1.12)	0	8	2.480
Peer	1(1.12)	0	0	23(25.84)	5(5.62)	0	19(21.35)	7(7.87)	4(4.49)	12(13.48)	5(5.62)	4(4.49)	6(6.74)	2(2.25)	0	8	8.729
Self	2(2.25)	0	0	26(29.21)	1(1.12)	1(1.12)	22(24.72)	4(4.49)	4(4.49)	16(17.98)	2(2.25)	3(3.37)	6(6.74)	2(2.25)	0	8	7.443

Note: Stu.=Student; Adm.=Administrative; S=Should Be Done; SN= Should Not Be Done; NR= No Response

Appendix E-M

A Comparison of Instructor Responses, by Age, to the Perceived Usage of Student Administrator, Peer and Self-Evaluation at Belmont Univ.

Student Evaluation (Pu

Use	21-30 IU	21-30 SBU	21-30 I/SBU	21-30 NR	31-40 IU	31-40 SBU	31-40 I/SBU	31-40 NR	41-50 IU	41-50 SBU	41-50 I/SBU	41-50 NR	51-60 IU	51-60 SBU	51-60 I/SBU	51-60 NR	61-70 IU	61-70 SBU	61-70 I/SBU	61-70 NR	df	X2
Imp Tch.	0	2(2.25)	0	0	1(1.12)	15(16.85)	12(13.48)	0	3(3.37)	15(16.85)	10(11.24)	2(2.25)	0	7(7.87)	10(11.24)	4(4.49)	2(2.25)	2(2.25)	3(3.37)	1(1.12)	12	17.107
Imp P/M	0	2(2.25)	0	0	1(1.12)	14(15.73)	7(7.87)	6(6.74)	2(2.25)	15(16.85)	6(6.74)	7(7.87)	0	11(12.36)	4(4.49)	6(6.74)	1(1.12)	2(2.25)	2(2.25)	3(3.37)	12	6.608
MAD	0	1(1.12)	0	1(1.12)	9(10.11)	4(4.49)	12(13.48)	3(3.37)	10(11.24)	4(4.49)	8(8.99)	8(8.99)	4(4.49)	1(1.12)	7(7.87)	9(10.11)	5(5.62)	1(1.12)	0	2(2.25)	12	17.496
D/K	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-	-	-	-	-

Administrative Evaluation (Per cent)

Use	21-30 IU	21-30 SBU	21-30 I/SBU	21-30 NR	31-40 IU	31-40 SBU	31-40 I/SBU	31-40 NR	41-50 IU	41-50 SBU	41-50 I/SBU	41-50 NR	51-60 IU	51-60 SBU	51-60 I/SBU	51-60 NR	61-70 IU	61-70 SBU	61-70 I/SBU	61-70 NR	df	X2
Imp. Tch.	0	2(2.25)	0	0	1(1.12)	14(15.73)	7(7.87)	6(6.74)	2(2.25)	10(11.24)	6(6.74)	12(13.48)	0	10(11.24)	6(6.74)	5(5.62)	0	2(2.25)	0	6(6.74)	12	15.214
Imp P/M	0	2(2.25)	0	0	2(2.25)	14(15.73)	5(5.62)	7(7.87)	1(1.12)	13(14.61)	4(4.49)	12(13.48)	0	10(11.24)	4(4.49)	7(7.87)	0	3(3.37)	0	5(5.62)	12	9.166
MAD	0	0	0	2(2.25)	6(6.74)	5(5.62)	13(14.61)	4(4.49)	6(6.74)	3(3.37)	8(8.99)	13(14.61)	3(3.37)	2(2.25)	9(10.11)	7(7.87)	1(1.12)	0	2(2.25)	5(5.62)	12	14.283
D/K	0	0	0	2(2.25)	6(6.74)	0	2(2.25)	20(22.47)	8(8.99)	0	0	22(24.72)	1(1.12)	0	0	20(22.47)	2(2.25)	0	1(1.12)	5(5.62)	8	10.128

Peer Evaluation (Per cent)

Use	21-30 IU	21-30 SBU	21-30 I/SBU	21-30 NR	31-40 IU	31-40 SBU	31-40 I/SBU	31-40 NR	41-50 IU	41-50 SBU	41-50 I/SBU	41-50 NR	51-60 IU	51-60 SBU	51-60 I/SBU	51-60 NR	61-70 IU	61-70 SBU	61-70 I/SBU	61-70 NR	df	X2
Imp. Tch.	0	1(1.12)	0	1(1.12)	1(1.12)	18(20.22)	2(2.25)	7(7.87)	1(1.12)	15(16.85)	3(3.37)	11(12.36)	0	8(8.99)	2(2.25)	11(12.36)	0	2(2.25)	0	6(6.74)	12	9.792
Imp P/M	0	1(1.12)	0	1(1.12)	1(1.12)	13(14.61)	2(2.25)	12(13.48)	0	12(13.48)	3(3.37)	15(16.85)	0	5(5.62)	2(2.25)	14(15.73)	0	3(3.37)	0	5(5.62)	12	6.444
MAD	0	0	0	2(2.25)	4(4.49)	12(13.48)	1(1.12)	11(12.36)	0	4(4.49)	3(3.37)	23(25.84)	2(2.25)	1(1.12)	4(4.49)	14(15.73)	0	2(2.25)	0	6(6.74)	12	23.722
D/K	0	0	0	2(2.25)	10(11.24)	0	4(4.49)	14(15.73)	9(10.11)	1(1.12)	0	20(22.47)	3(3.37)	1(1.12)	1(1.12)	16(17.98)	5(5.62)	0	0	3(3.37)	12	15.694

Self Evaluation (Per cent)

Use	21-30 IU	21-30 SBU	21-30 I/SBU	21-30 NR	31-40 IU	31-40 SBU	31-40 I/SBU	31-40 NR	41-50 IU	41-50 SBU	41-50 I/SBU	41-50 NR	51-60 IU	51-60 SBU	51-60 I/SBU	51-60 NR	61-70 IU	61-70 SBU	61-70 I/SBU	61-70 NR	df	X2
Imp. Tch.	0	2(2.25)	0	0	3(3.37)	8(8.99)	15(16.85)	2(2.25)	5(5.62)	8(8.99)	12(13.48)	5(5.62)	0	7(7.87)	9(10.11)	5(5.62)	1(1.12)	0	3(3.37)	4(4.49)	12	18.652
Imp P/M	0	2(2.25)	0	0	2(2.25)	8(8.99)	9(10.11)	9(10.11)	4(4.49)	10(11.24)	5(5.62)	11(12.36)	0	6(6.74)	6(6.74)	9(10.11)	1(1.12)	0	2(2.25)	5(5.62)	12	13.574
MAD	0	1(1.12)	0	1(1.12)	5(5.62)	4(4.49)	7(7.87)	12(13.48)	2(2.25)	5(5.62)	6(6.74)	17(19.10)	2(2.25)	1(1.12)	4(4.49)	14(15.73)	0	1(1.12)	1(1.12)	6(6.74)	12	9.225
D/K	0	0	0	2(2.25)	8(8.99)	0	1(1.12)	19(21.35)	3(3.37)	0	1(1.12)	26(29.21)	3(3.37)	0	0	18(20.22)	1(1.12)	0	0	7(7.87)	8	5.605

Note: Imp. Tch.=Improve Teaching, Imp. P/M=Improve Program/Major, MAD=Make Administrative Decisions, D/K=Don't Know
IU=Is Used, SBU=Should Be Used, I/SBU=Is and Should Be Used, NR=No Response,---=Insufficient numbers of responses for Chi-square analysis

Appendix E11. A Comparison of Instructor Responses, by Age, to the Perceived Disposition of the Four Types of Evaluation Practiced at Belmont University

Student Eval

Disp.	21-30 ID	21-30 SBD	21-30 I/SBD	21-30 NR	31-40 ID	31-40 SBD	31-40 I/SBD	31-40 NR	41-50 ID	41-50 SBD	41-50 I/SBD	41-50 NR	51-60 ID	51-60 SBD	51-60 I/SBD	51-60 NR	61-70 ID	61-70 SBD	61-70 I/SBD	61-70 NR	df	X2
RINR	2(2.25)	--	--	--	15(16.85	1(1.12)	10(11.24	2(2.25)	20(22.47	2(2.25)	5(5.62)	3(3.37)	9(10.11)	1(1.12)	7(7.87)	4(4.49)	5(5.62)	--	3(3.37)	--	12	8.693
RIFR	--	--	--	2(2.25)	1(1.12)	10(11.24	3(3.37)	14(15.73	1(1.12)	16(17.98	--	13(14.6)	--	7(7.87)	1(1.12)	12(14.61)	--	1(1.12)	--	7(7.87)	12	12.726
NOR	--	--	--	2(2.25)	2(2.25)	--	--	26(29.11	3(3.37)	--	--	27(30.34	--	1(1.12)	--	20(22.47)	--	1(1.12)	--	7(7.87)	8	8.635
NOTH.	--	--	--	2(2.25)	--	1(1.12)	--	27(30.34	--	2(2.25)	--	28(31.46	--	--	--	21(23.60)	--	1(1.12)	--	7(7.87)	4	2.662

Admin Eval

Disp.	21-30 ID	21-30 SBD	21-30 I/SBD	21-30 NR	31-40 ID	31-40 SBD	31-40 I/SBD	31-40 NR	41-50 ID	41-50 SBD	41-50 I/SBD	41-50 NR	51-60 ID	51-60 SBD	51-60 I/SBD	51-60 NR	61-70 ID	61-70 SBD	61-70 I/SBD	61-70 NR	df	X2
RINR	--	--	--	2(2.25)	6(6.74)	1(1.12)	6(6.74)	15(16.85	7(7.87)	--	2(2.25)	21(23.60	4(4.49)	3(3.37)	2(2.25)	12(13.48)	1(1.12)	1(1.12)	1(1.12)	5(5.62)	12	10.521
RIFR	--	--	--	2(2.25)	1(1.12)	9(10.11)	5(5.62)	13(14.61	1(1.12)	15(16.85	2(2.25)	12(13.48	--	6(6.74)	3(3.37)	12(13.48)	--	2(2.25)	--	6(6.74)	12	9.932
NOR	--	--	--	2(2.25)	2(2.25)	1(1.12)	--	25(28.09	5(5.62)	1(1.12)	1(1.12)	23(25.84	6(6.74)	--	1(1.12)	14(15.73)	3(3.37)	--	--	5(5.62)	12	9.084
NOTH.	1(1.12)	--	--	1(1.12)	1(1.12)	--	--	27(30.34	7(7.87)	1(1.12)	--	22(24.72	1(1.12)	--	--	20(22.47)	1(1.12)	--	1(1.12)	6(6.74)	12	21.581

Peer Eval.

Disp.	21-30 ID	21-30 SBD	21-30 I/SBD	21-30 NR	31-40 ID	31-40 SBD	31-40 I/SBD	31-40 NR	41-50 ID	41-50 SBD	41-50 I/SBD	41-50 NR	51-60 ID	51-60 SBD	51-60 I/SBD	51-60 NR	61-70 ID	61-70 SBD	61-70 I/SBD	61-70 NR	df	X2
RINR	--	--	--	2(2.25)	3(3.37)	4(4.49)	2(2.25)	19(21.35	1(1.12)	2(2.25)	3(3.37)	24(26.97	1(1.12)	3(3.37)	--	17(19.10)	--	1(1.12)	--	7(7.87)	12	6.870
RIFR	--	--	--	2(2.25)	--	9(10.11)	--	19(21.35	1(1.12)	10(11.24	--	19(21.35	--	4(4.49)	--	17(19.10)	--	--	--	8(8.99)	9	7.673
NOR	--	--	--	2(2.25)	1(1.12)	1(1.12)	1(1.12)	25(28.09	6(6.74)	--	--	24(26.97	3(3.37)	--	--	18(20.22)	1(1.12)	--	--	7(7.87)	12	8.040
NOTH.	1(1.12)	--	--	1(1.12)	10(11.24	--	2(2.25)	16(17.98	11(12.36	--	1(1.12)	18(20.22	3(3.37)	--	1(1.12)	17(19.10)	2(2.25)	--	1(1.12)	5(5.62)	8	5.272

Self Eval

Disp.	21-30 ID	21-30 SBD	21-30 I/SBD	21-30 NR	31-40 ID	31-40 SBD	31-40 I/SBD	31-40 NR	41-50 ID	41-50 SBD	41-50 I/SBD	41-50 NR	51-60 ID	51-60 SBD	51-60 I/SBD	51-60 NR	61-70 ID	61-70 SBD	61-70 I/SBD	61-70 NR	df	X2
RINR	--	--	--	2(2.25)	6(6.74)	--	3(3.37)	19(21.35	5(5.62)	2(2.25)	2(2.25)	21(23.60	3(3.37)	--	1(1.12)	17(19.10)	--	--	--	8(8.99)	12	8.798
RIFR	--	--	--	2(2.25)	--	7(7.87)	3(3.37)	18(20.22	1(1.12)	10(11.24	4(4.49)	15(16.85	--	5(5.62)	2(2.25)	14(15.73)	--	1(1.12)	--	7(7.87)	12	7.120
NOR	--	--	--	2(2.25)	6(6.74)	1(1.12)	1(1.12)	20(22.47	3(3.37)	--	--	27(30.34	3(3.37)	--	3(3.37)	15(16.85)	1(1.12)	--	--	7(7.87)	12	11.138
NOTH.	1(1.12)	--	--	1(1.12)	6(6.74)	1(1.12)	--	21(23.60	8(8.99)	--	1(1.12)	21(23.60	2(2.25)	--	1(1.12)	18(20.22)	2(2.25)	--	3(3.37)	3(3.37)	12	23.396

Note. Disp.=Disposition; RINR=Return to Instructor/No Response Expected; RIFR=Return to Instructor for Response; NOR=Not Returned; NOTH.=Nothing.
ID=Is Done, SBD=Should Be Done; I/SBD=Is & Should Be Done; NR= No Response

Appendix E12. A Comparison Of Instructor Responses, by Age, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching (Percent)

	21-30					31-40					41-50					51-60					61-70						
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu	1(1.12)	1(1.12)	--	--	--	--	12(13.48	14(15.73	2(2.25)	--	3(3.37)	10(11.24	12(13.48	3(3.47)	2(2.25)	1(1.12)	9(10.11)	6(6.74)	2(2.25)	3(3.37)	--	3(3.37)	4(4.49)	1(1.12)	--	16	18.375
Adm	--	1(1.12)	--	--	1(1.12)	--	12(13.48	8(8.99)	6(6.74)	2(2.25)	1(1.12)	12(13.48	7(7.87)	5(5.62)	5(5.62)	--	7(7.87)	7(7.87)	2(2.25)	5(5.62)	--	2(2.25)	3(3.37)	2(2.25)	1(1.12)	16	9.399
Peer	--	1(1.12)	--	--	1(1.12)	2(2.25)	4(4.49)	7(7.87)	5(5.62)	10(11.24	2(2.25)	5(5.62)	5(5.62)	8(8.99)	10(11.24	--	2(2.25)	4(4.49)	3(3.37)	12(13.48	--	2(2.25)	2(2.25)	1(1.12)	3(3.37)	16	9.497
Self	--	1(1.12)	--	--	1(1.12)	2(2.25)	14(15.73	10(11.24	1(1.12)	1(1.12)	3(3.37)	13(14.61	4(4.49)	5(5.62)	5(5.62)	1(1.12)	11(12.36	4(4.49)	--	5(5.62)	1(1.12)	2(2.25)	3(3.37)	1(1.12)	1(1.12)	16	17.838

Improving Program/Major

	21-30					31-40					41-50					51-60					61-70						
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu	--	--	1(1.12)	--	1(1.12)	--	10(11.24	14(15.73	4(4.49)	--	2(2.25)	12(13.48	10(11.24	4(4.49)	2(2.25)	--	2(2.25)	13(14.61	3(3.37)	3(3.37)	--	3(3.37)	2(2.25)	3(3.37)	--	16	24.638
Adm.	--	--	1(1.12)	--	1(1.12)	--	13(14.61	7(7.87)	6(6.74)	2(2.25)	--	9(10.11	7(7.87)	9(10.11	5(5.62)	1(1.12)	6(6.74)	7(7.87)	3(3.37)	4(4.49)	--	2(2.25)	2(2.25)	3(3.37)	1(1.12)	16	12.285
Peer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Self	--	--	1(1.12)	--	1(1.12)	1(1.12)	13(14.61	9(10.11	4(4.49)	1(1.12)	--	10(11.24	8(8.99)	7(7.87)	5(5.62)	2(2.25)	6(7.74	6(6.74	1(1.12	6(6.74	1(1.12	2(2.25	2(2.25	2(2.25	1(1.12	16	16.459

Making Administrative Decisions

	21-30					31-40					41-50					51-60					61-70						
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu	--	1(1.12)	--	--	1(1.12)	--	10(11.24	12(13.48	4(4.49)	2(2.25)	--	10(11.24	10(11.24	4(4.49	6(6.74)	--	6(6.74	7(7.87	5(5.62	3(3.37)	--	1(1.12	3(3.37	3(3.37	1(1.12	12	9.007
Adm	--	1(1.12)	--	--	1(1.12)	--	14(15.73	8(8.99)	2(2.25)	4(4.49)	1(1.12)	13(14.61	4(4.49)	4(4.49)	8(8.99)	--	10(11.24	4(4.49	3(3.37	4(4.49	--	2(2.25	2(2.25	2(2.25	2(2.25	16	8.981
Peer	--	--	1(1.12)	--	1(1.12)	--	8(8.99)	5(5.62)	5(5.62)	10(11.24	1(1.12)	6(6.74)	4(4.49)	7(7.87)	12(13.48	--	4(4.49	4(4.49	19(1.12	12(13.48	--	2(2.25	2(2.25	1(1.12	3(3.37	16	9.457
Self	--	--	1(1.12)	--	1(1.12)	--	13(12.36	8(8.99)	5(5.62)	4(4.49)	1(1.12)	8(8.99)	6(6.74)	6(6.74)	9(10.11	--	7(7.87	5(5.62	3(3.37	6(6.74	--	2(2.25	2(2.25	2(2.25	2(2.25	16	7.474

Note: Stu=Student, Adm.=Administrative, VE=Very Effective, E=Effective, INE=Ineffective, VIE=Very Ineffective, NR=No Response,--=Insufficient Responses in Cells to Make Chi Square Calculations

Appendix E13 A Comparison of Instructor Response, by Academic Rank, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practices at Belmont University

Fairness (Per cent)

Type	Instructor VF	F	UF	VUF	NR	Asst. Prof. VF	F	UF	VUF	NR	Assoc. Prof. VF	F	UF	VUF	NR	Full Prof. VF	F	UF	VUF	NR	df	X2
Stu	0	4(4.49)	3(3.37)	0	1(1.12)	1(1.12)	17(19.10)	4(4.49)	2(2.25)	2(2.25)	0	17(19.10)	7(7.87)	1(1.12)	2(2.25	2(2.25)	15(16.85)	11(12.36)	0	0	12	11.175
Adm.	0	6(6.74)	1(1.12)	0	1(1.12)	1(1.12)	15(16.85)	4(4.49)	0	6(6.74)	3(3.37)	15(16.85)	4(4.49)	1(1.12)	4(4.49	5(5.62)	16(17.98)	3(3.37)	1(1.12)	3(3.37)	12	6.846
Peer	0	5(5.62)	1(1.12)	0	2(2.25)	2(2.25)	11(12.36)	1(1.12)	1(1.12)	11(12.36)	2(2.25)	7(7.87)	2(2.25)	0	16(17.98	2(2.25)	11(12.36)	292.25)	0	13(14.68)	12	7.991
Self	1(1.12)	5(5.62)	1(1.12)	0	1(1.12)	2(2.25)	15(16.85)	1(1.12)	2(2.25)	6(6.74)	3(3.37)	15(16.85)	3(3.37)	2(2.25)	4(4.49)	3(3.37)	20(22.47)	3(3.37)	0	2(2.25)	12	7.087

Objectivity (Per cent)

Type	Instructor VO	O	UO	VUO	NR	Asst. Prof. VO	O	UO	VUO	NR	Assoc. Prof. VO	O	UO	VUO	NR	Full Prof. VO	O	UO	VUO	NR	df	X2
Stu	0	2(2.25)	4(4.49)	1(1.12)	1(1.12)	0	11(12.36)	9(10.11)	4(4.49)	2(2.25)	0	9(10.11)	14(15.73)	2(2.25)	2(2.25)	2(2.25)	8(8.99)	16(17.98)	2(2.25)	0	12	10.664
Adm.	0	11(12.36)	8(8.99)	1(1.12)	6(6.74)	1(1.12)	14(15.73)	7(7.87)	1(1.12)	4(4.49)	4(4.49)	13(14.61)	6(6.74)	2(2.25)	3(3.37)	4(4.49)	13(14.61)	6(6.74)	2(2.25)	3(3.37)	12	9.172
Peer	0	3(3.37)	3(3.37)	0	2(2.25)	1(1.12)	8(8.99)	5(5.62)	1(1.12)	11(12.36)	2(2.25)	5(5.62)	4(4.49)	0	16(17.98	1(1.12)	8(8.99)	6(6.74)	0	13(14.61)	12	7.987
Self	0	3(3.37)	1(1.12)	1(1.12)	3(3.37)	0	10(11.24)	7(7.87)	2(2.25)	7(7.87)	2(2.25)	10(11.24)	8(8.99)	2(2.25)	5(5.62)	2(2.25)	15(16.85)	7(7.87)	1(1.12)	3(3.37)	12	8.245

Reliability (Per cent)

Type	Instructor VR	R	UR	VUR	NR	Asst. Prof. VR	R	UR	VUR	NR	Assoc. Prof. VR	R	UR	VUR	NR	Full Prof VR	R	UR	VUR	NR	df	X2
Stu	0	4(4.49;	3(3.37)	0	1(1.12)	0	10(11.24)	10(11.24)	4(4.49)	2(2.25)	0	9(10.11)	12(13.48)	3(3.37)	3(3.37)	0	10(11.24)	14(15.73)	4(4.49)	0	9	5.360
Adm	0	5(5.62)	2(2.25)	0	1(1.12,	1(1.12)	10(11.24)	7(7.87)	2(2.25)	6(6.74)	2(2.25)	13(14.61)	6(6.74)	2(2.25)	4(4.49)	2(2.25)	19(21.35)	3(3.37)	1(1.12)	3(3.37)	12	7.723
Peer	0	4(4.49;	2(2.25)	0	2(2.25)	1(1.12)	10(11.24)	3(3.37)	1(1.12)	11(12.36)	2(2.25)	6(6.74)	1(1.12)	3(3.37)	15(16.85	1(1.12)	11(12.36)	2(2.25)	1(1.12)	13(14.61)	12	10.013
Self	0	4(4.49)	2(2.25)	0	2(2.25)	1(1.12)	13(14.61)	4(4.49)	1(1.12)	7(7.87)	2(2.25)	14(15.73)	4(4.49)	2(2.25)	5(5.62	2(2.25)	21(23.60)	2(2.25)	0	3(3.37)	12	9.071

Note: Stu=Student; Adm=Administrative; VF=Very Fair; F=Fair; UF= Unfair; VUF=Very Unfair; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR=Unreliable; VUR=Very Unreliable; NR=No Response

Appendix E14. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Response, by Academic Rank, to the Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University

(Percent)

	Instructor					Asst. Prof.					Assoc. Prof.					Full Prof.					
Type	VS	S	VS	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	U	VU	NR	df	X2
Stu.	0	3(3.37)	3(3.37)	1(1.12)	1(1.12)	0	12(13.48)	6(6.74)	6(6.74)	2(2.25)	0	10(11.24)	12(13.48)	3(3.37)	2(2.25	1(1.12)	8(8.99)	13(14.61)	6(6.74)	0	12	9.636
Adm.	0	5(5.62)	2(2.25)	0	1(1.12)	1(1.12)	11(12.36)	6(6.74)	5(5.62)	3(3.37)	1(1.12)	12(13.48)	6(6.74)	3(3.37)	5(5.62	3(3.37)	15(16.85)	7(7.87)	2(2.25)	1(1.12)	12	8.439
Peer	0	4(4.49)	2(2.25)	0	2(2.25)	1(1.12)	10(11.24)	2(2.25)	3(3.37)	10(11.24)	0	7(7.87)	3(3.37)	4(4.49)	13(14.61	2(2.25)	8(8.99)	4(4.49)	3(3.37)	11(12.36)	12	7.634
Self	0	4(4.49)	2(2.25)	0	2(2.25)	3(3.37)	11(12.36)	4(4.49)	2(2.25)	6(6.74)	2(2.25)	15(16.85)	3(3.37)	2(2.25)	5(5.62	4(4.49)	15(16.85)	6(6.74)	1(1.12)	2(2.25)	12	6.832

A Comparison of Instructor Response, by Academic Rank, to the Overall Satisfaction Level with the Evaluation System at Belmont University

(Percent)

	Instructor				Asst. Prof.					Assoc. Prof.					Full Professor					
VS	S	U	VU	NR	VS	S	U	VU	NR	VS	S	U	VU	NR	VS	S	U	VU	NR	df	X2
0	5(5.62)	2(2.25)	1(1.12)	0	1(1.12)	12(13.48)	9(10.11)	3(3.37)	1(1.12)	0	11(12.36)	12(13.48)	2(2.25)	2(2.25	0	13(14.61)	12(13.48)	3(3.37)	0	12	6.731

Note Stu=Student; Adm.=Administrative; VS=Very Satisfied; S=Satisfied; U=Unsatisfied; VU=Very Unsatisfied; Asst. Prof.=Assistant Professor; Assoc. Prof.=Associate Professor; Full Prof.=Full Professor

Appendix E15. A Comparison of Instructor Response, by Academic Rank, to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

	No./%	No./%	No./%	No./%	No./%	No./%	No./%	No./%	No./%	No./%	No./%	No./%		
	Instructor			Asst. Prof.			Assoc. Prof.			Full Prof.				
Type	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	df	X2
Stu.	6(6.74)	1(1.12)	1(1.12)	22(24.72)	3(3.37)	1(1.12)	24(26.97)	1(1.12)	2(2.25)	22(24.72)	5(5.62)	1(1.12)	6	3.900
Adm.	6(6.74)	1(1.12)	1(1.12)	20(22.47)	3(3.37)	3(3.37)	25(28.09)	0	2(2.25)	24(26.97)	4(4.49)	0	6	7.261
Peer	5(5.62)	2(2.25)	1(1.12)	20(22.47)	4(4.49)	2(2.25)	17(19.10)	6(6.74)	4(4.49)	20(24.47)	7(7.87)	1(1.12)	6	3.202
Self	6(6.74)	1(1.12)	1(1.12)	20(22.47)	3(3.37)	3(3.37)	24(24.72)	1(1.12)	4(4.49)	24(26.97)	4(4.49)	0	6	5.677

Note: Stu.=Student; Adm.=Administrative; S=Should; SN=Should Not; NR=No Response; Asst.Prof.=Assistant Professor; Assoc. Prof.=Associate Professor; Full Prof.=Full Professor

Appendix E16 A Comparison of Instructor Response, by Academic Rank, to the Perceived Usage of the Four Types of Evaluation Practiced at Belmont University

Use	Instructor				Asst. Prof.			Student (Per Evaluation cent)		Assoc. Prof.				Full Prof.				df	χ2
	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR		
Imp Tch.	1(1.12)	5(5.62)	2(2.25)	0	2(2.25)	15(16.85)	6(6.74)	3(3.37)	0	11(12.36)	13(14.61)	3(3.37)	3(3.37)	19(11.24)	14(15.73)	1(1.12)		9	10.555
Imp P/M	1(1.12)	4(4.49)	1(1.12)	2(2.25)	1(1.12)	10(11.24)	6(6.74)	9(10.11)	0	16(17.98)	5(5.62)	6(6.74)	2(2.25)	14(15.73)	7(7.87)	5(5.62)		9	6.216
MAD	4(4.49)	2(2.25)	1(1.12)	1(1.12)	8(8.99)	2(2.25)	8(8.99)	8(8.99)	4(4.49)	6(6.74)	9(10.11)	8(8.99)	12(13.48)	1(1.12)	9(10.11)	6(6.74)		9	11.886
DK	..	..	..	..	..	..	..	..	-	-	-	-	-	-	-	-		-	-

Use	Instructor				Asst. Prof.			Administrative (Per Evaluation cent)		Assoc. Prof.				Full Prof.				df	χ2
	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR		
Imp Tch.	0	5(5.62)	0	3(3.37)	1(1.12)	9(10.11)	4(4.49)	12(13.48)	0	13(14.61)	8(8.99)	6(6.74)	2(2.25)	11(12.36)	7(7.87)	8(8.99)		9	9.516
Imp P/M	0	4(4.49)	1(1.12)	3(3.37)	2(2.25)	9(10.11)	3(3.37)	12(13.48)	0	15(16.85)	4(4.49)	8(8.99)	1(1.12)	14(15.73)	5(5.62)	8(8.99)		9	5.849
MAD	2(2.25)	1(1.12)	1(1.12)	4(4.49)	6(6.74)	3(3.37)	8(8.99)	9(10.11)	2(2.25)	5(5.62)	12(13.48)	8(8.99)	6(6.74)	1(1.12)	11(12.36)	10(11.24)		9	8.002
DK	2(2.25)	0	0	6(6.74)	5(5.62)	0	2(2.25)	19(21.35)	6(6.74)	0	1(1.12)	20(22.47)	4(4.49)	0	0	24(26.97)		6	3.643

Use	Instructor				Asst. Prof.			Peer (Per Evaluation cent)		Assoc. Prof.				Full Prof.				df	χ2
	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR		
Imp Tch.	0	2(2.25)	1(1.12)	5(5.62)	2(2.25)	15(16.85)	1(1.12)	8(8.99)	0	13(14.61)	2(2.25)	12(13.48)	0	14(15.73)	3(3.37)	11(12.36)		9	8.910
Imp P/M	0	2(2.25)	1(1.12)	5(5.62)	1(1.12)	9(10.11)	2(2.25)	14(15.73)	0	12(13.48)	2(2.25)	14(15.73)	0	11(12.36)	2(2.25)	15(16.85)		9	3.687
MAD	0	0	0	8(8.99)	3(3.37)	8(8.99)	1(1.12)	14(15.73)	1(1.12)	9(10.11)	2(2.25)	15(16.85)	2(2.25)	2(2.25)	5(5.62)	19(21.35)		9	15.495
DK	1(1.12)	0	2(2.25)	5(5.62)	9(10.11)	0	1(1.12)	16(17.98)	1(11.24)	1(1.12)	1(1.12)	15(16.85)	7(7.87)	1(1.12)	1(1.12)	19(21.35)		9	9.113

Appendix E17. A Comparison of Instructor Responses, by Academic Rank, to the Perceived Disposition of Evaluation Information of the Four Types of Evaluation Practiced at Belmont University

		Instructor				Asst. Prof		Student (Per cent) Evaluation			Assoc. Prof.					Full Prof.			
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	4(4.49)	1(1.12)	1(1.12)	2(2.25)	20(22.47)	1(1.12)	2(2.25)	3(3.37)	11(12.36)	0	13(14.61)	3(3.37)	16(17.98)	2(2.25)	9(10.11)	1(1.12)	9	17.520
RIFR	0	393.37)	0	5(5.62)	1(1.12)	11(12.36)	1(1.12)	13(14.61)	1(1.12)	9(10.11)	2(2.25)	15(16.85)	0	11(12.36)	1(1.12)	16(17.98)	9	2.823
NOR	1(1.12)	0	0	7(7.87)	1(1.12)	0	0	25(28.09)	2(2.25)	0	0	25(28.09)	1(1.12)	2(2.25)	0	25(28.09)	6	5.629
NOTH	0	0	0	8(8.99)	0	3(3.37)	0	23(25.84)	0	1(1.12)	0	26(29.21)	0	0	0	28(31.46)	3	4.739

								Admin. (Per cent) Evaluation			Assoc. Prof.					Full Prof.		
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	1(1.12)	1(1.12)	0	6(6.74)	8(8.99)	0	4(4.49)	14(15.73)	4(4.49)	1(1.12)	4(4.49)	18(20.22	5(5.62)	3(3.37)	3(3.37)	17(19.10	9	7.714
RIFR	0	4(4.49)	0	4(4.49)	1(1.12)	8(8.99)	3(3.37)	14(15.73)	1(1.12)	10(11.24	3(3.37)	13(14.61	0	10(11.24	4(4.49)	14(15.73	9	3.224
NOR	1(1.12)	0	0	7(7.87)	2(2.25)	0	0	24(26.97)	6(6.74)	2(2.25)	0	19(21.35	7(7.87)	0	2(2.25)	19(21.35	9	13.003
NOTH	1(1.12)	0	0	7(7.87)	3(3.37)	0	1(1.12)	22(24.72)	4(4.49)	0	0	22(24.72	3(3.37)	0	0	25(28.09	9	5.031

		Instructor				Asst. Prof		Peer (Per cent) Evaluation			Assoc. Prof.					Full Prof.		
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	0	2(2.25)	0	6(6.74)	2(2.25)	2(2.25)	2(2.25)	20(22.47)	2(2.25)	1(1.12)	1(1.12)	23(25.84	1(1.12)	5(5.62)	2(2.25)	20(22.47	9	6.393
RIFR	0	3(3.37)	0	5(5.62)	0	6(6.74)		0	1(1.12)	9(10.11)	0	17(19.10	0	5(5.62)	0	23(25.84	6	4.940
NOR	1(1.12)	0	0	7(7.87)	2(2.25)	0	0	24(26.97)	2(2.25)	1(1.12)	1(1.12)	23(25.84	6(6.74)	0	0	22(24.72	9	7.752
NOTH	2(2.25)	0	0	6(6.74)	9(10.11)	0	2(2.25)	15(16.85)	10(11.24	0	2(2.25)	15(16.85	6(6.74)	0	1(1.12)	21(23.60	6	3.517

		Instructor				Asst. Prof		Self (Per cent) Evaluation			Assoc. Prof					Full Prof.		
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBDB	NR	df	X2
RINR	1(1.12)	1(1.12)	0	6(6.74)	5(5.62)	0	2(2.25)	19(21.35)	5(5.62)	0	2(2.25)	20(22.47	3(3.37)	1(1.12)	2(2.25)	22(24.72	9	6 650
RIFR	0	2(2.25)	0	6(6.74)	0	4(4.49)	3(3.37)	19(21.35)	1(1.12)	9(10.11)	2(2.25)	15(16.85)	0	8(8.99)	4(4.49)	16(17.98	9	6 598
NOR	1(1.12)	0	0	7(7.87)	2(2.25)	1(1.12)	1(1.12)	22(24.72)	5(5.62)	0	1(1.12)	21(23.60	5(5.62)	0	2(2.25)	21(23.60	9	4.072
NOTH	4(4.49)	0	0	4(4.49)	5(5.62)	0	1(1.12)	20(22.47)	6(6.74)	1(1.12)	2(2.25)	18(20.18	4(4.49)	0	2(2.25)	22(24.72	9	7 057

Note. Disp.=Disposition; ID=Is Done; SBD=Should Be Done; I/SBD=Is & Should Be Done; NR= No Response; Asst. Prof.=Assistant Professor; Assoc. Prof.=Associate Professor; Full Prof.= Full Professor; RINR=Return to Instructor, No Response Expected; RIFR=Return to Instructor for Response; NOR=Not Returned; NOTH.=Nothing

Appendix E18. A Comparison of Instructor Responses, by Academic Rank, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching

	Instructor					Asst. Prof.					Assoc. Prof.						Full Prof.					
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu.	0	5(5.62)	2(2.25)	0	1(1.12)	1(1.12)	12(13.48)	7(7.87)	4(4.49)	2(2.25)	2(2.25)	10(11.24)	11(12.36)	2(2.25)	2(2.25)	2(2.25)	8(8.99)	16(17.98)	2(2.25)	0	12	11.386
Adm.	0	4(4.49)	0	1(1.12)	3(3.37)	0	9(10.11)	6(6.74)	7(7.87)	4(4.49)	1(1.12)	10(11.24)	8(8.99)	5(5.62)	3(3.37)	0	11(12.36)	11(12.36)	2(2.25)	4(4.49)	12	12.526
Peer	0	2(2.25)	1(1.12)	0	5(5.62)	2(2.25)	6(6.74)	6(6.74)	3(3.37)	9(10.11)	1(1.12)	1(1.12)	4(4.49)	9(10.11)	12(13.48)	1(1.12)	5(5.62)	7(7.87)	5(5.62)	10(11.24)	12	12.616
Self	0	4(4.49)	1(1.12)	0	3(3.37)	2(2.25)	12(13.48)	5(5.62)	2(2.25)	5(5.62)	2(2.25)	10(11.24)	7(7.87)	4(4.49)	4(4.49)	3(3.37)	15(16.85)	8(8.99)	1(1.12)	1(1.12)	12	11.288

Improving Program/Major (Per cent)

	Instructor					Asst. Prof.					Assoc. Prof.						Full Prof.					
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu.	0	4(4.49)	2(2.25)	0	2(2.25)	1(1.12)	9(10.11)	8(8.99)	6(6.74)	2(2.25)	1(1.12)	6(6.74)	17(19.10)	1(1.12)	2(2.25)	0	8(8.99)	13(14.61)	7(7.87)	0	12	19.044
Adm.	0	4(4.49)	0	1(1.12)	3(3.37)	0	9(10.11)	6(6.74)	7(7.87)	4(4.49)	0	8(8.99)	9(10.11)	7(7.87)	3(3.37)	1(1.12)	9(10.11)	9(10.11)	6(7.74)	3(3.37)	12	10.088
Peer	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Self	0	3(3.37)	2(2.25)	0	3(3.37)	1(1.12)	10(11.24)	6(6.74)	4(4.49)	5(5.62)	1(1.12)	7(7.87)	10(11.24)	5(5.62)	4(4.49)	2(2.25)	11(12.36)	8(8.99)	5(5.62)	2(2.25)	12	8.115

Making Administrative Decisions (Per cent)

	Instructor					Asst. Prof.					Assoc. Prof.						Full Prof.					
Type	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	VE	E	INE	VIE	NR	df	X2
Stu	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Adm.	–	–	–	–	–	–	–	–	–	–	1(1.12)	10(11.24)	7(7.87)	3(3.37)	6(6.74)	0	15(16.85)	6(6.74)	4(4.49)	3(3.37)	12	10.065
Peer	0	2(2.25)	0	1(1.12)	5(5.62)	0	7(7.87)	6(6.74)	4(4.49)	9(10.11)	1(1.12)	3(3.37)	3(3.37)	6(6.74)	14(15.73)	0	8(8.99)	7(7.87)	3(3.37)	10(11.24)	12	11.062
Self	0	2(2.25)	1(1.12)	1(1.12)	4(4.49)	0	7(7.87)	6(6.74)	5(5.62)	8(8.99)	1(1.12)	8(8.99)	7(7.87)	4(4.49)	7(7.87)	0	11(12.36)	8(8.99)	6(6.74)	3(3.37)	12	9.025

Note. Stu=Student; Adm.=Administrative; VE=Very Effective; E=Effective; INE=Ineffective; VIE=Very Ineffective; NR=No Response

Appendix E19. A Comparison of Instructor Responses, by Tenure Status, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

						Fairness (Percent)						
	Tenured					Non-Tenured						
Type	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	df	X2
Stu.	2(2.25)	27(30.34)	16(17.98)	1(1.12)	1(1.12)	1(1.12)	24(26.97)	9(10.11)	2(2.25)	4(4.49)	8	5.549
Adm.	7(7.87)	28(31.46)	5(5.62)	2(2.25)	5(5.62)	1(1.12)	24(26.97)	6(6.74)	0	9(10.11)	8	14.068
Peer	3(3.37)	15(16.85)	4(4.49)	0	25(28.09)	2(2.25)	19(21.35)	2(2.25)	1(1.12)	16(17.98)	8	10.454
Self	6(6.74)	29(32.58)	6(6.74)	2(2.25)	4(4.49)	3(3.37)	24(26.97)	2(2.25)	2(2.25)	9(10.11)	8	6.239

						Objectivity Percent)						
	Tenured					Non-Tenured						
Type	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	X2
Stu.	2(2.25)	14(15.73)	24(29.21)	4(4.49)	1(1.12)	0	14(15.73)	17(19.10)	5(5.62)	4(4.49)	8	9.398
Adm.	4(4.49)	25(28.09)	10(11.24)	3(3.37)	5(5.62)	0	18(19.32)	12(13.48)	1(1.12)	9(10.11)	8	15.418
Peer	2(2.25)	11(12.36)	9(10.11)	0	25(28.09)	1(1.12)	13(14.61)	9(10.11)	1(1.12)	16(17.98)	8	13.349
Self	4(4.49)	22(24.72)	13(14.61)	3(3.37)	5(5.62)	0	15(16.85)	10(11.24)	3(3.37)	12(13.48)	8	9.707

						Reliability (Percent)						
	Tenured					Non-Tenured						
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	X2
Stu.	0	17(19.10)	21(23.60)	7(7.87)	2(2.25)	0	15(16.85)	17(19.10)	4(4.49)	4(4.49)	6	2.011
Adm.	3(3.37)	29(32.58)	7(7.87)	3(3.37)	5(5.62)	1(1.12)	18(20.22)	10(11.24)	2(2.25)	9(10.11)	8	14.407
Peer	2(2.25)	14(15.73)	3(3.37)	3(3.37)	25(28.09)	1(1.12)	17(19.10)	5(5.62)	2(2.25)	15(16.85)	8	13.724
Self	4(4.49)	31(34.83)	5(5.62)	2(2.25)	5(5.62)	1(1.12)	20(22.47)	7(7.87)	1(1.12)	11(12.36)	8	8.007

Note . Stu=Student; Adm.=Administrative; VF=Very Fair; F=Fair; UF=Unfair; VUF=Very Unfair; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR=Unreliable; VUR=Very Unreliable; NR=No Response

Appendix E20. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by Tenure Status, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University

(Percent)

	Tenured					Non-Tenured						
Type	VS	S	US	VUS	NR	VS	S	US	VUS	NR	df	X2
Stu.	1(1.12)	14(15.73)	23(25.84)	8(8.99)	1(1.12)	0	17(19.10)	11(12.36)	8(8.99)	4(4.49)	8	10.374
Adm.	3(3.37)	27(30.34)	10(11.24)	4(4.49)	3(3.37)	1(1.12)	16(17.98)	10(11.24)	6(6.74)	7(7.87)	8	14.481
Peer	1(1.12)	13(14.61)	6(6.74)	6(6.74)	21(23.60)	1(1.12)	16(17.98)	5(5.62)	4(4.49)	14(15.73)	8	16.061
Self	6(6.74)	25(28.09)	9(10.11)	3(3.37)	4(4.49)	3(3.37)	19(21.35)	6(6.74)	2(2.25)	10(11.24)	8	6.512

A Comparison of Instructor Responses, Compared by Tenure Status, to the Perception of Overall Satisfaction with the Evaluation System at Belmont University

(Percent)

Tenured					Non-Tenured						
VS	S	U	VU	NR	VS	S	U	VU	NR	df	X2
0	21(23.60)	20(22.47)	5(5.62)	1(1.12)	1(1.12)	20(22.47)	13(14.61)	4(4.49)	2(2.25)	8	5.584

Note. Stu.=Student; Adm.=Administrative; VS=Very Satisfied; S=Satisfied; U=Unsatisfied; VU=Very Unsatisfied; NR=No Response

Appendix E21. A Comparison of Instructor Responses, by Tenure Status, to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

				(Percent)				
	Tenured			Non-Tenured				
Type	S	SN	NR	S	SN	NR	df	X2
Stu.	39(43.82)	6(6.74)	2(2.25)	33(37.08)	4(4.49)	3(3.37)	4	.967
Adm.	42(47.19)	4(4.49)	1(1.12)	31(34.83)	4(4.49)	5(5.62)	4	4.245
Peer	33(37.08)	12(13.48)	2(2.25)	28(31.46)	7(7.87)	5(5.62)	4	6.666
Self	41(46.07)	5(5.62)	1(1.12)	30(33.71)	4(4.49)	6(6.74)	4	8.669

Note. Stu=Student; Adm.=Administrative; S=Should Be Done; SN=Should Not Be Done; NR=No Response

Appendix E22. A Comparison of Instructor Responses, by Tenure Status, to the Perceptions of Usage of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Percent)

Tenured					Non-Tenured					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
3(3.37)	18(20.22)	23(25.84)	3(3.37)	Imp. Tch.	3(3.37)	22(24.72)	11(12.36)	4(4.49)	6	4.598
2(2.25)	27(30.34)	10(11.24)	8(8.99)	Imp. P/M	2(2.25)	16(17.98)	8(8.99)	14(15.73)	6	5.584
14(15.73)	5(5.62)	17(19.16)	11(12.36)	MAD	13(14.61)	5(5.62)	10(11.24)	12(13.48)	6	5.081
--	--	--	--	DK	--	--	--	--	--	--

Administrative Evaluation (Percent)

Tenured					Non-Tenured					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
2(2.25)	19(21.35)	14(15.73)	12(13.48)	Imp. Tch.	1(1.12)	17(19.10)	5(5.62)	17(19.10)	6	7.900
1(1.12)	24(26.97)	8(8.99)	14(15.73)	Imp. P/M	2(2.25)	16(17.98)	5(5.62)	17(19.10)	6	4.684
6(6.74)	5(5.62)	21(23.60)	15(16.85)	MAD	9(10.11)	5(5.62)	10(11.24)	16(17.98)	6	6.161
9(10.11)	0	0	38(42.70)	DK	8(8.99)	3(3.37)	3(3.37)	29(32.58)	4	4.402

Peer Evaluation (Percent)

Tenured					Non-Tenured					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
0	24(26.97)	4(4.49)	19(21.35)	Imp. Tch.	2(2.25)	18(20.22)	3(3.37)	17(19.10)	6	4.708
0	20(22.47)	3(3.37)	24(26.97)	Imp. P/M	1(1.12)	12(13.48)	4(4.49)	23(25.84)	6	5.932
3(3.37)	9(10.11)	7(7.87)	28(31.46)	MAD	3(3.37)	9(10.11)	1(1.12)	27(30.34)	6	5.225
16(17.98)	2(2.25)	1(1.12)	28(31.46)	DK	10(11.24)	0	4(4.49)	26(29.21)	6	5.278

Self Evaluation (Percent)

Tenured					Non-Tenured					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
3(3.37)	12(13.48)	23(25.84)	9(10.11)	Imp. Tch.	6(6.74)	12(13.48)	15(16.85)	7(7.87)	6	3.360
2(2.25)	16(17.98)	13(14.61)	16(17.98)	Imp. P/M	5(5.62)	9(10.11)	8(8.99)	18(20.22)	6	5.819
5(5.62)	5(5.62)	14(15.73)	23(25.84)	MAD	4(4.49)	7(7.87)	4(4.49)	25(28.09)	6	7.250
7(7.87)	0	0	40(44.94)	DK	7(7.87)	0	2(2.25)	31(34.83)	4	4.627

Note . IU=Is Used; SBU=Should Be Used; I/SBU=Is & Should Be Used; NR=No Response; Imp. Tch.=Improve Teaching; Imp. P/M=Improve Program/Major; MAD=Make Administrative Decisions; DK=Don't Know

Appendix E23. A Comparison of Instructor Responses, by Tenure Status, to the Perceived Disposition of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Percent)

Tenured					Non-Tenured					
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
24(26.97)	2(2.25)	18(20.22)	3(3.37)	RINR	26(29.21)	2(2.25)	6(6.74)	6(6.74)	6	7.189
1(1.12)	18(20.22)	3(3.37)	25(28.09)	RIFR	1(1.12)	16(17.98)	1(1.12)	22(24.72)	6	2.438
3(3.37)	1(1.12)	0	43(48.31)	NOR	2(2.25)	0	0	38(42.70)	4	21.808
0	1(1.12)	0	46(51.69)	NOTH.	0	3(3.37)	0	37(41.57)	2	1.549

Administrative Evaluation (Percent)

Tenured					Non-Tenured					
ID	SBD	I/SBBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
7(7.87)	4(4.49)	7(7.87)	29(32.58)	RINR	11(12.36)	1(1.12)	4(4.49)	24(26.97)	6	4.781
1(1.12)	15(16,85)	7(7.87)	24(26.97)	RIFR	1(1.12)	15(16.85)	3(3.37)	21(23.60)	6	4.905
13(14.61)	1(1.12)	2(2.25)	31(34.83)	NOR	3(3.37)	1(1.12)	0	36(40.45)	6	8.844
5(5.62)	1(1.12)	0	41(46.07)	NOTH.	5(5.62)	0	1(1.12)	34(38.20)	6	4.837

Peer Evaluation (Percent)

Tenured					Non-Tenured					
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
3(3.37)	6(6.74)	3(3.37)	35(39.33)	RINR	2(2.25)	4(4.49)	2(2.25)	32(35.96)	6	0.973
1(1.12)	10(11.24)	0	36(40.45)	RIFR	0	12(13.48)	0	28(31.46)	4	2.267
7(7.87)	1(1.12)	1(1.12)	38(42.70)	NOR	4(4.49)	0	0	36(40.45)	6	2.722
13(14.61)	0	2(2.25)	32(35.96)	NOTH.	13(14.61)	0	3(3.37)	24(26.97)	4	1.231

Self-Evaluation (Percent)

Tenured					Non-Tenured					
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
6(6.74)	1(1.12)	4(4.49)	36(40.45)	RINR	8(8.99)	1(1.12)	2(2.25)	29(32.58)	6	1.847
1(1.12)	13(14.61)	6(6.74)	27(30.34)	RIFR	0	10(11.24)	3(3.37)	27(30.34)	6	3.075
10(11.24)	0	3(3.37)	34(38.20)	NOR	3(3.37)	1(1.12)	1(1.12)	35(39.33)	6	5.876
7(7.87)	1(1.12)	4(4.49)	35(39.33)	NOTH.	11(12.36)	0	1(1.12)	28(31.46)	6	4.990

Note . ID=Is Done; SBD=Should Be Done; I/SBD=Is & Should Be Done; NR=No Response; Disp.=Disposition; RINR=Return to Instructor/No Response Requested; RIFR=Return to Instructor for Response; NOR=Not Returned; NOTH.=Nothing

Appendix E24. A Comparison of Instructor Responses, by Tenure Status, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching
(Percent)

	Tenured					Non-Tenured						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	3(3.37)	15(16.85)	24(26.97)	4(4.49)	1(1.12)	2(2.25)	19(21.35)	11(12.36)	4(4.49)	4(4.49)	8	7.331
Adm.	1(1.12)	19(21.35)	16(17.98)	5(5.62)	6(6.74)	0	14(15.73)	8(8.99)	10(11.24)	8(8.99)	8	7.031
Peer	2(2.25)	5(5.62)	10(11.24)	10(11.24)	20(22.47)	2(2.25)	8(8.99)	8(8.99)	6(6.74)	16(17.98)	8	5.612
Self	5(5.62)	23(25.84)	13(14.61)	3(3.37)	3(3.37)	2(2.25)	17(19.10)	8(8.99)	3(3.37)	10(11.24)	8	12.334

Improving Program/Major
(Percent)

	Tenured					Non-Tenured						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	0	12(12.48)	26(29.21)	8(8.99)	4(4.49)	2(2.25)	14(15.73)	13(14.61)	6(6.74)	5(5.62)	8	9.829
Adm.	1(1.12)	15(16.85)	14(17.98)	10(11.24)	5(5.62)	0	14(15.73)	8(8.99)	10(11.24)	8(8.99)	8	5.575
Peer	3(3.37)	16(17.98)	17(19.10)	7(7.87)	4(4.49)	1(1.12)	14(15.73)	9(10.11)	6(6.74)	10(11.24)	8	8.522
--	--	--	--	--	--	--	--	--	--	--	--	--

Making Administrative Decisions
(Percent)

	Tenured					Non-Tenured						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	0	15(16.85)	20(22.47)	9(10.11)	3(3.37)	0	13(14.61)	10(11.24)	7(7.87)	10(11.24)	6	10.659
Adm.	1(1.12)	22(24.72)	11(12.36)	6(6.74)	7(7.87)	0	17(19.10)	6(6.74)	5(5.62)	12(13.48)	8	5.647
Peer	1(1.12)	10(11.24)	8(8.99)	7(7.87)	21(23.60)	0	9(10.11)	8(8.99)	6(6.74)	17(19.10)	8	4.498
Self	1(1.12)	17(19.10)	13(14.61)	9(10.11)	7(7.87)	0	10(11.24)	9(10.11)	6(6.74)	15(16.85)	8	9.060

Note . Stu.=Student; Adm.=Administrative; VE=Very Effective; E=Effective; IE=Ineffective; VIE=Very Ineffective; NR=No Response
--Indicates two few responses in cell to execute Chi Square computation

Appendix E25. A Comparison of Instructor Responses, by Teaching Status, of the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fairness (Percent)

	Teach Exc.					Teach/Adm. Duty						
Type	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	df	X2
Stu.	2(2.25)	43(48.31)	20(22.47)	3(3.37)	0	1(1.12)	10(11.24)	5(5.62)	0	0	8	90.185
Adm.	5(5.62)	43(48.31)	11(12.36)	2(2.25)	7(7.87)	4(4.49)	8(8.99)	1(1.12)	0	3(3.37)	8	23.030
Peer	6(6.74)	27(30.34)	5(5.62)	1(1.12)	29(32.58)	0	7(7.87)	1(1.12)	0	8(8.99)	8	7.901
Self	7(7.87)	44(49.44)	7(7.87)	4(4.49)	6(6.74)	2(2.25)	11(12.36)	1(1.12)	0	2(2.25)	8	32.420

Objectivity (Percent)

	Teach Exc.					Teach/Adm. Duty						
Type	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	X2
Stu.	1(1.12)	24(26.47)	34(38.20)	9(10.11)	0	1(1.12)	6(6.74)	9(10.11)	0	0	8	92.684
Adm.	3(3.37)	38(42.70)	17(19.10)	3(3.37)	7(7.87)	2(2.25)	4(4.49)	6(6.74)	1(1.12)	3(3.37)	8	22.229
Peer	4(4.49)	19(21.35)	15(16.85)	1(1.12)	29(32.58)	0	5(5.62)	3(3.37)	0	8(8.99)	8	7.446
Self	3(3.37)	31(34.83)	18(20.22)	5(5.62)	11(12.36)	1(1.12)	7(7.87)	5(5.62)	1(1.12)	2(2.25)	8	21.228

Reliability (Percent)

	Teach Exc.					Teach/Adm. Duty						
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	X2
Stu.	0	24(26.97)	33(37.08)	11(12.36)	0	0	9(10.11)	6(6.74)	0	1(1.12)	6	78.670
Adm.	4(4.49)	36(40.45)	16(17.98)	5(5.62)	7(7.87)	1(1.12)	10(11.24)	2(2.25)	0	3(3.37)	8	19.499
Peer	4(4.49)	25(28.09)	6(6.74)	5(5.62)	28(31.46)	0	6(6.74)	2(2.25)	0	8(8.99)	8	8.861
Self	3(3.37)	41(46.07)	11(12.36)	3(3.37)	10(11.24)	2(2.25)	11(12.36)	1(1.12)	0	2(2.25)	8	25.831

Note . Stu.=Student; Adm.=Administrative; Teach Exc.=Teach Exclusively; Teach/Adm. Duty=Teach +Have Administrative Duties; VF=Very Fair; F=Fair; UF=Unfair; VUF=Very Unfair; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR=Unreliable; VUR=Very Unreliable

Appendix E26. Instructor Perceptions of Satisfaction with The Belmont University Evaluation System

A Comparison of Instructor Responses, by Teaching Status, to the Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University
(Percent)

	Teach Exc.					Teach/Adm. Duty						
Type	VS	S	US	VUS	NR	VS	S	US	VUS	NR	df	X2
Stu.	1(1.12)	28(31.46)	26(29.21)	13(14.61)	0	0	5(5.62)	8(8.99)	3(3.37)	0	8	90.064
Adm.	2(2.25)	34(38.20)	18(20.22)	8(8.99)	6(6.74)	3(3.37)	8(8.99)	3(3.37)	2(2.25)	0	8	32.309
Peer	3(3.37)	22(24.72)	9(10.11)	7(7.87)	27(30.34)	0	7(7.87)	2(2.25)	3(3.37)	4(4.49)	8	10.586
Self	6(6.74)	36(40.45)	12(13.43)	4(4.49)	10(11.24)	3(3.37)	9(10.11)	3(3.37)	1(1.12)	0	8	29.100

A Comparison of Instructor Responses, by Teaching Status, to the Perception of Overall Satisfaction with the Evaluation System at Belmont
(Percent)

	Teach Exc.					Teach/Adm. Duty					
VS	S	U	VU	NR	VS	S	U	VU	NR	df	X2
1(1.12)	23(35.96)	27(30.34)	8(8.99)	0	0	8(8.99)	7(7.87)	1(1.12)	0	8	52.997*

Note. Stu.=Student; Adm.=Administrative; VS=Very Satisfied; S=Satisfied; US=Unsatisfied; VUS=Very Unsatisfied; NR=No Response
Teach Exc.=Teach Exclusively; Teach/Adm. Duty=Teach and Have Some Administrative Duties

Appendix E27. A Comparison of Instructor Responses, by Teaching Status, to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

(Percent)

	Teach Exc.			Teach/Adm. Duty				
Type	S	SN	NR	S	SN	NR	df	X2
Stu.	57(64.04)	10(11.24)	1(1.12)	16(17.98)	0	0	4	58.258*
Adm.	59(66.29)	6(6.74)	3(3.37)	14(15.73)	2(2.25)	0	4	24.567*
Peer	49(55.06)	16(17.98)	3(3.37)	12(13.48)	3(3.37)	1(1.12)	4	32.929*
Self	55(61.80)	9(10.11)	4(4.49)	16(17.98)	0	0	4	36.001*

Note . Teach Exc.=Teach Exclusively; Teach/Adm.Duty=Teach and Have Some Administrative Duties; Stu.=Student; Adm.=Administrative; S=Should Be Done; SN=Should Not Be Done; NR=No Response

Appendix E28. A Comparison of Instructor Responses, by Teaching Status, to the Perceived Usage of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Percent)

Teach Exc.					Teach/Adm. Duty					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
4(4.49)	35(39.33)	26(29.21)	3(3.37)	Imp. Tch.	2(2.25)	5(5.62)	9(10.11)	0	6	41.650*
3(3.37)	33(37.08)	15(16.85)	17(19.10)	Imp. P/M	1(1.12)	11(12.36)	4(4.49)	0	6	20.630*
24(26.97)	9(10.11)	21(23.60)	14(15.73)	MAD	4(4.49)	2(2.25)	6(6.74)	4(4.49)	6	15.928*
--	--	--	--	DK	--	--	--	--	--	--

Administrative Evaluation (Percent)

Teach Exc.					Teach/Adm. Duty					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
1(1.12)	33(37.08)	14(15.73)	20(22.47)	Imp. Tch.	2(2.25)	5(5.62)	5(5.62)	4(4.49)	6	17.308*
2(2.25)	35(39.33)	8(8.99)	23(25.84)	Imp. P/M	1(1.12)	7(7.87)	4(4.49)	4(4.49)	6	8.314
15(16.85)	8(8.99)	26(29.21)	19(21.35)	MAD	1(1.12)	2(2.25)	6(6.74)	7(7.87)	6	12.650
14(15.73)	0	3(3.37)	51(57.30)	DK	3(3.37)	0	13(14.61)	16(17.98)	4	2.372

Peer Evaluation (Percent)

Teach Exc.					Teach.Adm. Duty					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
2(2.25)	37(41.57)	6(6.74)	23(25.84)	Imp. Tch.	0	7(7.87)	1(1.12)	8(8.99)	6	9.543
1(1.12)	28(31.46)	6(6.74)	33(37.08)	Imp. P/M	0	6(6.74)	1(1.12)	9(10.11)	6	5.284
5(5.62)	16(17.98)	6(6.74)	41(46.07)	MAD	1(1.12)	3(3.37)	2(2.25)	10(11.24)	6	3.489
23(25.84)	1(1.12)	4(4.49)	40(44.94)	DK	3(3.37)	1(1.12)	1(1.12)	11(12.36)	6	3.384

Self-Evaluation (Percent)

Teach Exc.					Teach/Adm. Duty					
IU	SBU	I/SBU	NR	Use	IU	SBU	I/SBU	NR	df	X2
7(7.87)	21(23.60)	32(35.96)	8(8.99)	Imp. Tch.	2(2.25)	4(4.49)	7(7.87)	3(3.37)	6	24.776*
6(6.74)	20(22.47)	16(17.98)	26(29.21)	Imp. P/M	1(1.12)	6(6.74)	5(5.62)	4(4.49)	6	5.740
8(8.99)	9(10.11)	14(15.73)	37(41.57)	MAD	1(1.12)	3(3.37)	4(4.49)	8(8.99)	6	4.983
13(14.61)	0	1(1.12)	54(60.67)	DK	2(2.25)	0	1(1.12)	13(14.61)	4	2.909

Note. Teach Exc.=Teach Exclusively; Teach/Adm. Duty=Teach and Have Some Administrative Duties; IU=Is Used; SBU=Should Be Used; I/SBU=Is and Should Be Used; NR=No Response; Imp. Tch.=Improve Teaching; Imp. P/M=Improve Program/Major; MAD=Make Administrative Decisions; DK=Don't Know

Appendix E29. A Comparison of Instructor Responses, by Teaching Status, to the Perceived Disposition of Evaluation Information of the Four Types of Evaluation Practiced at Belmont University

				Student Evaluation (Percent)						
		Teach Exc.					Teach/Adm. Duty			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
41(46.07)	4(4.49)	21(23.60)	2(2.25)	RINR	10(11.24)	0	4(4.49)	2(2.25)	6	49.428
2(2.25)	26(29.21)	2(2.25)	38(42.70)	RIFR	0	7(7.87)	2(2.25)	7(7.87)	6	4.997
5(5.62)	2(2.25)	0	61(68.54)	NOR	0	0	0	16(17.98)	4	2.346
0	4(4.49)	0	64(71.91)	NOTH.	0	0	0	16(17.98)	2	1.293
				Administrative Evaluation (Percent)						
		Teach. Exc.					Teach/Adm. Duty			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBBD	NR	df	X2
16(17.98)	4(4.49)	10(11.24)	38(42.70)	RINR	2(2.25)	1(1.12)	1(1.12)	12(13.48)	6	5.572
2(2.25)	29(32.58)	6(6.74)	31(34.83)	RIFR	0	2(2.25)	4(4.49)	10(11.24)	6	9.519
14(15.73)	2(2.25)	2(2.25)	50(56.18)	NOR	2(2.25)	0	0	14(15.73)	6	3.330
9(10.11)	1(1.12)	1(1.12)	57(64.04)	NOTH.	2(2.25)	0	0	14(15.73)	6	1.431
				Peer Evaluation (Percent)						
		Teach Exc.					Teach/Adm. Duty			
ID	SBD	I/SBBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
4(4.49)	9(10.11)	5(5.62)	50(56.18)	RINR	1(1.12)	1(1.12)	0	14(15.73)	6	3.674
1(1.12)	19(21.35)	0	48(53.93)	RIFR	0	3(3.37)	0	13(14.61)	4	1.039
11(12.36)	1(1.12)	0	56(62.92)	NOR	0	0	1(1.12)	15(16.85)	6	8.598
22(24.72)	0	5(5.62)	41(46.07)	NOTH.	5(5.62)	0	0	11(12.36)	4	4.370
				Self-Evaluation (Percent)						
		Teach. Exc.					Teach/Adm. Duty			
ID	SBD	I/SBD	NR	Disp.	ID	SBD	I/SBD	NR	df	X2
12(13.48)	2(2.25)	6(6.74)	48(53.93)	RINR	2(2.25)	0	0	14(15.73)	6	4.444
1(1.12)	17(19.10)	5(5.62)	45(50.56)	RIFR	0	5(5.62)	4(4.49)	7(7.87)	6	6.380
12(13.48)	1(1.12)	4(4.49)	51(57.30)	NOR	1(1.12)	0	0	15(16.85)	6	4.312
15(16.85)	1(1.12)	5(5.62)	47(52.81)	NOTH.	4(4.49)	0	0	12(13.48)	6	3.680

Note. Teach Exc.=Teach Exclusively; Teach/Adm. Duty=Teach and Have Some Administrative Duties; ID=Is Done; SBD=Should Be Done; I/SBD=Is & Should Be Done; NR=No Response; Disp.=Disposition; RINR=Return to Instructor /No Response Expected; RIFR=Return to Instructor for Response; NOR=Not Returned; Noth.=Nothing

Appendix E30. A Comparison of Instructor Responses, by Teaching Status, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching
(Percent)

	Teach Exc.					Teach/Adm. Duty						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	4(4.49)	29(32.58)	28(31.46)	7(7.87)	0	1(1.12)	6(6.74)	8(8.99)	1(1.12)	0	8	89.575
Adm.	1(1.12)	26(29.21)	21(23.60)	14(15.73)	6(6.74)	0	8(8.99)	4(4.49)	1(1.12)	3(3.37)	8	31.649
Peer	4(4.49)	11(12.36)	14(15.73)	13(14.61)	26(29.21)	0	3(3.37)	4(4.49)	4(4.49)	5(5.62)	8	9.367
Self	6(6.74)	30(33.71)	19(21.35)	6(6.74)	7(7.87)	1(1.12)	11(12.36)	2(2.25)	1(1.12)	1(1.12)	8	34.349

Improving Program/Major
(Percent)

	Teach Exc.					Teach/Adm.Duty						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	2(2.25)	22(24.72)	33(37.08)	11(12.36)	0	0	5(5.62)	7(7.87)	3(3.37)	1(1.12)	8	74.658
Adm.	0	23(25.84)	20(22.47)	19(21.35)	6(6.74)	1(1.12)	7(7.87)	4(4.49)	2(2.25)	2(2.25)	8	37.374
Peer	--	--	--	--	--	--	--	--	--	--	--	--
Self	2(2.25)	25(28.09)	22(24.72)	12(13.48)	7(7.87)	2(2.25)	6(6.74)	4(4.49)	2(2.25)	2(2.25)	8	31.513

Make Administrative Decisions
(Percent)

	Teach Exc.					Teach/Adm. Duty						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	0	23(25.84)	25(28.09)	13(14.61)	7(7.87)	0	5(5.62)	7(7.87)	3(3.37)	1(1.12)	6	31.320
Adm.	1(1.12)	31(34.83)	15(16.85)	10(11.24)	11(12.36)	0	9(10.11)	3(3.37)	1(1.12)	3(3.37)	8	20.954
Peer	1(1.12)	17(19.10)	12(13.48)	11(12.36)	27(30.24)	0	3(3.37)	4(4.49)	3(3.37)	6(6.74)	8	8.043
Self	1(1.12)	20(22.47)	18(20.22)	15(16.85)	14(15.73)	0	8(8.99)	4(4.49)	1(1.12)	3(3.37)	8	19.958

Note . Stu.=Student; Adm.=Administrative; Teach Exc.=Teach Exclusively; Teach/Adm. Duty=Teach and Have Some Administrative Duties; VE=Very Effective; E=Effective; IE=Ineffective; VIE=Very Ineffective; NR=No Response; --=Insufficient Responses in Cells to make Chi Square Calculations;

Appendix E31. A Comparison of Instructor Responses, by School, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fairness (Per cent)

	Nur.					Rel					Hu./Ed.					Sci.					Bus.					Mus.				
Type	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	df	X2
Stu.	0	3(3.37)	0	1(1.12)	0	0	3(3.37)	0	0	0	1(1.12)	19(21.35)	10(11.24)	2(2.25)	0	1(1.12)	9(10.11)	4(4.49)	0	0	1(1.12)	11(12.36)	7(7.87)	0	0	8(8.99)	4(4.49)	0	0	24	101.980
Adm.	0	4(4.49)	0	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	1(1.12)	22(24.92)	6(6.74)	2(2.25)	1(1.12)	7(7.87)	6(6.74)	1(1.12)	0	0	0	10(11.24)	3(3.37)	0	6(6.74)	0	8(8.99)	2(2.25)	0	2(2.25)	24	62.646
Peer	1(1.12)	3(3.37)	0	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	4(4.49)	12(13.48)	2(2.25)	1(1.12)	13(14.61)	0	7(7.87)	0	0	7(7.87)	0	7(7.87)	1(1.12)	0	11(12.36)	0	4(4.49)	3(3.37)	0	5(5.62)	24	29.374
Self	1(1.12)	2(2.25)	0	1(1.12)	0	0	2(2.25)	0	0	1(1.12)	5(5.62)	23(25.41)	2(2.25)	1(1.12)	1(1.12)	3(3.37)	11(12.36)	0	0	0	0	10(11.24)	3(3.37)	1(1.12)	5(5.62)	0	7(7.87)	3(3.37)	1(1.12)	1(1.12)	24	57.784

Objectivity (Per cent)

	Nur.					Rel					Hu./Ed.					Sci.					Bus.					Mus.				
Type	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	X2
Stu.	0	2(2.25)	1(1.12)	1(1.12)	0	0	2(2.25)	1(1.12)	0	0	0	8(8.99)	18(20.22)	6(6.74)	0	2(2.25)	6(6.74)	6(6.74)	0	0	0	7(7.87)	11(12.36)	1(1.12)	0	0	5(5.62)	6(6.74)	1(1.12)	0	24	108.848
Adm.	0	3(3.37)	1(1.12)	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	1(1.12)	13(14.61)	14(15.73)	3(3.37)	1(1.12)	3(3.37)	8(8.99)	2(2.25)	1(1.12)	0	0	10(11.24)	3(3.37)	0	6(6.74)	0	7(7.87)	3(3.37)	0	2(2.25)	24	49.612
Peer	0	3(3.37)	1(1.12)	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	3(3.37)	6(6.74)	9(10.11)	1(1.12)	13(14.61)	0	6(6.74)	1(1.12)	0	7(7.87)	0	5(5.62)	3(3.37)	0	11(12.36)	0	3(3.37)	4(4.49)	0	5(5.62)	24	29.263
Self	0	4(4.49)	0	0	0	0	2(2.25)	0	0	1(1.12)	2(2.25)	14(15.73)	12(13.48)	3(3.37)	1(1.12)	2(2.25)	7(7.87)	4(4.49)	1(1.12)	0	0	8(8.99)	2(2.25)	1(1.12)	8(8.99)	0	3(3.37)	5(5.62)	1(1.12)	3(3.37)	24	49.666

Reliability (Per cent)

	Nur.					Rel					Hu./Ed.					Sci.					Bus.					Mus.				
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	X2
Stu.	0	2(2.25)	1(1.12)	1(1.12)	0	0	2(2.25)	1(1.12)	0	0	0	14(15.73)	13(14.61)	5(5.62)	0	0	5(5.62)	6(6.74)	2(2.25)	1(1.12)	0	6(6.74)	12(13.48)	1(1.12)	0	0	4(4.49)	6(6.74)	2(2.25)	0	18	79.731
Adm.	0	4(4.49)	0	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	2(2.25)	15(16.85)	10(11.24)	4(4.49)	1(1.12)	2(2.25)	10(11.24)	2(2.25)	0	0	0	10(11.24)	3(3.37)	0	6(6.74)	0	6(6.74)	3(3.37)	1(1.12)	2(2.25)	24	45.439
Peer	1(1.12)	2(2.25)	1(1.12)	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	2(2.25)	11(12.36)	4(4.49)	3(3.37)	12(13.48)	0	7(7.37)	0	0	7(7.87)	0	7(7.87)	0	1(1.12)	11(12.36)	0	3(3.37)	3(3.37)	1(1.12)	5(5.62)	24	32.453
Self	0	4(4.49)	0	0	0	0	2(2.25)	0	0	1(1.12)	3(3.37)	22(24.92)	5(5.62)	1(1.12)	1(1.12)	2(2.25)	11(12.36)	1(1.12)	0	0	0	8(8.99)	2(2.25)	1(1.12)	8(8.99)	0	5(5.62)	4(4.49)	1(1.12)	2(2.25)	24	50.431

Note: Nur.=Nursing; Rel.=Religion; Hu./Ed.=Humanities/Education; Stu.=Students; Adm.=Administrative; VF=Very Fair; F=Fair; U=Unfair; VUF=Very Unfair; UR=Unreliable, VUR=Very Unreliable

NR=No Response; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable, R=Reliable, Sci.=The Sciences; Bus.=Business; Mus.=Music

E 32

Appendix E . . Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by School, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont

(Per cent)

	Nur.					Rel.					Hum./Ed.					Sci.					Bus.					Mus.						
Type	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
Stu.	0	1(1.12)	2(2.25)	1(1.12)	0	0	2(2.25)	1(1.12)	0	0	0	10(11.24)	14(15.73)	8(8.99)	0	1(1.12)	4(4.49)	6(6.74)	3(3.37)	0	0	10(11.24)	6(6.74)	3(3.37)	0	0	6(6.74)	5(5.62)	1(1.12)	0	24	100.281
Adm.	0	2(2.25)	1(1.12)	0	1(1.12)	1(1.12)	1(1.12)	0	1(1.12)	0	0	17(19.10)	10(11.24)	5(5.62)	0	4(4.49)	7(7.87)	2(2.25)	0	1(1.12)	0	9(10.11)	5(5.62)	3(3.37)	2(2.25)	0	6(6.74)	3(3.37)	1(1.12)	2(2.25)	24	56.747
Peer	1(1.12)	1(1.12)	2(2.25)	0	0	1(1.12)	1(1.12)	0	0	1(1.12)	1(1.12)	13(14.61)	5(5.62)	3(3.37)	10(11.24	0	5(5.62)	2(2.25)	1(1.12)	6(6.74)	0	6(6.74)	0	4(4.49)	9(10.11)	0	3(3.37)	2(2.25)	2(2.25)	2(2.25)	24	36.807
Self	1(1.12)	3(3.37)	0	0	0	0	2(2.25)	0	0	1(1.12)	3(3.37)	18(20.22)	9(10.11)	1(1.12)	1(1.12)	5(5.62)	6(6.74)	2(2.25)	1(1.12)	0	0	8(8.99)	3(3.37)	1(1.12)	7(7.87)	0	8(8.99)	1(1.12)	2(2.25)	1(1.12)	24	59.663

A Comparison of Instructor Responses, by School, to the Perceived Overall Satisfaction with Evaluation Procedures at Belmont University

(Per cent)

	Nur.					Rel.					Hum./Ed.					Sci.					Bus.					Mus.						
VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
0	3(3.37)	1(1.12)	0	0	0	2(2.25)	1(1.12)	0	0	1(1.12)	11(12.36)	15(16.85)	5(5.62)	0	0	9(19.11)	5(5.62)	0	0	0	9(10.11)	8(8.99)	2(2.25)	0	0	6(6.74)	4(4.49)	2(2.25)	0	24	61.599

Note Nur.=Nursing; Rel.=Religion; Hum./Ed.=Humanities/Education; Sci.=The Sciences; Bus.=Business, Mus.=Music; Stu.=Student; Adm.=Administrative VS=Very Satisfied; S=Satisfied; U=Unsatisfied; VUS=Very Unsatisfied; NR=No Response

Appendix E33. A Comparison of Instructor Responses, by School, About Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

(Percent)

Type	Nur. S	Nur. N	Nur. NR	Rel. S	Rel. N	Rel. NR	Hum./Ed. S	Hum./Ed. N	Hum./Ed. NR	Sci. S	Sci. N	Sci. NR	Bus. S	Bus. N	Bus. NR	Mus S	Mus N	Mus NR	df	χ^2
Stu.	4(4.49)	0	0	3(3.37)	0	0	29(32.58	3(3.37)	0	11(12.36)	3(3.37)	0	16(17.98)	2(2.25)	1(1.12)	10(11.24)	2(2.25)	0	12	58.852
Adm.	3(3.37)	1(1.12)	0	2(2.25)	1(1.12)	0	27(30.34	4(4.49)	0	13(14.61)	1(1.12)	0	17(19.10)	1(1.12)	1(1.12)	11(12.36)	0	1(1.12)	12	30.188
Peer	4(4.49)	0	0	2(2.25)	1(1.12)	0	25(28.09	7(7.87)	0	9(10.11)	4(4.49)	1(1.12)	12(13.48)	4(4.49)	3(3.37)	9(10.11)	3(3.37)	0	12	39.064*
Self	4(4.49)	0	0	3(3.37)	0	0	29(32.58	2(2.25)	1(1.12)	10(11.24)	4(4.49)	0	15(16.85)	1(1.12)	3(3.37)	10(11.24)	2(2.25)	0	12	43.727*

Note. Stu.=Student; Adm.=Administrative; S=Should Be Done; N=Should Not Be Done; NR=No Response; Nur.=Nursing; Rel.=Religion; Hum./Ed.=Humanities/Education; Sci.=The Sciences; Bus.=Business; Mus.=Music

Appendix E34. A Comparison of Instructor Responses, by School, to the Perceived Usage of the Four Types of Evaluation Practiced at Belmont University

Student (Per Evaluation cent)

Use	Nur. IU	Nur. SBU	Nur. I/SBU	Nur. NR	Rel. IU	Rel. SBU	Rel. I/SBU	Rel. NR	Hum./Ed. IU	Hum./Ed. SBU	Hum./Ed. I/SBU	Hum./Ed. NR	Sci. IU	Sci. SBU	Sci. I/SBU	Sci. NR	Bus. IU	Bus. SBU	Bus. I/SBU	Bus. NR	Mus. IU	Mus. SBU	Mus. I/SBU	Mus. NR	df	χ2
Imp. Teach.	1(1.12)	0	3(3.37)	0	0	2(2.25)	1(1.12)	0	2(2.25)	15(16.85)	13(14.61)	2(2.25)	0	5(5.62)	9(10.11)	0	1(1.12)	11(12.36)	6(6.74)	1(1.12)	2(2.25)	7(7.87)	3(3.37)	0	18	51.823*
Imp. P/M	1(1.12)	0	3(3.37)	0	0	0	2(2.25)	1(1.12)	0	22(24.72)	5(5.62)	5(5.62)	0	8(8.99)	5(5.62)	1(1.12)	1(1.12)	9(10.11)	1(1.12)	8(8.99)	2(2.25)	5(5.62)	3(3.37)	2(2.25)	18	50.397*
MAD	0	1(1.12)	0	3(3.37)	2(2.25)	0	1(1.12)	0	9(10.11)	5(5.62)	14(15.73)	4(4.49)	6(6.74)	0	4(4.49)	4(4.49)	6(6.74)	3(3.37)	4(4.49)	6(6.74)	5(5.62)	2(2.25)	4(4.49)	1(1.12)	18	33.268*
DK	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Administrative (Per Evaluation cent)

Use	Nur. IU	Nur. SBU	Nur. I/SBU	Nur. NR	Rel. IU	Rel. SBU	Rel. I/SBU	Rel. NR	Hum./Ed. IU	Hum./Ed. SBU	Hum./Ed. I/SBU	Hum./Ed. NR	Sci. IU	Sci. SBU	Sci. I/SBU	Sci. NR	Bus. IU	Bus. SBU	Bus. I/SBU	Bus. NR	Mus. IU	Mus. SBU	Mus. I/SBU	Mus. NR	df	χ2
Imp. Teach.	1(1.12)	0	2(2.25)	1(1.12)	0	1(1.12)	1(1.12)	1(1.12)	1(1.12)	19(20.22)	7(7.87)	6(6.74)	0	6(6.74)	6(6.74)	2(2.25)	0	8(8.99)	1(1.12)	10(11.24)	1(1.12)	5(5.62)	2(2.25)	4(4.49)	18	34.082*
Imp. P/M	1(1.12)	1(1.12)	2(2.25)	0	0	1(1.12)	1(1.12)	1(1.12)	1(1.12)	20(22.47)	4(4.49)	7(7.87)	0	7(7.87)	4(4.49)	3(3.37)	0	9(10.11)	0	10(11.24)	1(1.12)	4(4.49)	1(1.12)	6(6.74)	18	31.240*
MAD	1(1.120	1(1.12)	2(2.25)	0	1(1.12)	0	0	2(2.25)	7(7.87)	4(4.49)	19(21.35)	2(2.25)	2(2.25)	1(1.12)	4(4.49)	7(7.87)	3(3.37)	2(2.25)	4(4.49)	10(11.24)	2(2.25)	2(2.25)	3(3.37)	5(5.62)	18	32.437
DK	0	0	0	4(4.49)	0	0	1(1.12)	2(2.25)	2(2.25)	0	1(1.12)	29(32.58)	4(4.49)	0	0	10(11.24)	6(6.74)	0	1(1.12)	12(13.48)	5(5.62)					

Peer (Per Evaluation cent)

Use	Nur. IU	Nur. SBU	Nur. I/SBU	Nur. NR	Rel. IU	Rel. SBU	Rel. I/SBU	Rel. NR	Hum./Ed. IU	Hum./Ed. SBU	Hum./Ed. I/SBU	Hum./Ed. NR	Sci. IU	Sci. SBU	Sci. I/SBU	Sci. NR	Bus. IU	Bus. SBU	Bus. I/SBU	Bus. NR	Mus. IU	Mus. SBU	Mus. I/SBU	Mus. NR	df	χ2
Imp. Teach.	1(1.12)	0	3(3.37)	0	0	2(2.25)	0	1(1.12)	1(1.12)	19(20.22)	4(4.49)	9(10.11)	0	8(8.99)	0	6(6.74)	0	9(10.11)	0	10(11.24)	0	7(7.87)	0	5(5.62)	18	51.749
Imp. P/M	1(1.12)	0	3(3.37)	0	0	1(1.12)	0	2(2.25)	0	14(15.73)	4(4.49)	14(15.73)	0	5(5.62)	0	9(10.11)	0	8(8.99)	0	11(12.36)	0	6(6.74)	0	6(6.74)	18	58.752
MAD	1(1.12)	2(2.25)	1(1.12)	0	0	1(1.12)	0	2(2.25)	3(3.37)	7(7.87)	5(5.62)	17(19.10)	0	2(2.25)	0	12(13.48)	1(1.12)	4(4.49)	1(1.12)	13(14.61)	1(1.12)	3(3.37)	1(1.12)	7(7.87)	18	17.488
DK	1(1.12)	0	0	4(4.49)	1(1.12)	0	1(1.12)	1(1.12)	8(8.99)	1(1.12)	1(1.12)	22(24.72)	5(5.62)	0	2(2.25)	7(7.87)	7(7.87)	0	1(1.12)	11(12.36)	5(5.62)	1(1.12)	0	6(6.74)	18	15.705

Self (Per Evaluation cent)

Use	Nur. IU	Nur. SBU	Nur. I/SBU	Nur. NR	Rel. IU	Rel. SBU	Rel. I/SBU	Rel. NR	Hum./Ed. IU	Hum./Ed. SBU	Hum./Ed. I/SBU	Hum./Ed. NR	IU	Sc		
Imp. Teach.	1(1.12)	0	3(3.37)	0	0	1(1.12)	2(2.25)	0	4(4.49)	9(10.11)	18(20.22)	1(1.12)	0	4(4		
Imp. P/M	1(1.12)	0	2(2.25)	1(1.12)	0	0	1(1.12)	2(2.25)	1(1.12)	13(14.61)	10(11.24)	8(8.99)	0	5(5		
MAD	0	1(1.12)	2(2.25)	1(1.12)	0	0	0	3(3.37)	4(4.49)	4(4.49)	8(8.99)	16(17.98)	1(1.12)	1(1		
DK	0	0	0	4(4.49)	1(1.12)	0	0	2(2.25)	5(5.62)	0	0	27(30.34)	2(2.25)	1		

χ2
501*
119
.53
692

Note. - Nur.=Nursing; Rel.=Religion, Hum./Ed.=Humanities/Education, Sci.=The Sciences, Bus.=Business; Mus.=Music; Imp. Teach.=Improve Teaching, I/SBU=Is & Should Be Used, NR=No Response

Imp. P/M=Improve

Appendix E35. A Comparison of Instructor Responses, by School, to the Perceptions about the Disposition of Evaluation Information of the Four Types Practiced at Belmont University

Student (Per Evaluation cont)

Disp.	Nur. ID	Nur. SBD	Nur. I/SBD	Nur. NR	Rel. ID	Rel. SBD	Rel. I/SBD	Rel. NR	Hum./Ed. ID	Hum./Ed. SBD	Hum./Ed. I/SBD	Hum./Ed. NR	Sci. ID	Sci. SBD	Sci. I/SBD	Sci. NR	Bus. ID	Bus. SBD	Bus. I/SBD	Bus. NR	Mus. ID	Mus. SBD	Mus. I/SBD	Mus. NR	df	χ2
RINR	4(4.49)	0	0	0	3(3.37)	0	0	0	18(20.22)	1(1.12)	12(13.48)	1(1.12)	5(5.62)	3(3.37)	5(5.62)	1(1.12)	12(13.48)	0	7(7.87)	0	9(10.11)	0	3(1.12)	2(2.25)	18	69.616
RIFR	0	3(3.37)	0	1(1.12)	0	2(2.25)	0	1(1.12)	1(1.12)	13(14.61)	1(1.12)	17(19.10)	1(1.12)	5(5.62)	1(1.12)	7(7.87)	0	3(3.37)	0	16(17.98)	0	7(7.87)	2(2.25)	3(3.37)	18	21.059
NOR	0	0	0	4(4.49)	0	1(1.12)	0	2(2.25)	3(3.37)	0	0	29(32.58)	1(1.12)	0	0	13(14.61)	0	1(1.12)	0	18(20.22)	1(1.12)	0	0	11(12.36)	12	18.231
NOTH.	0	0	0	4(4.39)	0	0	0	3(3.37)	0	2(2.25)	0	30(33.71)	0	0	0	14(15.73)	0	1(1.12)	0	18(20.22)	0	1(1.12)	0	11(12.36)	6	1.891

Administrative (Per Evaluation cont)

Disp.	Nur. ID	Nur. SBD	Nur. I/SBD	Nur. NR	Rel. ID	Rel. SBD	Rel. I/SBD	Rel. NR	Hum./Ed. ID	Hum./Ed. SBD	Hum./Ed. I/SBD	Hum./Ed. NR	Sci. ID	Sci. SBD	Sci. I/SBD	Sci. NR	Bus. ID	Bus. SBD	Bus. I/SBD	Bus. NR	Mus. ID	Mus. SBD	Mus. I/SBD	Mus. NR	df	χ2
RINR	2(2.25)	0	0	2(2.25)	0	0	0	3(3.37)	11(12.36)	1(1.12)	11(12.36)	9(10.11)	2(2.25)	2(2.25)	0	10(11.24)	2(2.25)	2(2.25)	0	15(16.85)	1(1.12)	0	0	11(12.36)	18	43.187*
RIFR	1(1.12)	2(2.25)	0	1(1.12)	0	1(1.12)	1(1.12)	1(1.12)	1(1.12)	15(16.85)	3(3.37)	13(14.61)	0	3(3.37)	5(5.62)	6(6.74)	0	5(5.62)	0	14(15.73)	0	5(5.62)	1(1.12)	6(6.74)	16	30.031
NOR	0	1(1.12)	0	3(3.37)	0	0	0	3(3.37)	3(3.37)	0	0	29(32.58)	3(3.37)	0	2(2.25)	9(10.11)	6(6.74)	0	0	13(14.61)	4(4.49)	1(1.12)	0	7(7.87)	18	32.961*
NOTH.	0	0	0	4(4.49)	0	0	0	3(3.37)	1(1.12)	0	1(1.12)	30(33.71)	1(1.12)	0	0	13(14.61)	6(6.74)	0	0	13(14.61)	3(3.37)	1(1.12)	0	8(8.99)	16	21.214

Peer (Per Evaluation cont)

Disp.	Nur ID	Nur SBD	Nur I/SBD	Nur NR	Rel. ID	Rel. SBD	Rel. I/SBD	Rel. NR	Hum./Ed. ID	Hum./Ed. SBD	Hum./Ed. I/SBD	Hum./Ed. NR	Sci ID	Sci SBD	Sci I/SBD	Sci NR	Bus. ID	Bus. SBD	Bus. I/SBD	Bus. NR	Mus. ID	Mus. SBD	Mus. I/SBD	Mus. NR	df	χ2
RINR	1(1.12)	0	1(1.12)	2(2.25)	0	0	0	3(3.37)	3(3.37)	6(6.74)	4(4.49)	19(21.35)	1(1.12)	3(3.37)	0	10(11.24)	0	1(1.12)	0	18(20.22)	0	0	0	12(13.48)	18	24.127
RIFR	0	2(2.25)	0	2(2.25)	0	0	0	3(3.37)	1(1.12)	12(13.48)	0	19(21.35)	0	1(1.12)	0	13(14.61)	0	2(2.25)	0	17(19.10)	0	5(5.62)	0	7(7.87)	12	13.242
NOR	1(1.12)	0	0	3(3.37)	1(1.12)	0	0	2(2.25)	3(3.37)	1(1.12)	1(1.12)	27(30.34)	1(1.12)	0	0	13(14.61)	3(3.37)	0	0	16(17.98)	2(2.25)	0	0	10(11.24)	18	7.068
NOTH.	0	0	0	4(4.49)	0	0	0	3(3.37)	9(10.11)	0	1(1.12)	22(24.72)	3(3.37)	0	2(2.25)	9(10.11)	9(10.11)	0	1(1.12)	9(10.11)	6(6.74)	0	1(1.12)	5(5.62)	12	14.763

Self (Per Evaluation cont)

Disp.	Nur. ID	Nur. SBD	Nur. I/SBD	Nur. NR	Rel. ID	Rel. SBD	Rel. I/SBD	Rel. NR	Hum./Ed. ID	Hum./Ed. SBD	Hum./Ed. I/SBD	Hum./Ed. NR	Sci. ID	Sci. SBD	Sci. I/SBD	Sci. NR	Bus. ID	Bus. SBD	Bus. I/SBD	Bus. NR	Mus. ID	Mus. SBD	Mus. I/SBD	Mus. NR	df	χ2
RINR	3(3.37)	0	1(1.12)	0	0	0	0	3(3.37)	9(10.11)	1(1.12)	4(4.49)	18(20.22)	1(1.12)	1(1.12)	1(1.12)	11(12.36)	0	0	0	19(21.35)	1(1.12)	0	0	11(12.36)	18	33.323
RIFR	0	2(2.25)	0	2(2.25)	0	1(1.12)	0	2(2.25)	1(1.12)	11(12.36)	2(2.25)	18(20.22)	3(3.37)	5(5.62)	5(5.62)	6(6.74)	0	1(1.12)	0	18(20.22)	0	4(4.49)	2(2.25)	6(6.74)	18	25.230
NOR	0	0	0	4(4.49)	1(1.12)	0	0	2(2.25)	4(4.49)	1(1.12)	1(1.12)	26(29.21)	3(3.37)	0	2(2.25)	9(10.11)	1(1.12)	0	1(1.12)	17(19.10)	4(4.49)	0	0	8(8.99)	18	14.038
NOTH.	0	0	0	4(4.49)	0	0	0	3(3.37)	5(5.62)	0	2(2.25)	25(28.09)	3(3.37)	1(1.12)	0	10(11.24)	8(8.99)	0	2(2.25)	9(10.11)	3(3.37)	0	1(1.12)	8(8.99)	18	17.902

Note: Nur.=Nursing; Rel.=Religion; Hum./Ed.=Humanities/Education; Sci.=The Sciences; Bus.=Business; Mus.=Music; Disp.=Disposition; ID=Is Done SBD=Should Be Done; I/SBD=Is & Should Be Done; NR=No Response
RINR=Return to Instructor/No Response Expected; RIFR=Return to Instructor for Response; NOR=Not Returned; Noth.=Nothing

Appendix E36. A Comparison of Instructor Responses, by School, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching (Percent)

	Nur.					Rel.					Hum./Ed.					Sci.					Bus.					Mus.						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	1(1.12)	1(1.12)	1(1.12)	1(1.12)	0	0	0	3(3.37)	0	0	2(2.25)	13(14.61)	14(15.73)	3(3.37)	0	1(1.12)	5(5.62)	7(7.87)	1(1.12)	0	1(1.12)	11(12.36)	5(5.62)	2(2.25)	0	0	5(5.62)	6(6.74)	1(1.12)	0	24	101.454
Adm.	0	2(2.25)	2(2.25)	0	0	0	2(2.25)	0	0	1(1.12)	1(1.12)	12(13.48)	11(12.36)	8(8.99)	0	0	8(8.99)	4(4.49)	0	2(2.25)	0	4(4.49)	5(5.62)	6(6.74)	4(4.49)	0	6(6.74)	3(3.37)	1(1.12)	2(2.25)	24	48.933
Peer	1(1.12)	0	2(2.25)	1(1.12)	0	0	2(2.25)	0	0	1(1.12)	3(3.37)	4(4.49)	10(11.24)	4(4.49)	11(12.36	0	2(2.25)	3(3.37)	3(3.37)	6(6.74)	0	3(3.37)	2(2.25)	6(6.74)	8(8.99)	0	3(3.37)	1(1.12)	3(3.37)	5(5.62)	24	32.604
Self	2(2.25)	2(2.25)	0	0	0	0	2(2.25)	0	0	1(1.12)	3(3.37)	17(19.10)	11(12.36)	1(1.12)	0	1(1.12)	9(10.11)	4(4.49)	0	0	1(1.12)	5(5.62)	3(3.37)	4(4.49)	6(6.74)	0	6(6.74)	3(3.37)	2(2.25)	1(1.12)	24	66.917*

Improving Program/Major (Percent)

	Nur.					Rel.					Hum./Ed.					Sci.					Bus.					Mus.						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	1(1.12)	1(1.12)	2(2.25)	0	0	0	1(1.12)	2(2.25)	0	0	1(1.12)	8(8.99)	13(14.61)	10(11.24)	0	0	5(5.62)	8(8.99)	1(1.12)	0	0	6(6.74)	8(10.11)	3(3.37)	1(1.12)	0	6(6.74)	6(6.74)	0	0	24	95.605*
Adm.	0	3(3.37)	1(1.12)	0	0	0	2(2.25)	0	0	1(1.12)	0	12(13.48)	9(10.11)	1(1.12)	0	1(1.12)	5(5.62)	7(7.87)	0	1(1.12)	0	3(3.37)	4(4.49)	8(8.99)	4(4.49)	0	5(5.62)	3(3.37)	2(2.25)	2(2.25)	24	60.181*
Peer	1(1.12)	2(2.25)	1(1.12)	0	0	0	2(2.25)	0	0	1(1.12)	1(1.12)	11(12.36)	11(12.36)	9(10.11)	0	1(1.12)	7(7.87)	6(6.74)	0	0	1(1.12)	3(3.37)	4(4.49)	4(4.49)	7(7.87)	0	6(6.74)	4(4.49)	1(1.12)	1(1.12)	24	60.446*
Self	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Making Administrative Decisions (Percent)

	Nur.					Rel.					Hum./Ed.					Sci.					Bus.					Mus.						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	0	1(1.12)	2(2.25)	1(1.12)	0	0	2(2.25)	2(2.25)	0	0	0	11(12.36)	11(12.36)	8(8.99)	2(2.25)	0	4(4.49)	6(6.74)	3(3.37)	1(1.12)	0	7(7.87)	6(6.74)	3(3.37)	3(3.37)	0	4(4.49)	5(5.62)	1(1.12)	2(2.25)	18	36.454*
Adm.	0	3(3.37)	1(1.12)	0	0	0	2(2.25)	0	0	1(1.12)	1(1.12)	15(16.85)	8(8.99)	6(6.74)	2(2.25)	0	8(8.99)	3(3.37)	1(1.12)	2(2.25)	0	6(6.74)	4(4.49)	3(3.37)	6(6.74)	0	6(6.74)	2(2.25)	1(1.12)	3(3.37)	24	32.058
Peer	0	1(1.12)	2(2.25)	1(1.12)	0	0	2(2.25)	0	0	1(1.12)	1(1.12)	8(8.99)	6(6.74)	6(6.74)	11(12.36	0	3(3.37)	4(4.49)	1(1.12)	6(6.74)	0	4(4.49)	3(3.37)	3(3.37)	9(10.11)	0	2(2.25)	1(1.12)	3(3.37)	6(6.74)	24	20.450
Self	0	4(4.49)	0	0	0	0	2(2.25)	0	0	1(1.12)	1(1.12)	9(10.11)	13(14.61)	7(7.87)	2(2.25)	0	7(7.87)	3(3.37)	1(1.12)	3(3.37)	0	2(2.25)	4(4.49)	4(4.49)	9(10.11)	0	4(4.49)	2(2.25)	4(4.49)	2(2.25)	24	47.825*

--=Too Few Responses in Cells to Allow for Chi Square Calculations

Note: Stu.=Student; Adm.=Administrative; Nur.=Nursing; Rel.=Religion; Hum./Ed.=Humanities/Education; Sci.=The Sciences; Bus.=Business; Mus.=Music; VE=Very Effective; E=Effective; IE=Ineffective; VIE=Very Ineffective; NR=No Response

Appendix E37. A Comparison of Instructor Responses, by Academic Degree, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fairness (Per cent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc. Can.					
Type	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	VF	F	U	VUF	NR	df	χ2
Stu.	2(2.25)	31(40.45)	20(22.47)	3(3.37)	0	1(1.12)	12(13.48)	4(4.49)	0	5(5.62)	0	1(1.12)	0	0	0	0	1(1.12)	0	0	0	0	3(3.37)	1(1.12)	0	0	16	19.527
Adm.	8(8.99)	36(40.45)	8(8.99)	2(2.25)	7(7.87)	1(1.12)	12(13.48)	3(3.37)	0	6(6.74)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	0	3(3.37)	1(1.12)	0	0	16	12.407
Peer	4(4.49)	23(25.84)	4(4.49)	1(1.12)	29(32.58)	2(2.25)	7(7.87)	2(2.25)	0	11(12.36)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	6(6.74)	34(38.20)	6(6.74)	1(1.12)	42(47.19)	16	6.181
Self	6(6.74)	42(47.19)	3(3.37)	4(4.49)	6(6.74)	3(3.37)	8(8.99)	4(4.49)	0	7(7.87)	0	1(1.12)	0	0	0	0	0	1(1.12)	0	0	0	4(4.49)	0	0	0	16	24.456

Objectivity (Per cent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc. Can.					
Type	VO	O	U	VUO	NR	VO	O	U	VUO	NR	VO	O	U	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	χ2
Stu.	2(2.25)	19(21.35)	33(37.08)	7(7.87)	0	0	8(8.99)	9(10.11)	0	5(5.62)	0	1(1.12)	0	0	0	0	0	1(1.12)	0	0	0	2(2.25)	0	2(2.25)	0	16	30.739
Adm.	4(4.49)	27(30.34)	19(21.35)	4(4.49)	7(7.87)	1(1.12)	12(13.48)	3(3.37)	0	6(6.74)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	3(3.37)	1(1.12)	0	0	16	14.641
Peer	3(3.37)	15(16.85)	13(14.61)	1(1.12)	29(32.58)	1(1.12)	5(5.62)	5(5.62)	0	11(12.36)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	3(3.37)	0	0	1(1.12)	16	9.533
Self	3(3.37)	26(29.21)	20(22.47)	4(4.49)	8(8.99)	1(1.12)	8(8.99)	3(3.37)	1(1.12)	9(10.11)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	3(3.37)	0	1(1.12)	0	16	19.242

Reliability (Per cent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc. Can.					
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	χ2
Stu.	0	21(23.60)	29(32.58)	10(11.24)	1(1.12)	0	9(10.11)	8(8.99)	0	5(5.62)	0	1(1.12)	0	0	0	0	1(1.12)	0	0	0	0	2(2.25)	1(1.12)	1(1.12)	0	12	19.161
Adm.	4(4.49)	32(35.96)	14(15.73)	4(4.49)	7(7.87)	1(1.12)	12(13.48)	3(3.37)	0	6(6.74)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	2(2.25)	1(1.12)	1(1.12)	0	16	14.579
Peer	2(2.25)	22(24.72)	4(4.49)	5(5.62)	29(31.46)	2(2.25)	6(6.74)	3(3.37)	0	11(12.36)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	2(2.25)	1(1.12)	0	1(1.12)	16	9.821
Self	4(4.49)	38(42.70)	9(10.11)	2(2.25)	8(8.99)	1(1.12)	9(10.11)	3(3.37)	1(1.12)	8(8.99)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	4(4.49)	0	0	0	16	13.987

Note. Stu.=Student; Adm.=Administrative; VF=Very Fair; F=Fair; U=Unfair; VUF=Very Unfair; VO=Very Objective; O=Objective; U=Unobjective; VUO=Very Unobjective VR=Very Reliable; R=Reliable; UR=Unreliable; VUR=Very Unreliable; NR=No Response

Appendix E38. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by Academic Degree, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at

(Per cent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc. Can.						
Type	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
Stu.	1(1.12)	20(22.47)	16(29.21)	14(15.73)	0	0	9(10.11)	7(7.87)	1(1.12)	5(5.62)	0	1(1.12)	0	0	0	0	1(1.12)	0	0	0	0	2(2.25)	1(1.12)	1(1.12)	0	16	23.402
Adm.	5(5.62)	29(32.58)	15(16.85)	9(10.11)	3(3.37)	0	12(13.48)	5(5.62)	0	5(5.62)	0	1(1.12)	0	0	0	0	0	0	0	1(1.12)	0	1(1.12)	1(1.12)	1(1.12)	1(1.12)	16	21.068
Peer	2(2.25)	21(23.60)	6(6.74)	9(10.11)	23(25.84)	1(1.12)	6(6.74)	3(3.37)	1(1.12)	11(12.36)	0	1(1.12)	0	0	0	0	0	0	0	1(1.12)	0	1(1.12)	2(2.25)	0	1(1.12)	16	11.864
Self	8(8.99)	31(34.83)	12(13.48)	4(4.49)	6(6.74)	1(1.12)	10(11.24)	2(2.25)	1(1.12)	8(8.99)	0	1(1.12)	0	0	0	0	0	0	0	1(1.12)	0	3(3.37)	1(1.12)	0	0	16	17.143

A Comparison of Instructor Responses, by Academic Degree, to the Perceived Overall Satisfaction with Evaluation Practiced at Belmont

(Per cent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc.Can.						
	VS	S	U	VUS	NR	VS	S	U	VSU	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
	1(1.12)	30(33.71)	23(25.84)	7(7.87)	0	0	8(8.99)	10(11.24)	1(1.12)	3(3.37)	0	1(1.12)	0	0	0	0	0	1(1.12)	0	0	0	2(2.25)	1(1.12)	1(1.12)	0	16	14.931

Note. Stu.=Student; Adm.=Administrative; VS=Very Satisfied; S=Satisfied; U=Unsatisfied; VUS=Very Unsatisfied; NR=No Response

Appendix E39. A Comparison of Instructor Responses, by Academic Degree, about Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

(Per cent)

Type	Doctors			Masters			Bachelors			2yr./Assoc.			ABD/Doc.Can.			df	X2
	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR		
Stu.	52(58.43	8(8.99)	1(1.12)	16(17.98	2(2.25)	4(4.49)	1(1.12)	0	0	1(1.12)	0	0	4(4.49)	0	0	8	9.718
Adm.	53(59.55	7(7.87)	1(1.12)	18(20.22	1(1.12)	3(3.37)	1(1.12)	0	0	0	0	1(1.12)	3(3.37)	0	1(1.12)	8	21.273
Peer	44(49.44	15(16.85	2(2.25)	13(14.61	3(3.37)	6(6.74)	1(1.12)	0	0	0	1(1.12)	0	4(4.49)	0	0	8	17.576
Self	51(57.30	8(8.99)	2(2.25)	16(17.98	0	6(6.74)	1(1.12)	0	0	0	1(1.12)	0	4(4.49)	0	0	8	23.470

Note. Stu.=Student; Adm.=Administrative; S=Should Be Done; SN=Should Not Be Done; NR=No Response

Appendix E40. A Comparison of Instructor Responses, by Academic Degree, to the Perceptions of the Usage of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Per cent)

	Doctors				Masters				Bachelors				2yr./Assoc.				ABD/Doc.Can.						
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	5(5.62)	29(32.58)	25(28.09)	2(2.25)	0	10(11.24)	7(7.87)	5(5.62)	0	0	1(1.12)	0	1(1.12)	0	0	0	0	2(2.25)	2(2.25)	0	12	25.963
Imp. P/M	3(3.37)	30(33.71)	15(16.85)	13(14.61)	0	12(13.48)	2(2.25)	8(8.99)	0	0	1(1.12)	0	1(1.12)	0	0	0	0	2(2.25)	1(1.12)	1(1.12)	12	29.611
MAD	22(24.72)	6(6.74)	21(23.60)	12(13.48)	4(4.49)	5(5.62)	4(4.49)	9(10.11)	0	0	1(1.12)	0	1(1.12)	0	0	0	1(1.12)	0	1(1.12)	2(2.25)	12	14.074
DK	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Administrative Evaluation (Per cent)

	Doctors				Masters				Bachelors				2yr./Assoc.				ABD/Doc.Can.					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	3(3.37)	25(28.09)	14(15.73)	19(21.35)	0	11(12.36)	3(3.37)	8(8.99)	0	1(1.12)	0	0	0	0	0	1(1.12)	0	1(1.12)	2(2.25)	1(1.12)	12	7.761
Imp. P/M	2(2.25)	27(30.34)	11(12.36)	21(23.60)	0	12(13.48)	1(1.12)	9(10.11)	0	1(1.12)	0	0	0	0	0	1(1.12)	1(1.12)	2(2.25)	1(1.12)	0	12	13.594
MAD	12(13.48)	5(5.62)	22(24.72)	22(24.72)	3(3.37)	3(3.37)	8(8.99)	8(8.99)	0	1(1.12)	0	0	0	0	0	1(1.12)	1(1.12)	1(1.12)	2(2.25)	0	12	13.153
DK	13(14.61)	0	3(3.37)	45(50.56)	2(2.25)	0	0	20(22.47)	1(1.12)	0	0	0	1(1.12)	0	0	0	0	0	0	4(4.49)	8	12.734

Peer Evaluation (Per cent)

	Doctors				Masters				Bachelors				2yr./Assoc.				ABD/Doc.Can.					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBD	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	2(2.25)	34(38.20)	3(3.37)	22(24.72)	0	8(8.99)	3(3.37)	11(12.36)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	2(2.25)	1(1.12)	1(1.12)	12	9.189
Imp. P/M	1(1.12)	26(29.21)	3(3.37)	31(34.83)	0	7(7.87)	3(3.37)	12(13.48)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	1(1.12)	1(1.12)	2(2.25)	12	6.116
MAD	4(4.49)	14(15.73)	6(6.74)	37(41.57)	1(1.12)	4(4.49)	2(2.25)	15(16.85)	0	0	0	1(1.12)	0	0	0	1(1.12)	1(1.12)	1(1.12)	0	2(2.25)	12	4.100
DK	21(23.60)	1(1.12)	3(3.37)	36(40.45)	6(6.74)	1(1.12)	1(1.12)	14(15.73)	0	0	1(1.12)	0	0	0	0	1(1.12)	0	0	0	4(4.49)	12	21.099

Self Evaluation (Per cent)

	Doctors				Masters				Bachelors				2yr./Assoc				ABD/Doc.Can.					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	6(6.74)	18(20.22)	27(30.34)	10(11.24)	2(2.25)	6(6.74)	9(10.11)	5(5.62)	0	0	0	1(1.12)	1(1.12)	0	0	0	0	1(1.12)	3(3.37)	0	12	15.959
Imp. P/M	4(4.49)	19(21.35)	14(15.73)	24(26.97)	2(2.25)	6(6.74)	7(7.87)	7(7.87)	0	0	0	1(1.12)	1(1.12)	0	0	0	0	1(1.12)	1(1.12)	2(2.25)	12	14.879
MAD	6(6.74)	8(8.99)	13(14.61)	34(38.20)	3(3.37)	3(3.37)	4(4.49)	12(13.48)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	1(1.12)	1(1.12)	2(2.25)	12	2.805
DK	11(12.36)	0	0	50(56.18)	4(4.49)	0	1(1.12)	17(19.10)	0	0	1(1.12)	0	0	0	0	1(1.12)	0	0	0	4(4.49)	8	46.686

Note: Imp. Tch.=Improve Teaching; Imp. P/M=Improve Program/Major; MAD=Make Administrative Decisions; DK=Don't Know; IU=Is Used, SBU=Should Be Used; I/SBU=Is & Should Be Used; NR=No Response; 2yr./Assoc.=Two Year/Associates Degree; ABD/Doc.Can.=All But Dissertation/Doctoral Candidate

Appendix E41. A Comparison of Instructor Responses, by Academic Degree, to the Perceptions of the Disposition of Evaluation Information from the Four Types of Evaluation Practiced at Belmont University

Student (Per Evaluation cont)

	Doctors					Masters				Bachelors				2yr./Assoc.				ABD/Doc.Can.			
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	37(41.57)	4(4.49)	18(20.22)	2(2.25)	8(8.99)	0	7(7.87)	7(7.87)	1(1.12)	0	0	0	1(1.12)	0	0	0	4(4.49)	0	0	0	12	21.060
RIFR	2(2.25)	23(25.84)	3(3.37)	33(37.08)	0	8(8.99)	1(1.12)	13(14.61)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	3(3.37)	0	1(1.12)	12	4.883
NOR	5(5.62)	2(2.25)	0	54(60.67)	0	0	0	22(24.72)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	0	0	4(4.49)	8	3.487
NOTH.	0	4(4.49)	0	57(64.04)	0	0	0	22(24.72)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	0	0	4(4.49)	4	1.922

Administrative (Per Evaluation cont)

	Doctors					Masters				Bachelors				2yr/Assoc.				ABD/Doc.Can.			
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	13(14.61)	3(3.37)	8(8.99)	37(41.57)	3(3.37)	2(2.25)	3(3.37)	14(15.73)	0	0	0	1(1.12)	0	0	0	1(1.12)	2(2.25)	0	0	2(2.25)	12	4.914
RIFR	2(2.25)	20(22.47)	9(10.11)	30(33.71)	0	10(11.24)	0	12(13.48)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	2(2.25)	1(1.12)	1(1.12)	12	8.198
NOR	12(13.48)	1(1.12)	2(2.25)	46(51.69)	4(4.49)	0	0	18(20.22)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	1(1.12)	0	3(3.37)	12	12.170
NOTH.	7(7.87)	1(1.12)	1(1.12)	52(58.43)	4(4.49)	0	0	18(20.22)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	0	0	4(4.49)	12	2.485

Peer (Per Evaluation cont)

	Doctors					Masters				Bachelors				2yr./Assoc.				ABD/Doc.Can.			
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	3(3.37)	9(10.11)	3(3.37)	46(51.69)	1(1.12)	1(1.12)	2(2.25)	18(20.22)	0	0	0	1(1.12)	0	0	0	1(1.12)	1(1.12)	0	0	3(3.37)	12	6.192
RIFR	1(1.12)	13(14.61)	0	47(52.81)	0	7(7.87)	0	15(16.85)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	3(3.37)	0	1(1.12)	8	7.173
NOR	8(8.99)	0	1(1.12)	52(58.43)	2(2.25)	1(1.12)	0	19(21.35)	0	0	0	1(1.12)	0	0	0	1(1.12)	1(1.12)	0	0	3(3.37)	12	4.590
NOTH.	20(22.47)	0	5(5.62)	36(40.45)	5(5.62)	0	0	17(19.10)	1(1.12)	0	0	0	0	0	0	0	1(1.12)	1(1.12)	0	3(3.37)	8	6.535

Self (Per Evaluation cont)

	Doctors					Masters				Bachelors				2yr./Assoc.				ABD/Doc.Can			
Disp.	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	8(8.99)	2(2.25)	4(4.49)	47(52.81)	3(3.37)	0	2(2.25)	17(19.10)	0	0	0	1(1.12)	0	0	0	1(1.12)	3(3.37)	0	0	1(1.12)	12	12.622
RIFR	1(1.12)	16(17.98)	8(8.99)	36(40.45)	0	6(6.74)	1(1.12)	15(16.85)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	1(1.12)	0	3(3.37)	12	3.597
NOR	11(12.36)	1(1.12)	3(3.37)	46(51.69)	2(2.25)	0	1(1.12)	19(21.35)	0	0	0	1(1.12)	0	0	0	1(1.12)	0	0	0	4(4.49)	12	3.151
NOTH.	13(14.61)	1(1.12)	4(4.49)	43(48.31)	5(5.62)	0	1(1.12)	16(17.58)	1(1.12)	0	0	0	0	0	0	1(1.12)	0	0	0	4(4.49)	12	6.209

Note. Disp.=Disposition; RINR=Return to Instructor/No Response Requested; RIFR=Return to Instructor for Response; NOR=Not Returned; NOTH.=Nothing; 2yr./Assoc.=Two Year/Associates Degree; ABD/Doc.Can.=All But Dissertation/Doctoral Candidate; ID=Is Done; SBD=Should Be Done; I/SBD=Is &Should Be Done; NR=No Response

Appendix E42. A Comparison of Instructor Responses, by Academic Degree, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching (Percent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc.Can.						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	2(2.25)	21(23.60)	30(33.71)	8(8.99)	0	2(2.25)	10(11.24)	5(5.62)	0	5(5.62)	0	1(1.12)	0	0	0	0	1(1.12)	0	0	0	1(1.12)	2(2.25)	1(1.12)	0	0	16	29.226
Adm.	0	24(26.97)	19(21.35)	11(12.36)	7(7.87)	1(1.12)	7(7.87)	5(5.62)	3(3.37)	6(6.74)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	2(2.25	1(1.12)	1(1.12)	0	16	14.378
Peer	2(2.25)	9(10.11)	15(16.85)	14(15.73)	21(23.60)	2(2.25)	2(2.25)	3(3.37)	2(2.25)	13(14.61)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	2(2.25	0	1(1.12)	1(1.12)	16	18.113
Self	4(4.49)	29(32.58)	19(21.35)	5(5.62)	4(4.49)	3(3.37)	7(7.87)	2(2.25)	2(2.25)	8(8.99)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	4(4.49	0	0	0	16	27.084

Improving Program/Major (Percent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc.Can.						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	1(1.12)	20(22.47)	28(31.46)	12(13.48)	0	0	5(5.62)	11(12.36)	0	6(6.74)	0	1(1.12)	0	0	0	0	0	1(1.12)	0	0	1(1.12)	1(1.12)	0	2(2.25	0	16	41.032*
Adm.	1(1.12)	20(22.47)	19(21.35)	15(16.85)	6(6.74)	0	6(6.74)	5(5.62)	5(5.62)	6(6.74)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	3(3.37)	0	1(1.12	0	16	16.166
Peer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Self	2(2.25)	22(24.92)	19(21.35)	13(14.61)	5(5.62)	2(2.25)	5(5.62)	6(6.74)	1(1.12)	8(8.99)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	3(3.37	1(1.12)	0	0	16	23.950

Making Administrative Decisions (Percent)

	Doctors					Masters					Bachelors					2yr./Assoc.					ABD/Doc.Can.						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	0	19(21.35)	25(28.09)	14(15.73)	3(3.37)	0	7(7.87)	5(5.62)	2(2.25)	8(8.99)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	1(1.12)	2(2.25	0	1(1.12)	12	23.670
Adm.	0	30(33.71)	13(14.61)	9(10.11)	9(10.11)	1(1.12)	7(7.87)	4(4.49)	2(2.25)	8(8.99)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	2(2.25	1(1.12)	0	1(1.12)	16	13.726
Peer	0	16(17.98)	12(13.48)	11(12.36)	22(24.72)	1(1.12)	2(2.25)	4(4.49)	2(2.25)	13(14.61)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	1(1.12)	0	1(1.12)	2(2.25	16	13.883
Self	0	21(23.60)	18(20.22)	12(13.48)	10(11.24)	1(1.12)	5(5.62)	3(3.37)	3(3.37)	10(11.24)	0	1(1.12)	0	0	0	0	0	0	1(1.12)	0	1(1.12)	1(1.12)	1(1.12)	1(1.12)	16	16.638

Note: Stu.=Student; Adm.=Administrateave; VE=Very Effective; E=Effective; IE=Ineffective; VIE=Very Ineffective; NR=No Response

2yr./Assoc.=Two Year/Associates Degree; ABD/Doc.Can.=All But Dissertation/Doctoral Candidate;

Appendix E43. A Comparison of Instructor Responses, by Years of Teaching Experience at Belmont University, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fair (Percentages)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	df	χ2
Stu.	0	18(20.22)	7(7.87)	2(2.25)	2(2.25)	2(2.25)	15(16.85)	10(11.24)	0	0	1(1.12)	6(6.74)	1(1.12)	1(1.12)	0	0	6(6.74)	4(4.49)	0	1(1.12)	0	8(8.99)	3(3.37)	0	0	20	30.604
Adm.	2(2.25)	21(23.60)	4(4.49)	0	2(2.25)	1(1.12)	15(16.85)	5(5.62)	0	6(6.74)	1(1.12)	4(4.49)	1(1.12)	1(1.12)	2(2.25)	0	7(7.87)	2(2.25)	0	2(2.25)	5(5.62)	5(5.62)	0	1(1.12)	1(1.12)	20	33.464
Peer	1(1.12)	18(20.22)	1(1.12)	0	9(10.11)	3(3.37)	7(7.87)	2(2.25)	1(1.12)	14(15.73)	0	2(2.25)	1(1.12)	0	6(6.74)	0	4(4.49)	2(2.25)	0	5(5.62)	2(2.25)	3(3.37)	0	0	7(7.87)	20	20.743
Self	2(2.25)	21(23.60)	1(1.12)	1(1.12)	4(4.49)	2(2.25)	14(15.73)	3(3.37)	3(3.37)	5(5.62)	2(2.25)	5(5.62)	2(2.25)	0	0	0	7(7.87)	2(2.25)	0	2(2.25)	3(3.37)	8(8.99)	0	0	1(1.12)	20	23.617

Objectivity (Percentages)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	χ2
Stu.	0	8(8.99)	14(15.73)	5(5.62)	2(2.25)	0	11(12.36)	14(15.73)	2(2.25)	0	1(1.12)	3(3.37)	4(4.49)	1(1.12)	0	0	2(2.25	7(7.87	1(1.12	1(1.12	1(1.12	1(1.12	6(6.74	4(4.49	0	1(1.12	20	32.458
Adm.	0	16(17.98)	10(11.24)	1(1.12)	2(2.25)	0	13(14.61)	8(8.99)	0	6(6.74)	1(1.12)	4(4.49)	1(1.12)	1(1.12)	2(2.25)	0	6(6.74	5(5.62	0	2(2.25	4(4.49	4(4.49	1(1.12)	2(2.25	1(1.12	20	39.618
Peer	1(1.12)	10(11.24)	9(10.11)	0	9(10.11)	1(1.12)	7(7.87)	4(4.49)	1(1.12)	14(15.73)	0	2(2.25)	1(1.12)	0	6(6.74)	0	2(2.25	4(4.49	0	5(5.62	2(2.25	3(3.37	0	0	7(7.87	20	18.447
Self	0	15(16.85)	8(8.99)	2(2.25)	4(4.49)	1(1.12)	7(7.87)	8(8.99)	2(2.25)	9(10.11)	1(1.12)	393.37)	3(3.37)	1(1.12)	1(1.12)	0	6(6.74	3(3.37	0	2(2.25	2(2.25	7(7.87	1(1.12	1(1.12	1(1.12	20	21.169

Reliability (Percentages)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	χ2
Stu.	0	10(11.24)	14(15.73)	3(3.37)	2(2.25)	0	9(10.11)	13(14.61)	4(4.49)	1(1.12)	0	4(4.49)	4(4.49)	1(1.12)	0	0	3(3.37	5(5.62	292.25	1(1.12	0	7(7.87	3(3.37	1(1.12	1(1.12	15	18.901
Adm.	1(1.12)	16(17.98)	8(8.99)	2(2.25)	2(2.25)	1(1.12)	14(15.73)	6(6.74)	0	6(6.74)	0	3(3.37)	2(2.25)	2(2.25)	2(2.25)	0	8(8.99	1(1.12	0	2(2.25	3(3.37	6(6.74	1(1.12	1(1.12	1(1.12	20	28.467
Peer	0	15(16.85)	5(5.62)	1(1.12)	8(8.99)	2(2.25)	7(7.87)	2(2.25)	2(2.25)	14(15.73)	0	2(2.25)	0	1(1.12)	6(6.74)	0	4(4.49	1(1.12	1(1.12	5(5.62	2(2.25)	3(3.37	0	0	7(7.87	20	21.099
Self	1(1.12)	20(22.47)	4(4.49)	0	4(4.49)	1(1.12)	10(11.24)	6(6.74)	2(2.25)	8(8.99)	0	5(5.62)	2(2.25)	1(1.12)	1(1.12)	0	9(10.11	0	0	2(2.25	3(3.37	8(8.99	0	0	1(1.12	20	30.479

Note: Stu.=Student; Adm.=Administrative; VF=Very Fair; F=Fair; UF=Unfair; VUF=Very Unfair; NR=No Response; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR=Unreliable; VUR=Very Unreliable

Appendix E44. Instructor Perceptions of Satisfaction with The Belmont University Evaluation System

A Comparison of Instructor Responses, by Years Teaching Experience at Belmont, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	df	X2
Stu.	0	11(12.36)	9(10.11)	7(7.87)	2(2.25)	0	10(11.24)	13(14.61)	4(4.49)	0	1(1.12)	3(3.37)	4(4.49)	1(1.12)	0	0	4(4.49)	4(4.49)	2(2.25)	1(1.12)	0	5(5.62)	4(4.49)	2(2.25)	1(1.12)	20	30.563
Adm.	1(1.12)	12(13.48)	9(10.11)	5(5.62)	2(2.25)	1(1.12)	14(15.73)	6(6.74)	3(3.37)	3(3.37)	0	5(5.62)	2(2.25)	1(1.12)	1(1.12)	0	6(6.74)	2(2.25)	1(1.12)	2(2.25)	3(3.37)	6(6.74)	2(2.25)	0	1(1.12)	20	22.482
Peer	0	14(15.73)	6(6.74)	3(3.37)	6(6.74)	2(2.25)	8(8.99)	1(1.12)	3(3.37)	13(14.61)	0	1(1.12)	2(2.25)	2(2.25)	4(4.49)	0	3(3.37)	2(2.25)	1(1.12)	5(5.62)	1(1.12)	3(3.37)	0	1(1.12)	7(7.87)	20	20.278
Self	2(2.25)	19(21.35)	5(5.62)	1(1.12)	2(2.25)	4(4.49)	8(8.99)	5(5.62)	2(2.25)	8(8.99)	1(1.12)	4(4.49)	2(2.25)	1(1.12)	1(1.12)	0	8(8.99)	1(1.12)	0	2(2.25)	2(2.25)	6(6.74)	2(2.25)	1(1.12)	1(1.12)	20	19.486

A Comparison of Instructor Responses, by Years Teaching Experience at Belmont, to the Perceived Overall Satisfaction with the Evaluation System at Belmont University

1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
VS	S	US	VUS	NR	VS	S	US	VSU	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	df	X2
0	15(16.85)	10(11.24)	3(3.37)	1(1.12)	1(1.12)	15(16.98)	9(10.11)	2(2.25)	0	0	2(2.25)	5(5.62)	2(2.25)	0	0	4(4.49)	5(5.62)	1(1.12)	1(1.12)	0	5(5.62)	5(5.62)	1(1.12)	1(1.12)	20	11.797

Note. Stu.=Student; Adm.=Administration; VS=Very Satisfied; S=Satisfied; US=Unsatisfied; VUS=Very Unsatisfied; NR=No Response

Appendix E45. A Comparison of Instructor Responses, by Years Teaching Experience at Belmont University, to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

(Percent)

Type	1-5 Years			6-10 Years			11-15 Years			16-20 Years			20+ Years			df	χ2
	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR		
Stu.	25(28.09	2(2.25)	2(2.25)	22(24.72	4(4.49)	1(1.12)	8(8.99)	1(1.12)	0	8(8.99)	2(2.25)	1(1.12)	10(11.24)	1(1.12)	1(1.12)	10	2.940
Adm.	24(26.97	3(3.37)	2(2.25)	22(24.72	4(4.49)	1(1.12)	8(8.99)	0	1(1.12)	9(10.11)	1(1.12)	1(1.12)	11(12.36)	0	1(1.12)	10	4.055
Peer	24(26.97	3(3.37)	2(2.25)	17(19.10	7(7.87)	3(3.37)	7(7.87)	2(2.25)	0	5(5.62)	5(5.62)	1(1.12)	8(8.99)	2(2.25)	2(2.25)	10	9.252
Self	25(28.09	1(1.12)	3(3.37)	20(22.47	4(4.49)	3(3.37)	7(7.87)	2(2.25)	0	9(10.11)	1(1.12)	1(1.12)	10(11.24)	1(1.12)	1(1.12)	10	4.741

Note. Stu.=Student; Adm.=Administrative; S=Should Be Done; SN=Should Not Be Done; NR=No Response

Appendix E46. A Comparison of Instructor Responses, by Years Teaching Experience at Belmont University, to the Perceived Usage of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	χ2
Imp. Tch.	1(1.12)	15(16.85)	12(13.48)	1(1.12)	1(1.12)	14(15.73)	10(11.24)	2(2.25)	1(1.12)	4(4.49)	4(4.49)	0	3(3.37)	5(5.62)	1(1.12)	2(2.25)	0	3(3.37)	8(8.99)	1(1.12)	15	29.943
Imp. P/M	1(1.12)	14(15.73)	6(6.74)	8(8.99)	0	13(14.61)	7(7.87)	7(7.87)	1(1.12)	6(6.74)	2(2.25)	0	2(2.25)	3(3.37)	2(2.25)	4(4.49)	0	0	2(2.25)	2(2.25)	15	16.406
MAD	10(11.24)	4(4.49)	8(8.99)	7(7.87)	6(6.74)	4(4.49)	8(8.99)	9(10.11)	2(2.25)	2(2.25)	5(5.62)	0	7(7.87)	1(1.12)	1(1.12)	2(2.25)	3(3.37)	5(5.62)	2(2.25)	4(4.49)	15	17.739
DK	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Administrative Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	χ2
Imp. Tch.	1(1.12)	13(14.61)	8(8.99)	7(7.87)	0	12(13.48)	3(3.37)	12(13.48)	0	6(6.74)	2(2.25)	1(1.12)	2(2.25)	2(2.25)	3(3.37)	4(4.49)	0	5(5.62)	3(3.37)	4(4.49)	15	18.896
Imp. P/M	2(2.25)	13(14.61)	7(7.87)	7(7.87)	0	13(14.61)	1(1.12)	13(14.61)	0	6(6.74)	2(2.25)	1(1.12)	1(1.12)	4(4.49)	1(1.12)	5(5.62)	0	6(6.74)	2(2.25)	4(4.49)	15	15.449
MAD	7(7.87)	2(2.25)	12(13.48)	8(8.99)	3(3.37)	7(6.74)	7(7.87)	11(12.36)	1(1.12)	1(1.12)	4(4.49)	3(3.37)	3(3.37)	1(1.12)	2(2.25)	5(5.62)	2(2.25)	0	7(7.87)	3(3.37)	15	13.961
DK	5(5.62)	0	1(1.12)	23(25.84)	7(7.87)	0	2(2.25)	18(20.22)	3(3.37)	0	0	6(6.74)	2(2.25)	0	0	9(10.11)	0	0	0	12(13.48)	10	8.004

Peer Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	χ2
Imp. Tch.	1(1.12)	15(16.85)	3(3.37)	10(11.24)	1(1.12)	13(14.61)	2(2.25)	11(12.36)	0	5(5.62)	0	4(4.49)	0	6(6.74)	0	5(5.62)	0	5(5.62)	2(2.25)	5(5.62)	15	6.298
Imp. P/M	1(1.12)	11(12.36)	4(4.49)	13(14.61)	0	10(11.24)	1(1.12)	16(17.98)	0	5(5.62)	0	4(4.49)	0	4(4.49)	0	7(7.87	0	4(4.49)	2(2.25)	6(6.74)	15	9.344
MAD	2(2.25)	10(11.24)	1(1.12)	16(17.98)	3(3.37)	5(5.62)	3(3.37)	16(17.98)	0	3(3.37)	1(1.12)	5(5.62)	1(1.12)	0	1(1.12)	9(10.11)	0	1(1.12)	2(2.25)	9(10.11)	15	12.484
DK	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Self Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	χ2
Imp. Tch.	4(4.49)	9(10.11)	13(14.61)	3(3.37)	1(1.12)	11(12.36)	9(10.11)	6(6.74)	1(1.12)	1(1.12)	6(6.74)	1(1.12)	3(3.37)	1(1.12)	4(4.49)	3(3.37)	0	3(3.37)	7(7.87)	2(2.25)	15	18.936
Imp. P/M	2(2.25)	9(10.11)	8(8.99)	10(11.24)	2(2.25)	10(11.24)	2(2.25)	13(14.61)	1(1.12)	2(2.25)	5(5.62)	1(1.12)	2(2.25)	2(2.25)	2(2.25)	5(5.62)	0	3(3.37)	5(5.62)	4(4.49)	15	16.668
MAD	3(3.37)	6(6.74)	6(6.74)	14(15.73)	4(4.49)	4(4.49)	2(2.25)	17(19.10)	0	1(1.12)	6(6.74)	2(2.25)	0	1(1.12)	0	10(11.24)	2(2.25)	0	4(4.49)	6(6.74)	15	26.742
DK	5(5.62)	0	1(1.12)	23(25.84)	8(8.99)	0	1(1.12)	18(20.22)	0	0	0	9(10.11)	1(1.12)	0	0	10(11.24)	1(1.12)	0	0	11(12.36)	10	7.840

Note. Imp. Tch.=Improve Teaching; Imp. P/M=Improve Program/Major; MAD=Make Administrative Decisions; DK=Don't Know; IU=Is Used; SBU=Should Be Used; I/SBU=Is & Should Be Used; NR=No Response

Appendix E47. A Comparison of Instructor Responses, by Years Teaching Experience at Belmont University, to the Perceived Disposition of the four Types of Evaluation Practiced at Belmont University

Student Evaluation (Per cent)

Disp.	1-5 Years ID	1-5 Years SBD	1-5 Years I/SBD	1-5 Years NR	6-10 Years ID	6-10 Years SBD	6-10 Years I/SBD	6-10 Years NR	11-15 Years ID	11-15 Years SBD	11-15 Years I/SBD	11-15 Years NR	16-20 Years ID	16-20 Years SBD	16-20 Years I/SBD	16-20 Years NR	20+ Years ID	20+ Years SBD	20+ Years I/SBD	20+ Years NR	df	X2
RINR	18(20.22)	2(2.25)	7(7.87)	2(2.25)	17(19.10)	0	8(8.99)	2(2.25)	6(6.74)	0	2(2.25)	1(1.12)	6(6.74)	1(1.12)	3(3.37)	1(1.12)	4(4.49)	1(1.12)	5(5.62)	2(2.25)	15	15.518
RFR	1(1.12)	13(14.61)	0	15(16.85)	0	9(10.11)	1(1.12)	17(19.10)	1(1.12)	5(5.62)	2(2.25)	1(1.12)	0	3(3.37)	0	8(8.99)	0	4(4.49)	1(1.12)	7(7.87)	15	19.593
NOR	2(2.25)	0	0	17(30.34)	0	1(1.12)	0	26(29.21)	2(2.25)	0	0	7(7.87)	1(1.12)	0	0	10(11.24)	0	1(1.12)	0	11(12.36	10	10.605
NOTH.	0	1(1.12)	0	28(31.46)	0	3(3.37)	0	24(26.47)	0	0	0	9(10.11)	0	0	0	11(12.36	0	0	0	12(13.48	5	4.381

Administrative Evaluation (Per cent)

Disp.	1-5 Years ID	1-5 Years SBD	1-5 Years I/SBD	1-5 Years NR	6-10 years ID	6-10 years SBD	6-10 years I/SBD	6-10 years NR	11-15 Years ID	11-15 Years SBD	11-15 Years I/SBD	11-15 Years NR	16-20 Years ID	16-20 Years SDB	16-20 Years I/SBD	16-20 Years NR	20+ Years ID	20+ Years SBD	20+ Years I/SBD	20+ Years NR	df	X2
RINR	10(11.24)	1(1.12)	5(5.62)	13(14.61)	2(2.25)	0	3(3.37)	22(24.72)	2(2.25)	0	0	7(7.87)	1(1.12)	1(1.12)	1(1.12)	8(8.99)	3(3.37)	3(3.37)	2(2.25)	4(4.49)	15	24.262
RFR	1(1.12)	11(12.36)	4(4.49)	13(14.61)	0	11(12.36)	1(1.12)	15(16.85)	1(1.12)	5(5.62)	1(1.12)	2(2.25)	0	1(1.12)	2(2.25)	8(8.99)	0	4(4.49)	2(2.25)	6(6.74)	15	13.971
NOR	0	1(1.12)	0	28(31.46)	9(10.11)	0	1(1.12)	17(19.10)	3(3.37)	1(1.12)	0	5(5.62)	1(1.12)	0	1(1.12)	9(10.11)	3(3.37)	0	0	9(10.11)	15	21.854
NOTH.	3(3.37)	0	0	26(29.21)	4(4.49)	1(1.12)	1(1.12)	21(23.60)	2(2.25)	0	0	7(7.87)	2(2.25)	0	0	9(10.11)	0	0	0	12(13.48	15	8.084

Peer Evaluation (Per cent)

Disp.	1-5 Years ID	1-5 Years SBD	1-5 Years I/SBD	1-5 Years NR	6-10 Years ID	6-10 Years SBD	6-10 Years I/SBD	6-10 Years NR	11-15 Years ID	11-15 Years SBD	11-15 Years I/SBD	11-15 Years NR	16-20 Years ID	16-20 Years SBD	16-20 Years I/SBD	16-20 Years NR	20+ Years ID	20+ Years SBD	20+ Years I/SBD	20+ Years NR	df	X2
RINR	3(3.37)	4(4.49)	1(1.12)	21(23.60)	0	1(1.12)	3(3.37)	23(25.84)	1(1.12)	0	0	8(8.99)	0	2(2.25)	0	9(10.11)	1(1.12)	3(3.37)	1(1.12)	7(7.87)	15	13.280
RFR	0	11(12.36)	0	18(20.22)	0	5(5.62)	0	22(24.72)	1(1.12)	3(3.37)	0	5(5.62)	0	2(2.25)	0	9(10.11)	0	2(2.25)	0	10(11.24)	10	13.692
NOR	2(2.25)	1(1.12)	0	26(29.12)	5(5.62)	0	1(1.12)	21(23.60)	1(1.12)	0	0	8(8.99)	3(3.37)	0	0	8(8.99)	0	0	0	12(13.48	15	10.281
NOTH.	11(12.36)	0	2(2.25)	16(17.98)	8(8.99)	0	3(3.37)	16(17.98)	2(2.25)	0	0	7(7.87)	4(4.49)	0	0	7(7.87)	2(2.25)	0	0	10(11.24)	10	6.931

Self Evaluation (Per cent)

Disp.	1-5 Years ID	1-5 Years SBD	1-5 Years I/SBD	1-5 Years NR	6-10 Years ID	6-10 Years SBD	6-10 Years I/SBD	6-10 Years NR	11-15 Years ID	11-15 Years SBD	11-15 Years I/SBD	11-15 Years NR	16-20 Years ID	16-20 Years SBD	16-20 Years I/SBD	16-20 Years NR	20+ Years ID	20+ Years SBD	20+ Years I/SBD	20+ Years NR	df	X2
RINR	9(10.11)	1(1.12)	1(1.12)	18(20.22)	1(1.12)	0	2(2.25)	24(26.97)	3(3.37)	0	1(1.12)	5(5.62)	0	1(1.12)	0	10(11.24)	1(1.12)	0	2(2.25)	9(10.1)	15	20.000
RFR	0	9(10.11)	2(2.25)	18(20.22)	0	5(5.62)	2(2.25)	20(22.47)	1(1.12)	4(4.49)	2(2.25)	2(2.25)	0	1(1.12)	1(1.12)	9(10.11)	0	4(4.49)	2(2.25)	6(6.74)	15	19.157
NOR	3(3.37)	1(1.12)	1(1.12)	24(26.97)	3(3.37)	0	1(1.12)	23(25.84)	3(3.37)	0	0	6(6.74)	2(2.25)	0	2(2.25)	7(7.87)	2(2.25)	0	0	10(11.24)	15	11.530
NOTH.	9(10.11)	1(1.12)	0	19(21.35)	6(6.74)	0	3(3.37)	18(20.22)	1(1.12)	0	0	8(8.99)	2(2.25)	0	0	9(10.11)	1(1.12)	0	2(2.25)	9(10.11)	15	12.784

ID=Is Done; SBD=Should Be Done; I/SBD=Is & Should Be Done; NR=No Response

Note . Disp.=Disposition; RINR=Return to Instructor/No Response Expected, RFR=Return to Instructor for Response; NOR=Not Returned; NOTH.=Nothing;

Appendix E48. A Comparison of Instructor Responses, by Years Teaching Experience at Belmont University, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program and Major, and in Making Administrative Decisions

Improving Teaching (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	1(1.12)	13(14.61)	10(11.24)	3(3.37)	2(2.25)	2(2.25)	10(11.24)	13(14.61)	2(2.25)	0	1(1.12)	3(3.37)	4(4.49)	1(1.12)	0	0	4(4.49	5(5 62	1(1.12)	1(1.12)	1(1 12)	5(5 62	4(4.49	1(1.12)	1(1 12)			
Adm.	0	12(13.48)	7(7.87)	7(7.87)	3(3.37)	0	11(12.36)	8(8.99)	4(4.49)	4(4.49)	0	2(2.25)	4(4.49)	1(1.12)	2(2.25)	0	6(6.74	2(2 25	1(1.12)	2(2 25	1(1.12)	3(3.37	4(4.49	2(2 25	2(2.25	20	22.429
Peer	1(1.12)	6(6.74)	7(7.87)	6(6.74)	9(10.11)	2(2.25)	4(4.49)	7(7.87)	3(3.37)	11(12.36)	0	0	2(2.25)	3(3.37)	4(4.49)	0	3(3.37	1(1.12)	3(3.37	4(4.49	1(1.12)	1(1 12)	1(1.12)	2(2 25	7(7 87	20	17.734
Self	1(1.12)	16(17.98)	7(7.87)	2(2.25)	3(3.37)	3(3.37)	9(10.11)	7(7.87)	3(3.37)	5(5.62)	1(1.12)	4(4.49)	2(2.25)	1(1.12)	1(1.12)	0	8(8.99	0	1(1.12)	2(2.25	2(2.25	4(4 49	5(5 62	0	1(1.12)	20	12.745
																										20	19.641

Improving Program/Major (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	1(1.12)	11(12.36)	9(10.11)	5(5.62)	3(3.37)	1(1.12)	7(7.87)	16(17.98)	3(3.37)	0	0	1(1.12)	7(7.87)	1(1.12)	0	0	4(4.49	4(4.49	2(2.25	1(1.12)	0	4(4.49	4(4.49	3(3 37	1(1 12)	20	26.509
Adm.	0	14(15.73)	5(5.62)	7(7.87)	3(3.37)	0	7(7.87)	10(11.24)	6(6.74)	4(4.49)	0	2(2.25)	3(3.37)	2(2.25)	2(2.25)	0	3(3.37	3(3.37	3(3.37	2(2.25	1(1.12)	4(4.49	3(3.37	3(3 37	1(1.12)		16.480
Peer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Self	0	14(15.73)	7(7.87)	5(5.62)	3(3.37)	2(2.25)	5(5.62)	10(11.24)	4(4.49)	6(6.74)	0	3(3.37)	4(4.49)	1(1.12)	1(1.12)	0	5(5.62	2(2.25	2(2.25	2(2.25	2(2.25	4(4.49	3(3.37	2(2 25	1(1.12	20	20 097

Making Administrative Decisions (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	χ2
Stu.	0	10(11.24)	8(8.99)	5(5.62)	6(6.74)	0	11(12.36)	11(12.36)	4(4.49)	1(1.12)	0	2(2.25)	4(4.49)	1(1.12)	2(2.25)	0	2(2.25	4(4.49	1(1.12)	2(2.25	0	3(3.37	6(6.74	1(1.12)	1(1 12)	15	18.362
Adm.	0	15(16.85)	5(5.62)	3(3.37)	6(6.74)	0	12(13.48)	6(6.74)	3(3.37)	6(6.74)	0	2(2.25)	4(4.49)	1(1.12)	2(2.25)	0	7(7.87	1(1.12)	1(1.12)	2(2.25	1(1.12)	4(4.49	2(2 25	3(3.37	2(2.25	20	18.158
Peer	0	10(11.24)	5(5.62)	5(5.62)	9(10.11)	0	6(6.74)	5(5.62)	3(3.37)	13(14.61)	0	0	2(2.25)	3(3.37)	4(4.49)	0	3(3.37	2(2.25	2(2.25	4(4.49	1(1.12)	1(1.12)	2(2.25	1(1.12)	7(7 87	20	17.477
Self	0	11(12.36)	6(6.74)	5(5.62)	7(7.87)	0	7(7.87)	8(8.99)	4(4.49)	8(8.99)	0	2(2.25)	4(4.49)	1(1.12)	2(2.25)	0	4(4.49	2(2.25	3(3.37	2(2.25	1(1.12)	4(4 47	2(2.25	3(3.37	2(2 25	20	14.865

Note: Stu.=Student; Adm.=Administrative; VE=Very Effective; E=Effective; IE=Ineffective; VIE=Very Ineffective; NR=No Response

--=Insufficient Numbers in Cells to Allow Chi Square Calculations

Appendix E49. A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fairness (Per cent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	VF	F	UF	VUF	NR	df	X2
Stu.	3(3.37)	20(22.47)	6(6.74)	3(3.37)	2(2.25)	0	10(11.24)	8(8.99)	0	0	0	4(4.49)	2(2.25)	0	0	0	1(1.12)	3(3.37	0	0	0	2(2.25	1(1.12)	0	0	20	21.314
Adm.	1(1.12)	22(24.72)	3(3.37)	1(1.12)	7(7.87)	2(2.25)	13(14.61)	3(3.37)	0	0	1(1.12)	3(3.37)	2(2.25)	0	0	0	1(1.12)	1(1.12)	0	2(2.25	0	3(3.37	0	0	0	20	21.688
Peer	1(1.12)	11(12.36)	3(3.37)	0	19(21.35)	3(3.37)	11(12.36)	1(1.12)	0	3(3.37)	0	2(2.25)	0	0	4(4.49)	0	1(1.12)	1(1.12)	0	2(2.25	0	1(1.12)	0	0	2(2.25	20	18.143
Self	5(5.62)	17(19.10)	3(3.37)	2(2.25)	7(7.87)	2(2.25)	15(16.85)	1(1.12)	0	0	0	6(6.74)	0	0	0	0	1(1.12)	1(1.12)	0	2(2.25	0	2(2.25	1(1.12)	0	0	20	21.095

Objectivity (Per cent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	VO	O	UO	VUO	NR	df	X2
Stu.	1(1.12)	12(13.48)	14(15.73)	5(5.62)	2(2.25)	1(1.12)	5(5.62)	10(11.24)	2(2.25)	0	0	1(1.12)	4(4.49)	1(1.12)	0	0	1(1.12)	2(2.25	1(1.12)	0	0	1(1.12)	2(2.25	0	0	20	12.625
Adm.	1(1.12)	16(17.98)	8(8.99)	2(2.25)	7(7.87)	0	12(13.48)	5(5.62)	1(1.12)	0	0	3(3.37)	3(3.37)	0	0	0	2(2.25	0	0	2(2.25	0	2(2.25	1(1.12)	0	0	20	21.465
Peer	1(1.12)	8(8.99)	6(6.74)	0	19(21.35)	1(1.12)	9(10.11)	5(5.62)	0	3(3.37)	0	1(1.12)	1(1.12)	0	4(4.49)	0	1(1.12)	1(1.12)	0	2(2.25	0	0	1(1.12)	0	2(2.25	20	15.359
Self	2(2.25)	15(16.85)	8(8.99)	1(1.12)	8(8.99)	0	9(10.11)	7(7.87)	2(2.25)	0	0	1(1.12)	4(4.49)	1(1.12)	0	0	2(2.25	0	0	2(2.25	0	2(2.25	0	0	1(1.12)	20	22.513

Reliability (Per cent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	VR	R	UR	VUR	NR	df	X2
Stu.	0	15(16.85)	13(14.61)	4(4.49)	2(2.25)	0	6(6.74)	9(10.11)	3(3.37)	0	0	1(1.12)	4(4.49)	1(1.12)	0	0	1(1.12)	2(2.25	1(1.12)	0	0	0	3(3.37	0	0	15	13.721
Adm.	0	16(17.98)	9(10.11)	2(2.25)	7(7.87)	0	15(16.85)	1(1.12)	2(2.25)	0	2(2.25)	1(1.12)	3(3.37)	0	0	0	1(1.12)	1(1.12)	0	2(2.25	0	3(3.37	0	0	0	20	36.422*
Peer	1(1.12)	11(12.36)	2(2.25)	1(1.12)	19(21.35)	1(1.12)	11(12.36)	2(2.25)	1(1.12)	3(3.37)	0	1(1.12)	1(1.12)	1(1.12)	3(3.37)	0	1(1.12)	0	1(1.12)	2(2.25	0	0	1(1.12)	0	2(2.25	20	19.521
Self	1(1.12)	19(21.35)	5(5.62)	1(1.12)	8(8.99)	2(2.25)	14(15.73)	2(2.25)	0	0	0	2(2.25)	4(4.49)	0	0	0	2(2.25	0	0	2(2.25	0	2(2.25	0	0	1(1.12)	20	30.135

Note. Stu.=Student; Adm.=Administrative; VF=Very Fair; F=Fair; UF=Unfair; VUF=Very Unfair; NR=No Response; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR=Unreliable;VUR=Very Unreliable

Appendix E5Q. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	df	χ2
Stu.	1(1.12)	11(12.36)	14(15.73)	6(6.74)	2(2.25)	0	8(8.99)	6(6.74)	4(4.49)	0	0	2(2.25)	4(4.49)	0	0	0	1(1.12)	1(1.12)	2(2.25	0	0	0	3(3.37	0	0	20	17.488
Adm.	0	17(19.10)	10(11.24)	4(4.49)	3(3.37)	2(2.25)	8(8.99)	3(3.37)	4(4.49)	1(1.12)	0	3(3.37)	3(3.37)	0	0	0	2(2.25	0	2(2.25	0	0	2(2.25	1(1.12)	0	0	20	27.331
Peer	0	10(11.24)	3(3.37)	5(5.62)	16(17.98)	2(2.25)	9(10.11)	3(3.37)	1(1.12)	3(3.37)	0	1(1.12)	2(2.25)	1(1.12)	2(2.25)	0	1(1.12)	0	2(2.25	1(1.12)	0	0	1(1.12)	0	2(2.25	20	24.794
Self	4(4.49)	15(16.85)	6(6.74)	3(3.37)	6(6.74)	3(3.37)	11(12.36)	3(3.37)	1(1.12)	0	0	3(3.37)	2(2.25)	0	1(1.12)	0	2(2.25	1(1.12)	0	1(1.12)	0	2(2.25	0	0	1(1.12)	20	11.065

A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perception of Overall Satisfaction with the Evaluation System at Belmont University

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	VS	S	US	VUS	NR	df	χ2
0	16(17.98)	14(15.73)	3(3.37)	1(1.12)	0	11(12.36)	6(6.74)	1(1.12)	0	1(1.12)	2(2.25)	2(2.25)	1(1.12)	0	0	1(1.12)	1(1.12)	2(2.25	0	0	2(2.25	1(3.12)	0	0	20	26.893

Note: Stu.=Student; Adm.=Administrative; VS=Very Satisfied; S=Satisfied; US=Unsatisfied; VUS=Very Unsatisfied; NR=No Response

Appendix E51. A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University Should or Should Not Be Done

Type	1-5 Years			6-10 Years			11-15 Years			16-20 Years			20+ Years				
	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	df	χ2
Stu.	28(31.46	4(4.49)	2(2.25)	16(17.98	2(2.25)	0	5(5.62)	1(1.12)	0	3(3.37)	1(1.12)	0	2(2.25)	0	1(1.12)	10	7.556
Adm.	28(31.46	3(3.37)	3(3.37)	17(19.10	1(1.12)	0	6(6.74)	0	0	3(3.37)	1(1.12)	0	1(1.15)	1(1.12)	0	10	8.033
Peer	25(28.09	6(6.74)	3(3.37)	14(15.73	4(4.49)	0	5(5.62)	1(1.12)	0	3(3.37)	1(1.12)	0	2(2.25)	1(1.12)	0	10	8.448
Self	30(33.71	1(1.12)	3(3.37)	14(15.73	3(3.37)	1(1.12)	4(4.49)	2(2.25)	0	4(4.49)	0	0	3(3.37)	0	0	10	10.530

Note. Stu.=Student; Administrative; S=Should Be Done; SN=Should Not Be Done; NR=No Response

Appendix E52. A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perceived Usage of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	X2
Imp Tch	2(2.25)	22(24.72)	9(10.11)	1(1.12)	1(1.12)	6(6.74)	11(12.36)	0	0	3(3.37)	2(2.25)	1(1.12)	0	3(3.37)	1(1.12)	0	0	0	2(2.25)	1(1.12)	15	21.622
Imp P/M	2(2.25)	18(20.22)	8(8.99)	6(6.74)	0	10(11.24)	4(4.49)	4(4.49)	0	3(3.37)	1(1.12)	2(2.25)	0	2(2.25)	0	2(2.25)	0	1(1.12)	1(1.12)	1(1.12)	15	6.430
MAD	10(11.24)	4(4.49)	14(15.73)	6(6.74)	5(5.62)	2(2.25)	6(6.74)	5(5.62)	3(3.37)	2(2.25)	0	1(1.12)	2(2.25)	1(1.12)	1(1.12)	0	1(1.12)	0	0	2(2.25)	15	14.329
DK	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Administrative Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	X2
Imp Tch	1(1.12)	18(20.22)	5(5.62)	10(11.24)	1(1.12)	6(6.74)	6(6.74)	5(5.62)	0	4(4.49)	1(1.12)	1(1.12)	0	1(1.12)	0	3(3.37)	0	1(1.12)	1(1.12)	1(1.12)	15	9.805
Imp P/M	2(2.25)	16(17.98)	7(7.87)	9(10.11)	0	9(10.11)	3(3.37)	6(6.74)	0	4(4.49)	1(1.12)	1(1.12)	0	1(1.12)	0	3(3.37)	0	2(2.25)	1(1.12)	0	15	12.796
MAD	8(8.99)	4(4.49)	12(13.48)	10(11.24)	2(2.25)	3(3.37)	8(8.99)	5(5.62)	2(2.25)	1(1.12)	1(1.12)	2(2.25)	1(1.12)	0	0	3(3.37)	1(1.12)	0	1(1.12)	1(1.12)	15	10.564
DK	8(8.99)	0	1(1.12)	25(28.09)	2(2.25)	0	1(1.12)	15(16.85)	1(1.12)	0	0	5(5.62)	1(1.12)	0	0	3(3.37)	0	0	0	3(3.37)	10	2.803

Peer Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	X2
Imp Tch	1(1.12)	20(22.47)	1(1.12)	12(13.48)	0	11(12.36)	3(3.37)	4(4.49)	1(1.12)	3(3.37)	0	2(2.25)	0	1(1.12)	0	3(3.37)	0	0	1(1.12)	2(2.25)	15	21.158
Imp P/M	1(1.12)	14(15.73)	2(2.25)	17(19.10)	0	7(7.87)	2(2.25)	9(10.11)	0	3(3.37)	0	3(3.37)	0	1(1.12)	0	3(3.37)	0	1(1.12)	1(1.12)	1(1.12)	15	6.759
MAD	4(4.49)	9(10.11)	1(1.12)	20(22.47)	1(1.12)	5(5.62)	2(2.25)	10(11.24)	0	1(1.12)	0	5(5.62)	1(1.12)	0	0	3(3.37)	0	1(1.12)	1(1.12)	1(1.12)	15	15.593
DK	11(12.36)	2(2.25)	1(1.12)	20(22.47)	6(6.74)	0	1(1.12)	11(12.96)	2(2.25)	0	0	4(4.49)	1(1.12)	0	0	3(3.37)	0	0	1(1.12)	2(2.25)	15	9.922

Self Evaluation (Per cent)

Use	1-5 Years IU	SBU	I/SBU	NR	6-10 Years IU	SBU	I/SBU	NR	11-15 Years IU	SBU	I/SBU	NR	16-20 Years IU	SBU	I/SBU	NR	20+ Years IU	SBU	I/SBU	NR	df	X2
Imp Tch	4(4.49)	12(13.48)	13(14.61)	5(5.62)	1(1.12)	4(4.49)	11(12.36)	2(2.25)	1(1.12)	3(3.37)	0	2(2.25)	0	2(2.25)	1(1.12)	1(1.12)	0	1(1.12)	1(1.12)	1(1.12)	15	13.381
Imp P/M	3(3.37)	12(13.48)	10(11.24)	9(10.11)	0	5(5.62)	4(4.49)	9(10.11)	0	3(3.37)	0	3(3.37)	0	1(1.12)	0	3(3.37)	0	1(1.12)	1(1.12)	1(1.12)	15	14.148
MAD	7(7.87)	3(3.37)	7(7.87)	17(19.10)	2(2.25)	3(3.37)	6(6.74)	7(7.87)	0	1(1.12)	1(1.12)	4(4.49)	0	2(2.25)	0	2(2.25)	0	1(1.12)	1(1.12)	1(1.12)	15	20.693
DK	9(10.11)	0	0	25(28.09)	1(1.12)	0	0	17(19.10)	1(1.12)	0	0	5(5.62)	0	0	0	4(4.49)	5	0	0	2(2.25)	10	10.979

Note: Imp Tch/=Improve Teaching; Imp P/M=Improve Program/Major; MAD=Make Administrative Decisions; DK= Don't Know, IU=Is Used; SBU=Should Be Used, I/SBU=Is & Should Be Used; NR=No Response

Appendix ES3 A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perceived Disposition of Evaluation of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years			16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBC	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	22(24.72)	1(1.12)	7(7.87)	4(4.49)	10(11.24)	1(1.12)	7(6.74)	1(1.12)	4(4.49)	1(1.12)	1(1.12)	0	2(2.25)	0	2(2.25)	0	2(2.25)	0	1(1.12)	0	15	9.466
RIFR	1(1.12)	15(16.85)	3(3.37)	15(16.85)	1(1.12)	6(6.74)	1(1.12)	10(11.24)	0	2(2.25)	0	4(4.49)	0	2(2.25)	0	2(2.25)	0	1(1.12)	0	2(2.25)	15	7.247
NOR	2(2.25)	0	0	32(35.96)	1(1.12)	1(1.12)	0	16(17.98)	1(1.12)	0	0	5(5.62)	0	0	0	4(4.49)	0	0	0	3(3.37)	10	4.251
NOTH	0	2(2.25)	0	32(35.96)	0	1(1.12)	0	17(19.10)	0	0	0	6(6.74)	0	0	0	4(4.49)	0	0	0	3(3.37)	5	0.818

Administrative Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years			16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	9(10.11)	2(2.25)	4(4.49)	19(21.35)	3(3.37)	0	4(4.49)	11(12.36)	2(2.25)	0	0	4(4.49)	0	1(1.12)	0	3(3.37)	1(1.12)	0	0	2(2.25)	15	11.028
RIFR	2(2.25)	12(13.48)	2(2.25)	18(20.22)	0	8(8.99)	4(4.49)	6(6.74)	0	4(4.49)	1(1.12)	1(1.12)	0	0	0	4(4.49)	0	1(1.12)	0	2(2.25)	15	15.537
NOR	8(8.99)	1(1.12)	0	25(28.09)	2(2.25)	1(1.12)	0	15(16.85)	2(2.25)	0	0	4(4.49)	1(1.12)	0	0	3(3.37)	1(1.12)	0	1(1.12)	1(1.12)	15	21.362
NOTH	4(4.49)	0	0	30(33.71)	0	1(1.12)	0	17(19.10)	2(2.25)	0	0	4(4.49)	1(1.12)	0	0	3(3.37)	0	0	0	3(3.37)	15	12.957

Peer Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years			16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	2(2.25)	4(4.49)	0	28(31.46)	2(2.25)	2(2.25)	3(3.37)	11(12.36)	0	1(1.12)	0	5(5.62)	0	1(1.12)	0	3(3.37)	0	0	0	3(3.37)	15	10.977
RIFR	1(1.12)	7(7.87)	0	26(29.21)	0	6(6.74)	0	12(13.48)	0	2(2.25)	0	4(4.49)	0	0	0	4(4.49)	0	1(1.12)	0	2(2.25)	10	4.322
NOR	4(4.49)	0	1(1.12)	29(32.58)	3(3.37)	1(1.12)	0	14(15.73)	0	0	0	6(6.74)	1(1.12)	0	0	3(3.37)	0	0	0	2(2.25)	15	9.040
NOTH	13(14.61)	0	0	21(23.60)	4(4.49)	0	3(3.37)	11(12.36)	3(3.37)	0	0	3(3.37)	1(1.12)	0	3(3.37)	4(4.49)	0	1(1.12)	2(2.25)	3(3.37)	10	14.090

Self Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years			16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	7(7.87)	0	3(3.37)	24(26.97)	3(3.37)	0	1(1.12)	14(15.73)	2(2.25)	1(1.12)	0	3(3.37)	0	0	0	4(4.49)	1(1.12)	0	0	2(2.25)	15	14.641
RIFR	1(1.12)	9(10.11)	4(4.49)	20(22.47)	0	6(6.74)	2(2.25)	10(11.24)	0	2(2.25)	0	4(4.49)	0	0	0	4(4.49)	1(1.12)	0	0	2(2.25)	15	6.226
NOR	8(8.99)	0	1(1.12)	25(28.09)	1(1.12)	1(1.12)	0	16(17.98)	0	0	0	6(6.74)	0	0	0	3(3.37)	1(1.12)	0	1(1.12)	1(1.12)	15	21.144
NOTH	6(6.74)	0	0	28(31.46)	3(3.37)	0	3(3.37)	12(13.48)	3(3.37)	1(1.12)	0	2(2.25)	1(1.12)	0	0	3(3.37)	0	0	0	3(3.37)	15	26.478

Note: RINR=Return to Instructor/No Response Requested; RIFR=Return to Instructor for Response; NOR=Not Returned; NOTH.=Nothing; ID=Is Done; SBD=Should Be Done; I/SBD=Is & Should Be Done; NR=No Response

Appendix E54 A Comparison of Instructor Responses, by Years Teaching Experience at Other Institutions, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu	3(3.37)	10(11.24)	14(15.73)	5(5.62)	2(2.25)	1(1.12)	10(11.24)	6(6.74)	1(1.12)	0	0	2(2.25)	4(4.49)	0	0	0	1(1.12)	2(2.25	1(1.12)	0	0	0	3(3.37	0	0	20	18.174
Adm	0	10(11.24)	11(12.36)	7(7.87)	6(6.74)	0	8(8.99)	4(4.49)	4(4.49)	2(2.25)	0	4(4.49)	2(2.25)	0	0	0	1(1.12)	1(1.12)	0	2(2.25	0	1(1.12)	2(2.25	0	0	20	15.201
Peer	0	5(5.62)	6(6.74)	7(7.87)	16(17.98)	3(3.37)	5(5.62)	3(3.37)	3(3.37)	4(4.49)	0	0	3(3.37)	2(2.25)	1(1.12)	0	1(1.12)	0	1(1.12)	2(2.25	0	0	2(2.25	0	1(1.12)	20	23.348
Self	3(3.37)	13(14.61)	9(10.11)	3(3.37)	6(6.74)	2(2.25)	9(10.11)	4(4.49)	2(2.25)	1(1.12)	0	2(2.25)	4(4.49)	0	0	0	3(3.37	0	0	1(1.12)	0	2(2.25	1(1.12)	0	0	20	15.068

Improving Program/Major (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Year						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu	0	10(11.24)	17(19.10)	5(5.62)	2(2.25)	1(1.12)	6(6.74)	5(5.62)	5(5.62)	1(1.12)	1(1.12)	2(2.25)	3(3.37)	0	0	0	1(1.12)	1(1.12)	2(2.25	0	0	0	3(3.37	0	0	20	21.830
Adm.	0	11(12.36)	10(11.24)	7(7.87)	6(6.74)	0	8(8.99)	1(1.12)	8(8.99)	1(1.12)	0	3(3.37)	3(3.37)	0	0	0	1(1.12)	0	1(1.12)	2(2.25	0	0	3(3.37	0	0	20	28.125
Peer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Self	1(1.12)	12(13.48)	10(11.24)	5(5.62)	6(6.74)	1(1.12)	7(7.87)	4(4.49)	5(5.62)	1(1.12)	0	1(1.12)	4(4.49)	1(1.12)	0	0	2(2.25	0	0	2(2.25	0	1(1.12)	2(2.25	0	0	20	17.484

Making Administrative Decisions (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu	0	12(13.48)	12(13.48)	6(6.74)	4(4.49)	0	8(8.99)	4(4.49)	3(3.37)	3(3.37)	0	2(2.25)	4(4.49)	0	0	0	2(2.25	1(1.12)	1(1.12)	0	0	0	3(3.37	0	0	15	16.624
Adm.	0	13(14.61)	8(8.99)	6(6.74)	7(7.87)	0	12(13.48)	1(1.12)	2(2.25)	3(3.37)	0	4(4.49)	2(2.25)	0	0	0	2(2.25	0	0	2(2.25	0	1(1.12)	2(2.25	0	0	20	19.762
Peer	0	7(7.87)	3(3.37)	8(8.99)	16(17.98,	0	6(6.74)	4(4.49)	3(3.37)	5(5.62)	0	3(3.37)	2(2.25)	0	1(1.12)	0	2(2.25	0	0	2(2.25	0	0	2(2.25	0	1(1.12)	20	23.646
Self	0	9(10.11)	9(10.11)	7(7.87)	9(10.11)	0	7(7.87)	4(4.49)	4(4.49)	3(3.37)	0	2(2.25)	4(4.49)	0	0	0	2(2.25	0	1(1.12)	2(2.25	0	2(2.25	1(1.12)	0	0	20	17.196

Note Stu.=Student:: Adm.=Administratorv; VE=Very Effective; E=Effective, IE=Ineffective, VIE=Very Ineffective; NR=No Response

Appendix E55. A Comparison of Instructor Responses, by Total Years Teaching Experience, to the Perceived Fairness, Objectivity and Reliability of the Four Types of Evaluation Practiced at Belmont University

Fairness (Percent)

Type	1-5 Years VF	F	UF	VUF	NR	6-10 Years VF	F	UF	VUF	NR	11-15 Years VF	F	UF	VUF	NR	16-20 Years VF	F	UF	VUF	NR	20+ Years VF	F	UF	VUF	NR	df	X2
Stu.	0	4(4.49)	2(2.25)	0	1(1.12)	0	12(13.48)	7(7.87)	2(2.25)	1(1.12)	2(2.25)	13(14.61)	6(6.74)	1(1.12)	0	1(1.12)	13(14.61)	1(1.12)	0	1(1.12)	0	11(12.36)	8(10.11)	0	1(1.12)	20	32.987
Adm.	0	5(5.62)	1(1.12)	0	1(1.12)	2(2.25)	10(11.24)	6(6.74)	0	4(4.49)	1(1.12)	16(17.98)	1(1.12)	1(1.12)	3(3.37)	1(1.12)	12(13.48)	1(1.12)	0	2(2.25)	5(5.62)	9(10.11)	3(3.37)	1(1.12)	3(3.37)	20	21.572
Peer	0	3(3.37)	1(1.12)	0	3(3.37)	1(1.12)	11(12.36)	1(1.12)	1(1.12)	8(8.99)	2(2.25)	9(10.11)	2(2.25)	0	9(10.11)	1(1.12)	5(5.62)	1(1.12)	0	9(10.11)	2(2.25)	6(6.74)	1(1.12)	0	12(13.48)	20	9.339
Self	1(1.12)	4(4.49)	0	0	2(2.25)	2(2.25)	11(12.36)	1(1.12)	4(4.49)	4(4.49)	2(2.25)	16(17.98)	2(2.25)	0	2(2.25)	1(1.12)	10(11.24)	3(3.37)	0	2(2.25)	3(3.37)	14(15.73)	2(2.25)	0	2(2.25)	20	24.595

Objectivity (Percent)

Type	1-5 Years VO	O	UO	VUO	NR	6-10 Years VO	O	UO	VUO	NR	11-15 Years VO	O	UO	VUO	NR	16-20 Years VO	O	UO	VUO	NR	20+ Years VO	O	UO	VUO	NR	df	X2
Stu.	0	2(2.24)	4(4.49)	0	1(1.12)	0	8(8.99)	8(8.99)	5(5.62)	1(1.12)	0	8(8.99)	11(12.36)	3(3.37)	0	1(1.12)	4(4.49)	10(11.24)	0	1(1.12)	1(1.12)	8(8.99)	10(11.24)	1(1.12)	1(1.12)	20	30.658
Adm.	0	4(4.49)	2(2.25)	0	1(1.12)	0	9(10.11)	8(8.99)	1(1.12)	4(4.49)	0	14(15.73)	4(4.49)	1(1.12)	3(3.37)	1(1.12)	7(7.87)	6(6.74)	0	2(2.25)	4(4.49)	9(10.11)	3(3.37)	2(2.25)	3(3.37)	20	22.499
Peer	0	2(2.25)	2(2.25)	0	3(3.37)	0	7(7.87)	6(6.74)	1(1.12)	8(8.99)	2(2.25)	7(7.87)	4(4.49)	0	9(10.11)	0	4(4.49)	3(3.37)	0	9(10.11)	2(2.25)	4(4.49)	3(3.37)	0	12(13.48)	20	11.982
Self	0	3(3.37)	2(2.25)	0	2(2.25)	0	9(10.11)	4(4.49)	3(3.37)	6(6.74)	1(1.12)	8(8.99)	8(8.99)	1(1.12)	4(4.49)	1(1.12)	6(6.74)	6(6.74)	1(1.12)	2(2.25)	2(2.25)	12(13.48)	3(3.37)	1(1.12)	3(3.37)	20	15.253

Reliability (Percent)

Type	1-5 Years VR	R	UR	VUR	NR	6-10 Years VR	R	UR	VUR	NR	11-15 Years VR	R	UR	VUR	NR	16-20 Years VR	R	UR	VUR	NR	20+ Years VR	R	UR	VUR	NR	df	X2
Stu.	0	3(3.37)	2(2.25)	1(1.12)	1(1.12)	0	6(6.74)	11(12.36)	3(3.37)	2(2.25)	0	10(11.24)	9(10.11)	3(3.37)	0	0	7(7.87)	6(6.74)	2(2.25)	1(1.12)	0	7(7.87)	11(12.36)	2(2.25)	1(1.12)	15	18.855
Adm.	0	4(4.49)	2(2.25)	0	1(1.12)	0	9(10.11)	8(8.99)	1(1.12)	4(4.49)	1(1.12)	11(12.36)	4(4.49)	3(3.37)	3(3.37)	1(1.12)	12(13.48)	1(1.12)	0	2(2.25)	3(3.37)	11(12.36)	3(3.37)	1(1.12)	3(3.37)	20	21.194
Peer	0	4(4.49)	0	0	3(3.37)	1(1.12)	8(8.99)	4(4.49)	1(1.12)	8(8.99)	1(1.12)	8(8.99)	2(2.25)	2(2.25)	9(10.11)	0	6(6.74)	1(1.12)	1(1.12)	8(8.99)	2(2.25)	5(5.62)	1(1.12)	1(1.12)	12(13.48)	20	10.585
Self	0	5(5.62)	0	0	2(2.25)	0	11(12.36)	4(4.49)	2(2.25)	5(5.62)	2(2.25)	12(13.48)	4(4.49)	0	4(4.49)	0	10(11.24)	3(3.37)	1(1.12)	2(2.25)	3(3.37)	14(15.73)	1(1.12)	0	3(3.37)	20	19.367

Note. Stu.=Student; Adm.=Administrative; VF=Very Fair; F=Fair; UF=Unfair; VUF=Very Unfair; VO=Very Objective; O=Objective; UO=Unobjective; VUO=Very Unobjective; VR=Very Reliable; R=Reliable; UR=Unreliable; VUR=Very Unreliable

Appendix E56. Instructor Perceptions of Satisfaction with the Belmont University Evaluation System

A Comparison of Instructor Responses, by Total Years Teaching Experience, to the Perceived Satisfaction Level of the Four Types of Evaluation Practiced at Belmont University

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	US	VUS	NR	VS	S	U	VSU	NR	VS	S	U	VUS	NR	df	χ2
Stu.	0	1(1.12)	2(2.25)	3(3.37)	1(1.12)	0	8(8.99)	9(10.11)	4(4.49)	1(1.12)	0	12(13.48)	6(6.74)	4(4.49)	0	1(1.12)	7(7.87	6(6.74	1(1.12)	1(1.12)	0	5(5.62	11(12.36	4(4.49	1(1.12)	20	33.246
Adm.	0	4(4.49)	2(2.25)	1(1.12)	0	2(2.25)	8(8.99)	6(6.74)	3(3.37)	3(3.37)	0	9(10.11)	6(6.74)	4(4.49)	3(3.37)	0	11(12.36	3(3.37	0	2(2.25	*3(3.37	11(12.36	4(4.49	2(2.25	1(1.12)	20	21.632
Peer	0	3(3.37)	1(1.12)	1(1.12)	2(2.25)	1(1.12)	8(8.99)	2(2.25)	2(2.25)	9(10.11)	0	10(11.24)	4(4.49)	2(2.25)	6(6.74)	1(1.12)	4(4.49	2(2.25	2(2.25	7(7.87	1(1.12)	4(4.49	2(2.25	3(3.37	11(12.36	20	9.317
Self	1(1.12)	3(3.37)	1(1.12)	1(1.12)	1(1.12)	3(3.37)	9(10.11)	4(4.49)	1(1.12)	5(5.62)	1(1.12)	14(15.73)	3(3.37)	1(1.12)	3(3.37)	2(2.25	8(8.99	3(3.37	1(1.12)	2(2.25	2(2.25	11(12.36	4(4.49	1(1.12)	3(3.37	20	9.561

A Comparison of Instructor Responses, by Total Years Teaching Experience, to the Perception of Overall Satisfaction with the Evaluation System at Belmont University

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	VS	S	U	VUS	NR	df	χ2
0	4(4.49)	3(3.37)	0	0	0	11(12.36)	7(7.87)	3(3.37)	1(1.12)	0	10(11.24)	10(11.24)	2(2.25)	0	1(1.12)	8(8.99	5(5.62	1(1.12	1(1.12)	0	8(8.99	9(10.11	3(3.37	1(1.12)	20	10.759

Note. Stu.=Student; Adm.=Administrative; VS=Very Satisfied; S=Satisfied; US=Unsatisfied; VUS=Very Unsatisfied; NR=No Response

Appendix E57. A Comparison of Instructor Responses, by Total Years Teaching Experience to the Perception of Whether the Four Types of Evaluation Practiced at Belmont University

(Per cent)

Type	1-5 Years			6-10 Years			11-15 Years			16-20 Years			20+ Years			df	χ2
	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR	S	SN	NR		
Stu.	5(5.62)	1(1.12)	1(1.12)	18(20.22	3(3.37)	1(1.12)	21(23.60	1(1.12	0	13(14.61)	2(2.25)	1(1.12)	16(17.98)	3(3.37)	2(2.25)	10	4.895
Adm.	6(6.74)	1(1.12)	0	17(19.10	3(3.37	2(2.25)	20(22.47	0	2(2.25	13(14.61)	2(2.25)	1(1.12)	18(20.22)	2(2.25)	1(1.12)	10	4.365
Peer	5(5.62)	1(1.12)	1(1.12)	15(16.85	4(4.49	3(3.37)	16(17.98	5(5.62)	1(1.12)	10(11.24)	5(5.62)	1(1.12)	15(16.85)	4(4.49)	2(2.25)	10	3.046
Self	6(6.74)	0	1(1.12)	17(19.10	2(2.25)	3(3.37)	18(20.22	2(2.25)	2(2.25)	12(13.48)	3(3.37)	1(1.12)	16(20.22)	2(2.25)	1(1.12)	10	3.663

Note. Stu.=Student; Adm.=Administrative; S=Should Be Done; SN=Should Not Be Done; NR=No Response

Appendix E5E A Comparison of Instructor Responses by Total Years Teaching Experience, to the Perceived Usage of the Four Types of Evaluation Practiced at Belmont University

Student Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	0	5(5.62)	2(2.25)	0	2(2.25)	13(14.61)	6(6.74)	1(1.12)	1(1.12)	12(13.48)	9(10.11)	0	2(2.25)	4(4.49)	7(7.87)	3(3.37)	1(1.12)	7(7.87)	11(12.36	2(2.25)	15	24.939
Imp. P/M	0	5(5.62)	1(1.12)	1(1.12)	1(1.12)	7(7.87)	6(6.74)	8(8.99)	1(1.12)	14(15.73)	5(5.62)	2(2.25)	2(2.25)	4(4.49)	4(4.49)	6(6.74)	0	14(15.73	3(3.37)	4(4.49)	15	18.896
MAD	3(3.37)	1(1.12)	2(2.25)	1(1.12)	6(6.74)	3(3.37)	6(6.74)	7(7.87)	7(7.87)	1(1.12)	10(11.24)	4(4.49)	4(4.49)	5(5.62)	3(3.37)	4(4.49)	8(8.99)	1(1.12)	6(6.74)	6(6.74)	15	14.218
DK	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Administrative Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	0	4(4.49)	1(1.12)	2(2.25)	1(1.12)	7(7.87)	4(4.49)	10(11.24)	0	14(15.73)	4(4.49)	4(4.49)	1(1.12)	4(4.49)	6(6.74)	5(5.62)	1(1.12)	9(10.11)	4(4.49)	7(7.87)	15	12.979
Imp. P/M	0	3(3.37)	2(2.25)	2(2.25)	2(2.25)	8(8.99)	2(2.25)	10(11.24)	0	14(15.73)	3(3.37)	5(5.62)	1(1.12)	5(5.62)	3(3.37)	7(7.87)	0	12(13.48	3(3.37)	6(6.74)	15	13.138
MAD	1(1.12)	1(1.12)	3(3.37)	2(2.25)	4(4.49)	5(5.62)	4(4.49)	9(10.11)	4(4.49)	1(1.12)	12(13.48)	5(5.62)	3(3.37)	1(1.12)	5(5.62)	7(7.87)	4(4.49)	2(2.25)	8(8.99)	7(7.87)	15	11.754
DK	2(2.25)	0	0	5(5.62)	5(5.62)	0	3(3.37)	14(15.73)	5(5.62)	0	0	17(19.10)	4(4.49)	0	0	12(13.48	1(1.12)	0	0	20(22.47	10	14.020

Peer Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	0	2(2.25)	1(1.12)	4(4.49)	1(1.12)	12(13.48)	1(1.12)	8(8.99)	0	16(17.98)	1(1.12)	5(5.62)	1(1.12)	6(6.74)	1(1.12)	8(8.99)	0	8(8.99)	3(3.37)	10(11.24	15	13.454
Imp. P/M	0	2(2.25)	1(1.12)	4(4.49)	1(1.12)	7(7.87)	2(2.25)	12(13.48)	0	14(15.73)	1(1.12)	7(7.87)	0	3(3.37)	0	13(14.61	0	8(8.99)	3(3.37)	10(11.24	15	17.150
MAD	0	2(2.25)	0	5(5.62)	2(2.25)	6(6.74)	0	14(15.73)	3(3.37)	8(8.99)	3(3.37)	8(8.99)	0	1(1.12)	2(2.25)	13(14.61	1(1.12)	2(2.25)	3(3.37)	15(16.85	15	17.588
DK	2(2.25)	0	2(2.25)	3(3.37)	7(7.87)	0	2(2.25)	13(14.61)	9(10.11)	1(1.12)	0	12(13.48)	5(5.62)	1(1.12)	0	10(11.24	4(4.49)	0	1(1.12)	16(17.98	15	15.644

Self Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Use	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	IU	SBU	I/SBU	NR	df	X2
Imp. Tch.	1(1.12)	2(2.25)	2(2.25)	2(2.25)	3(3.37)	7(7.87)	10(11.24)	2(2.25)	2(2.25)	8(8.99)	10(11.24)	2(2.25)	2(2.25)	2(2.25)	8(8.99)	4(4.49)	1(1.12)	6(6.74)	9(10.11	5(5.62)	15	11.441
Imp. P/M	0	3(3.37)	2(2.25)	2(2.25)	4(4.49)	5(5.62)	4(4.49)	9(10.11)	1(1.12)	9(10.11)	6(6.74)	6(6.74)	2(2.25)	2(2.25)	4(4.49)	8(8.99)	0	7(7.87)	6(6.74)	8(8.99)	15	13.077
MAD	1(1.12)	2(2.25)	0	4(4.49)	3(3.37)	4(4.49)	2(2.25)	13(14.61)	2(2.25)	2(2.25)	8(8.99)	10(11.24)	1(1.12)	2(2.25)	3(3.37)	10(11.24	2(2.25)	2(2.25)	5(5.62)	12(13.48	15	9.899
DK	2(2.25)	0	1(1.12)	4(4.49)	6(6.74)	0	1(1.12)	15(16.85)	3(3.37)	0	0	19(21.35)	2(2.25)	0	0	14(15.73	2(2.25)	0	0	19(21.35	10	10.899

Note: Imp. Tch.=Improve Teaching; Imp.P/M=Improve Program/Major; MAD=Make Administrative Decisions; DK=Don't Know; IU=Is Used; SBU=Should Be Used; I/SBU=Is & Should Be Used; NR=No Response; --=Insufficient Items in Cells to Make Chi Square Calculations

Appendix E59. A Comparison of Instructor Responses, by Total Years Teaching Experience, to the Perceived Disposition of Information from the Four Types Evaluation Practiced at Belmont University

Student (Per Evaluation cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	4(4.49)	0	2(2.25)	1(1.12)	14(15.73)	1(1.12)	4(4.49)	3(3.37)	12(13.48)	1(1.12)	9(10.11)	0	11(12.36	1(1.12)	2(2.25)	2(2.25)	10(11.24	1(1.12)	8(8.99)	2(2.25)	15	17.112
RFR	0	4(4.49)	0	3(3.37)	1(1.12)	8(8.99)	1(1.12)	12(13.48)*	1(1.12)	7(7.87)	1(1.12)	13(14.61)	0	7(7.87)	1(1.12)	8(8.99)	0	8(8.99)	1(1.12)	12(13.48	15	4.654
NOR	0	0	0	7(7.87)	0	0	0	22(24.72)	4(4.49)	0	0	18(20.22)	1(1.12)	1(1.12)	0	14(15.73)	0	1(1.12)	0	20(22.47	10	12.473
NOTH.	0	0	0	7(7.87)	0	3(3.37)	0	19(21.35)	0	1(1.12)	0	21(23.60)	0	0	0	16(17.98	0	0	0	21(23.60	5	6.401

Administrative Evaluation (Per cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	2(2.25)	1(1.12)	1(1.12)	3(3.37)	4(4.49)	0	2(2.25)	16(17.98)	5(5.62)	0	5(5.62)	12(13.48)	3(3.37)	0	1(1.12)	12(13.48	4(4.49)	4(4.49)	2(2.25)	11(12.36	15	16.569
RFR	0	3(3.37)	0	4(4.49)	1(1.12)	9(10.11)	3(3.37)	9(10.11)	1(1.12)	7(7.87)	2(2.25)	12(13.48)	0	5(5.62)	3(3.37)	8(8.99)	0	8(8.99)	2(2.25)	11(12.36	15	5.871
NOR	0	0	0	7(7.87)	3(3.37)	0	0	19(21.35)	4(4.49)	2(2.25)	0	16(17.98)	2(2.25)	0	1(1.12)	13(14.61	7(7.87)	0	1(1.12)	13(14.61	15	15.180
NOTH.	1(1.12)	0	0	6(6.74)	3(3.37)	0	1(1.12)	18(20.22)	2(2.25)	1(1.12)	0	19(21.35)	4(4.49)	0	0	12(13.48	1(1.12)	0	0	20(22.47	15	10.005

Peer (Per Evaluation cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	0	1(1.12)	1(1.12)	5(5.62)	1(1.12)	1(1.12)	1(1.12)	19(21.35)	3(3.37)	3(3.37)	1(1.12)	15(16.85)	0	1(1.12)	1(1.12)	14(15.73	1(1.12)	4(4.49)	1(1.12)	15(16.85	15	8.588
RFR	0	2(2.25)	0	5(5.62)	0	7(7.87)	0	15(16.85)	1(1.12)	6(6.74)	0	15(16.85)	0	4(4.49)	0	12(13.48	0	4(4.49)	0	17(19.10	10	4.458
NOR	0	0	0	7(7.87)	3(3.37)	0	0	19(21.35)	2(2.25)	1(1.12)	1(1.12)	18(20.22)	4(4.49)	0	0	12(13.48	2(2.25)	0	0	19(21.35	15	10.005
NOTH	2(2.25)	0	0	5(5.62)	8(8.99)	0	3(3.37)	11(12.36)	8(8.99)	0	1(1.12)	13(14.61)	6(6.74)	0	0	10(11.24	3(3.37)	0	1(1.12)	17(19.10	10	8.797

Self (Per Evaluation cent)

	1-5 Years				6-10 Years				11-15 Years				16-20 Years				20+ Years					
Disp	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	ID	SBD	I/SBD	NR	df	X2
RINR	1(1.12)	0	0	6(6.74)	2(2.25)	0	2(2.25)	18(20.22)	7(7.87)	1(1.12)	1(1.12)	13(14.61)	2(2.25)	1(1.12)	1(1.12)	12(13.48	2(2.25)	0	2(2.25)	17(19.10	15	10.114
RFR	0	2(2.25)	0	5(5.62)	0	5(5.62)	4(4.49)	13(14.61)	1(1.12)	6(6.74)	0	15(16.85)	0	3(3.37)	0	10(11.24	7(7.87)	2(2.25)	2(2.25)	12(13.48	15	10.388
NOR	1(1.12)	0	1(1.12)	5(5.62)	3(3.37)	0	0	19(21.35)	4(4.49)	1(1.12)	0	17(19.10)	2(2.25)	0	1(1.12)	13(14.61	3(3.37)	0	2(2.25)	16(17.98	15	8.544
NOTH.	3(3.37)	0	0	4(4.49)	5(5.62)	0	1(1.12)	16(17.98)	5(5.62)	1(1.12)	1(1.12)	15(16.85)	3(3.37)	0	1(1.12)	12(13.48	3(3.37)	0	2(2.25)	16(17.98	15	6.919

Note: RINR=Return to Instructor/No Response Requested; RFR=Return to Instructor for Response; NOR=Not Returned, NOTH.=Nothing; ID=Is Done; SBD=Should Be Done; I/SBD=Is & Should Be Done; NR=No Response

Appendix E60. A Comparison of Instructor Responses, by Total Years Teaching Experience, to the Perceived Effectiveness of the Four Types of Evaluation Practiced at Belmont University in Improving Teaching, Improving Program/Major, and in Making Administrative Decisions

Improving Teaching (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	0	2(2.25)	3(3.37)	1(1.12)	1(1.12)	0	11(12.36)	7(7.87)	3(3.37)	1(1.12)	3(3.37)	9(10.11)	8(8.99)	2(2.25)	0	1(1.12)	7(7.87	7(7.87)	0	1(1.12)	1(1.12)	6(6.74	11(12.36	2(2.25	1(1.12)	20	28.277
Adm.	0	3(3.37)	1(1.12)	2(2.25)	1(1.12)	0	6(6.74)	6(6.74)	6(6.74)	4(4.49)	0	9(10.11)	7(7.87)	4(4.49)	2(2.25)	0	10(11.24	3(3.37	1(1.12)	2(2.25	1(1.12)	6(6.74	8(8.99	2(2.25	4(4.49	20	18.881
Peer	0	1(1.12)	1(1.12)	1(1.12)	4(4.49)	1(1.12)	4(4.49)	4(4.49)	4(4.49)	9(10.11)	1(1.12)	3(3.37)	7(7.87)	5(5.62)	6(6.74)	1(1.12)	4(4.49	2(2.25	4(4.49	5(5.62	1(1.12)	2(2.25	4(4.49	3(3.37	11(12.36	20	8.053
Self	0	3(3.37)	2(2.25)	1(1.12)	1(1.12)	2(2.25)	8(8.99)	4(4.49)	3(3.37)	5(5.62)	2(2.25)	11(12.36)	5(5.62)	1(1.12)	3(3.37)	1(1.12)	9(10.11	3(3.37	2(2.25	1(1.12)	2(2.25	10(11.24	7(7.87	0	2(2.25	20	17.714

Improving Program/Major (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	0	2(2.25)	3(3.37)	1(1.12)	1(1.12)	0	8(8.99)	8(8.99)	4(4.49)	2(2.25)	1(1.12)	7(7.87)	11(12.36)	3(3.37)	0	1(1.12)	6(6.74	8(8.99	0	1(1.12)	0	4(4.49)	10(11.24	6(6.74	1(1.12)	20	25.926
Adm.	0	2(2.25)	2(2.25)	2(2.25)	1(1.12)	0	8(8.99)	3(3.37)	7(7.87)	4(4.49)	0	8(8.99)	7(7.87)	5(5.62)	2(2.25)	0	7(7.87	5(5.62	2(2.25	2(2.25	1(1.12)	5(5.62	7(7.87	5(5.62	3(3.37	20	14.629
Peer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Self	0	3(3.37)	2(2.25)	1(1.12)	1(1.12)	1(1.12)	7(7.87)	4(4.49)	5(5.62)	5(5.62)	1(1.12)	6(6.74)	9(10.11)	3(3.37)	3(3.37)	0	8(8.99	5(5.62	2(2.25	1(1.12)	2(2.25	7(7.87	6(6.74	3(3.37	3(3.37	20	13.751

Making Administrative Decisions (Percent)

	1-5 Years					6-10 Years					11-15 Years					16-20 Years					20+ Years						
Type	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	VE	E	IE	VIE	NR	df	X2
Stu.	0	2(2.25)	1(1.12)	3(3.37)	1(1.12)	0	4(4.49)	9(10.11)	4(4.49)	5(5.62)	0	12(13.48)	5(5.62)	3(3.37)	2(2.25)	0	7(7.87	7(7.87	0	2(2.25	0	3(3.37	10(11.24	6(6.74	2(2.25	15	24.462
Adm.	0	3(3.37)	1(1.12)	2(2.25)	1(1.12)	0	7(7.87)	4(4.49)	3(3.37)	8(8.99)	0	12(13.48)	5(5.62)	2(2.25)	3(3.37)	0	10(11.24	3(3.37	1(1.12	2(2.25	1(1.12)	8(8.99	5(5.62	3(3.37	4(4.49	20	15.906
Peer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Self	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Note. Stu.=Student; Adm.=Administrative; VE=Very Effective; E=Effective; IE=Ineffective; VIE=Very Ineffective; NR=No response

--Insufficient Number of Items in Cells to Make Chi Square Calculations

VITA

Sheridan Clinton Barker was born in Bulls Gap, Tennessee. There he attended public schools, graduating from Bulls Gap High School in 1968.

He then enrolled at Carson-Newman College in Jefferson City, Tennessee where he pursued a degree in English. He was graduated Magna Cum Laude, With Honors in 1972, and enrolled in the Communications program at the University of Tennessee, Knoxville, in the Fall of 1972. He received the Master of Science degree in Communication in 1973, and began teaching in the Morristown, Tennessee School System. He taught at West High School in Morristown, Tennessee for six years, and in the Virginia Community College System for one year.

In 1980, he began a teaching career a Carson-Newman College, Jefferson City, Tennessee, teaching in the Communication Arts Department, and has continued in that capacity until the present.

He has served as national 2nd Vice President, 1st Vice President and President of the Society for Collegiate Journalists, the oldest national honor society for mass communication students in the United States. He is also very active in civic organizations and in the choir and other functions of the church.