

1990

# Instructor performance evaluation: a nationwide survey of performance appraisal in two-year colleges

Lloyd Oral Roettger  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Educational Administration and Supervision Commons](#), and the [Higher Education and Teaching Commons](#)

## Recommended Citation

Roettger, Lloyd Oral, "Instructor performance evaluation: a nationwide survey of performance appraisal in two-year colleges " (1990). *Retrospective Theses and Dissertations*. 11217.  
<https://lib.dr.iastate.edu/rtd/11217>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

5

90

3

5

1

1

2

U·M·I

MICROFILMED 1990

## **INFORMATION TO USERS**

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

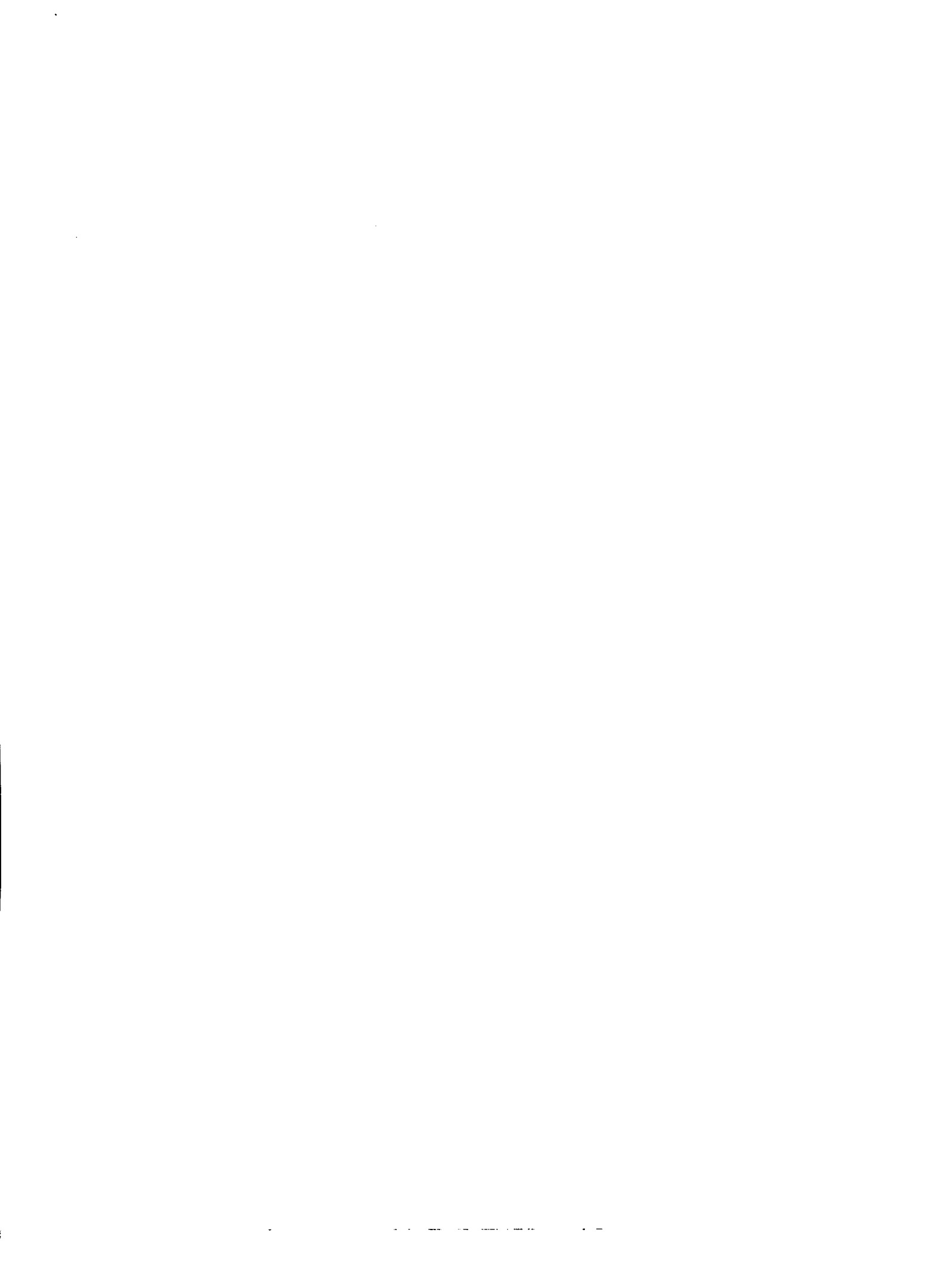
In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# **U·M·I**

University Microfilms International  
A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
313/761-4700 800/521-0600



**Order Number 9035112**

**Instructor performance evaluation: A nationwide survey of  
performance appraisal in two-year colleges**

**Roettger, Lloyd Oral, Ph.D.**

**Iowa State University, 1990**

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



**Instructor performance evaluation: A nationwide survey of  
performance appraisal in two-year colleges**

**By**

**Lloyd Oral Roettger.**

**A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of the  
Requirements for the Degree of  
DOCTOR OF PHILOSOPHY**

**Department: Professional Studies in Education**

**Co-Majors: Education: (Educational Administration and Higher Education)**

**~~Approved:~~**

Signature was redacted for privacy.

**~~In Charge of Major Work~~**

Signature was redacted for privacy.

**~~For the Major Department~~**

Signature was redacted for privacy.

**~~For the Graduate College~~**

**Iowa State University**

**Ames, Iowa**

**1990**

## TABLE OF CONTENTS

	<u>Page</u>
<b>CHAPTER I. THE PROBLEM</b>	<b>1</b>
Introduction	1
Problem Statement	2
Purpose of the Study	3
Objectives of the Study	4
Research Questions	5
Hypotheses	7
Assumptions	8
Delimitations	8
Terminology	9
<b>CHAPTER II. REVIEW OF LITERATURE</b>	<b>21</b>
Introduction	21
Background of the Problem	21
Historical Considerations	21
Legal Considerations	26
Ethical Considerations	31
Applicational Considerations	33
McGreal's Models	40
Common Law Model	41
Goal-setting Model	41
Product Model	42
Clinical Supervision Model	42
Artistic or Naturalistic Model	43



The Personnel Evaluation Standards	44
Approaches to Instructor Performance Appraisal	47
Purposes of Instructor Evaluation Systems	48
Elements of Instructor Evaluation	49
Self-evaluation	50
Peer Review	52
Student Feedback	53
Student Achievement	54
Feedback Conference	54
Supervisor Observation	55
Competency Tests	56
Teacher Artifacts/Portfolios	57
Outside Evaluators	58
Indirect Measures	58
Summary of the Review of Literature	59
CHAPTER III. METHODOLOGY	61
Introduction	61
Population	61
Sample Selection	62
Methodology and Procedures	62
Data Analysis	65
Instrumentation	65
Summary	66
CHAPTER IV. ANALYSIS OF DATA	68
Introduction	68

Data Analysis Results	68
Phase One Questionnaire	68
Phase Two Questionnaire	70
The First Research Question	72
The Second Research Question	86
The Third Research Question	102
The Fourth Research Question	120
The Fifth Research Question	128
The Sixth Research Question	141
The Seventh Research Question	143
The Eighth Research Question	143
The Ninth Research Question	145
The Tenth Research Question	151
Summary	162
CHAPTER V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	167
Summary	167
Findings	169
Conclusions	187
Limitations	194
Discussion	195
Recommendations for Practice	199
Recommendations for Research	200
BIBLIOGRAPHY	203
ACKNOWLEDGMENT	222

APPENDIX A. SUMMARY OF THE STANDARDS	223
APPENDIX B. THE INITIAL QUESTIONNAIRE	234
APPENDIX C. THE PRIMARY DATA GATHERING QUESTIONNAIRE	236
APPENDIX D. THE LETTER THAT ACCOMPANIED THE PRIMARY DATA GATHERING QUESTIONNAIRE	241

## CHAPTER I. THE PROBLEM

### Introduction

American higher education is in a continuing state of evolution. As with any evolution, the process of continuous change from one state to another is evident throughout the education community. A new vitality has emerged on the scene in the form of the two-year colleges. To be sure, they've been around for sometime but this past decade has seen them mature into the institutions of higher learning they were intended to be. The excitement, vitality, change, and growth in these colleges can also pose certain problems. With demands for accountability, improved productivity, and quality; the evaluation of two-year college instructors is one of the most challenging of those problems.

The need for instructor evaluation is one of the few areas in which there's agreement among educators. Therefore, there is little need to offer an extensive justification for its existence (McGreal, 1983). While there may be some argument about stated versus implied purposes of evaluation, educators generally agree regarding its general purpose: to safeguard and improve the quality of instruction received by students (Bolton, 1973). The quality of the students' education is directly proportional to the quality of instruction. There is, however, some diversity of thought on how to evaluate the quality of instruction and instructional performance. However, the four key questions that must always be answered are: what are the criteria, how high are the standards, how do you monitor and report progress, and how do you improve performance (Manatt, 1988). Questions arise regarding the effectiveness of the current evaluation practices and approaches; the environmental factors influencing instructor evalua-

tion; the instrumentation presently being used; the instructor evaluation training being conducted, required, and needed; the use of incentive pay schemes; and the efficacy of the overall process. This diversity of thought and the questions that arise lead to a myriad of different evaluation models with an even greater array of instruments and levels of effectiveness. Despite the obvious difficulties, most two-year colleges must have a functioning evaluation system. Whether the mandate is professional, political, contractual, or legislative the two-year college is now expected to be able to point to some systematic procedure to monitor the performance of its instructors. The ultimate question for two-year colleges is then: is there a totally reliable instructor evaluation system and if there is not, what can be done to develop the most effective system possible? It is the purpose of this study to attempt to provide the baseline data necessary to answer that question.

#### Problem Statement

At the time of this study, no evidence can be located that indicates that anyone knows what the two-year colleges in this country are doing in instructor performance appraisal. With the nation concerned with the quality of both public school and postsecondary education, this study was undertaken to produce a set of data which would be used in the development of "state-of-the-art" instructor evaluation systems for the nation's two-year colleges. The problem of this study was to analyze the procedures and processes of instructor evaluation as they exist in 1989 in the target educational institutions and to classify them according to the evaluation models established by McGreal (1983). Additionally, the relationship between the model used and the extent to which it satisfies the

personnel evaluation standards established by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 1988) will be determined and analyzed. A wide variety of variables will be used to maximize the utility of the results.

### Purpose of the Study

The purposes of this study were:

1. To identify who supervises/evaluates whom, in a two-year college. To determine the normal span of control in a two-year college. To identify the duties of instructional leaders in two-year colleges.
2. To determine the types of evaluations being done in the two-year college and classify them according to the McGreal taxonomy.
3. To describe the instructor evaluation processes carried out by the two-year post secondary institutions during the 1988-89 school year.
4. To determine the educational, and technical backgrounds of those doing the supervision/evaluation of instructors.
5. To determine the average amount of evaluator training evaluators have.
6. To establish a listing of the most prevalent guidelines or criteria for instructor evaluation used by local two-year post secondary institutions based on a review of the literature and an analysis of what is currently being used.
7. To establish the relationships, if any, between the evaluation model used and its efficacy at meeting the personnel evaluation standards established by the Joint Committee (1988).

8. To provide a further disaggregation of the data gathered by use of selected variables to maximize the utility of the results.

### Objectives of the Study

The objectives of this study were to:

1. describe and analyze the procedures and processes of instructor evaluation in the U.S.A. according to the following criteria:
  - a. the legislative involvement via state mandates, if any;
  - b. the professional, political, or contractual mandates, if any;
  - c. the purposes and/or uses;
  - d. the span of control of evaluators;
  - e. the instructional leadership duties performed;
  - f. the percentage of time evaluators spent on evaluations;
  - g. the instruments, if any;
  - h. the training, if any, of the evaluator;
  - i. the methodology of data collection;
  - j. the number and types of observations;
  - k. the types and frequency of conferences, if any;
  - l. the demographics about the evaluator, i.e. preparation; highest degree earned, area of degree, perceived need for more evaluator training, etc. ;
  - m. the type of institution;
  - n. the size of institution.
2. classify and evaluate models of instructor evaluation according to those categories established by McGreal (1983).

3. established a rating of quality for each model relative to the specific standards for personnel evaluation established by the Joint Committee (1988).
4. develop a data base of the most prominent practices and approaches used in "state-of-the-art" instructor performance appraisal.
5. determine the variability of the sample, that is; the number supervised and evaluated, the types of instructors supervised and evaluated, and the level of expertise of those supervising and evaluating.

#### Research Questions

1. Who supervises whom and what? What is the span of control? What are each supervisors duties?
  - a. Specifically - Who are the instructional leaders?
  - b. How many instructors do they supervise and/or evaluate?
  - c. What other duties do these instructional leaders have responsibility for?
2. What are the models of evaluation being utilized in the two-year college?
  - a. Do the models used differ by type of institution or institutional size?
  - b. What percentage of two-year colleges are utilizing a "pay for performance", "merit pay", "career ladder", or other type of incentive program?
3. What are the elements of evaluation being used in instructor performance evaluation in two-year colleges?



- a. What percentage of states mandate two-year college instructor evaluation?
  - b. What percentage of two-year colleges require instructor evaluation?
  - c. Is the same instrument utilized for all instructors?
  - d. To what extent is peer evaluation being utilized?
  - e. To what extent is an academic council being utilized in two-year college evaluation?
  - f. To what extent is student feedback being used for evaluation?
  - g. To what extent are student achievement data being used for evaluation?
  - h. Are outside evaluators being used?
4. What are the educational and technical backgrounds of those doing the supervision of instructors?
- a. What percentage of the evaluators have had previous training in pedagogy?
  - b. What are the highest degrees held by those doing instructor evaluation?
5. What training does the evaluator have in evaluation?
- a. How many hours of evaluator training has the average evaluator had?
  - b. What differences, if any, are there between the size and types of institutions in regard to the amount of training evaluators have?
  - c. Is certification in evaluation held by the majority of evaluators?

- d. What is the perceived need for additional training in evaluation?
6. Is the frequency of occurrence of the evaluation models in use dependent on the type of the two-year college?
7. Is the frequency of occurrence of the evaluation models in use dependent on the size of the two-year college?
8. Is there a relationship between the model in use and the type or size of two-year college?
9. What are the efficacy ratings for each personnel standard?
10. Are differences in the efficacy ratings on each personnel standard attributable to the model in use, or the type or size of two-year college?

### Hypotheses

The following hypotheses were tested by this study:

1. The distribution of models in use among the two-year colleges is independent of the five types of two-year colleges.
2. The distribution of models in use among the two-year colleges is independent of the size of two-year colleges.
3. There is no relationship between the model in use and the type or size of two-year college.
4. There is no significant difference in the efficacy rating for each personnel standard between the evaluation models in use.
5. There is no significant difference in the efficacy rating for each personnel standard between the types of two-year colleges.

6. There is no significant difference in the efficacy rating for each personnel standard between the sizes of two-year colleges.

### Assumptions

This study was based upon the following assumptions:

1. The act of imparting knowledge or skill to students, be it called teaching or instruction, is basically the same activity regardless of the grade level attainment of the students receiving the schooling.
2. The respondents to the questionnaire provided complete and accurate information.
3. Instructor evaluation is important in improving instructional effectiveness in the various two-year college teaching settings.
4. A good instructor evaluation procedure is vital in identifying differences in instructional quality.
5. The college presidents and chief executive officers surveyed would be knowledgeable about the most appropriate evaluator to contact.
6. The evaluator contacted was knowledgeable about the evaluation system in their two-year college.

### Delimitations

The following delimitations apply to this study:

1. All the respondents in this study were be administrators in post-secondary two-year colleges. Thus generalizations of the findings to other populations or settings may not be appropriate.

2. This study was be limited to the evaluation documents, evaluation procedures, and evaluation instruments of the instructor evaluation systems for the randomly selected sample of all the two-year colleges as they exist in the 1988-89 school year, as well as any additional data gathered from the questionnaire and the review of literature.

### Terminology

For the purposes of this study, the following definitions apply:

Accountability - the responsibility to be answerable for specific happenings, costs, or performances.

Administration - management of an organization through such actions as planning, staffing, motivating, directing, controlling, communicating, and evaluating.

Artistic or Naturalistic Model - teacher evaluation model that views the act of teaching as an art. Teachers make judgments based largely on qualities that unfold during the course of teaching; teaching is influenced by contingencies that are unpredictable; the ends achieved in teaching are often created in process (Eisner, 1982).

Assessment - the act of rating or describing a subject on some variable of interest (Joint Committee, 1988).

Assessment procedure - any method used to rate or describe some characteristic of a subject (Joint Committee, 1988).

Audit (of an evaluation) - an independent examination and verification of the quality of an evaluation plan, the adequacy of its implementation, the accuracy of results, and the validity of conclusions.

Behavior - specific, observable actions of an individual in response to internal and external stimuli (Joint Committee, 1988).

Career ladder scale - an incremental pay scale through which a teacher advances as a result of favorable evaluations (Joint Committee, 1988).

Classroom Instructor - any individual assigned to teach students in a specific curriculum area where the primary teaching activity occurs in a schoolroom fitted with traditional desks, tables, blackboards, etc.

Clinical Instructor - any individual assigned to teach students in a specific curriculum area where the primary teaching activity is associated with a hospital or medical setting.

Clinical Supervision Model - teacher evaluation model that heavily emphasizes a collegial relationship between the teacher and the supervisor. It is an in-class approach which seeks to improve instruction and teacher development. It seeks to assist teachers to modify existing patterns of teaching in ways which the teacher observes (Sullivan, 1980).

Code (data) - to translate information into a set of quantitative or qualitative symbols (Norusis, 1983).

Coefficient - a value expressing the degree to which some characteristic or relation exists in specified instances; e.g., the coefficient of correlation is

a value expressing the degree to which two variables vary concomitantly (Joint Committee, 1988).

Common Law Model - teacher evaluation model characterized by high supervisor-low teacher involvement. It usually contains standardized criteria, forces comparative judgments, and emphasizes summative evaluation (McGreal, 1983).

Community College - a comprehensive, community-centered two-year college that sees its mission as that of providing educational opportunities and community services for all students. This includes occupational career education and college transfer education.

Competency - a skill, knowledge, or experience that is suitable or sufficient for some purpose.

Conclusions (of an evaluation) - final judgments and recommendations (Joint Committee, 1988).

Conferencing - the skills of holding a conference with an employee for the purposes of discussing, deliberating, and sharing gathered performance appraisal data. These skills may include praise, offering positive and negative feedback, counseling, conversing, conferring, and reprimand.

Contextual variables - indicators or dimensions that are useful in describing the facts or circumstances that surround a particular job situation and influence a person's performance on that job (Joint Committee, 1988).

Correlation - the degree to which two or more sets of measurements vary together; e.g., a positive correlation exists when high values on one scale are associated with high values on another (Hinkle et al., 1979).

Criteria of the evaluation - factors on which an employee is evaluated (Glueck, 1982).

Criterion - a measure of job performance, such as productivity, accident rate, absenteeism, reject rate, or training score. It also includes subjective measures such as supervisory ratings (Joint Committee, 1988).

Data - material gathered during the course of an evaluation which serves as the basis for information, discussion, and inference (Joint Committee, 1988).

Data analysis - the process of studying data to arrive at answers to questions (Joint Committee, 1988).

Data collection procedures - any set of steps used to obtain quantitative or qualitative information about the qualifications or performance of an individual (Joint Committee, 1988).

Dependent variable - a measure (e.g., a student's performance on a test) that is assumed to vary as a result of some influence (often taken to be the independent variable), such as a student's instructional experience (Hinkle et al., 1979).

Development and implementation - refers to the steps used in the evolution and execution of an instructor evaluation program.

**Evaluated** - the person whose performance is being evaluated.

**Evaluation** - systematic investigation of the worth or merit of something; e.g., a person's qualifications or performance in a given role (Joint Committee, 1988).

**Evaluation system** - a regularized structure and set of procedures by which an institution initiates, designs, implements, and uses evaluations of its personnel or programs (Joint Committee, 1988).

**Evaluator** - the person who has responsibility for planning, conducting, and reporting evaluations.

**Feasibility** - the extent to which an evaluation is appropriate and practical for implementation (Joint Committee, 1988).

**Formative evaluation** - is ongoing, descriptive, nonjudgmental, and performed to help to teachers teach better (Manatt, 1988).

**Gain scores** - the difference between a person's performance on a test and his or her performance on a subsequent administration of the same test (Joint Committee, 1988).

**Generalizability** - the extent to which information collected in one setting about a program, project, or instructional material can be used to reach a valid prediction of its utility and reliability in other settings (Joint Committee, 1988).

**Goals** - statements cited as goals, objectives, and purposes of a teacher evaluation program.



Goal-setting Model - teacher evaluation model that emphasizes an individualized approach to evaluation. Its steps include setting goals, working toward those goals, and reviewing progress toward the goals (McGreal, 1983).

Incentive pay - compensation paid to employees for doing different kinds or amounts of work. Incentive pay plans may open new opportunities for professional development, or they may increase the volume of work tasks. Although some such plans require that employees be judged meritorious in order to participate, incentive pay differs from merit pay (Joint Committee, 1988).

Informed consent - agreement by the participants in an evaluation that their names and/or confidential information supplied by them may be used in specified ways, for stated purposes, and in light of possible consequences prior to the collection and/or release of this information in evaluation reports (Joint Committee, 1988).

Instructor - a teacher.

Instructor evaluation process - the formal and systematic approach utilized in assessing the competence of an instructor.

Instrument - an assessment device adopted, adapted, or constructed for the purposes of the evaluation (Joint Committee, 1988).

Junior College - an educational institution offering a two-year course that is generally the equivalent of the first two-years of a four-year undergraduate course.

Laboratory Instructor - any individual assigned to teach students in a specific curriculum area where the primary teaching activity occurs in a

schoolroom equipped for instruction in scientific experimentation, research, or testing.

Level of significance - a predetermined probability value used to decide whether the results occurred by chance (Joint Committee, 1988).

Mean - the arithmetic average of a set of numbers.

Measurement - a process of determining the level of performance on specific variables or criteria.

Merit - excellence as assessed by intrinsic qualities or performance (Joint Committee, 1988).

Merit pay - monetary compensation in the form of higher wages or salaries awarded to deserving employees - who may have the same job descriptions and responsibilities as other employees not receiving merit pay - on the basis of verifiable superiority in the quality of their work performance. The differences in compensation, which may be one time bonuses or permanent pay increases, are usually based on annual systematic evaluations of employee performance (Joint Committee, 1988).

Norm - a single value, or a distribution of values, constituting the typical performance of a given group (Joint Committee, 1988).

Objective evaluation - evaluation carried out in a way that minimizes error or bias due to the predilections of the evaluator (Joint Committee, 1988).

Observability - the extent to which one can make systematic or scientific observation of phenomenon about personnel performance.

Outside evaluator - an evaluator from outside the two-year college.

Peer - person working with and at the same level as the employee (Glueck, 1982).

Performance standard - a formal specification of the expected level of achievement in fulfilling a performance objective (Joint Committee, 1988).

Personnel evaluation - the systematic assessment of a person's performance and/or qualifications in relation to a role and some specified, defensible institutional purpose (Joint Committee, 1988).

Personnel evaluation system - all of the rules, procedures, assignments, and other elements that an institution uses to evaluate its personnel (Joint Committee, 1988).

Post-test - an instrument or process designed to assess performance after the administration of a program or instructional material.

Pre-test - an instrument or process designed to assess performance before the administration of a program or instructional material.

Predictor - a measurable characteristic used to predict criterion performance; e.g., scores on a test or the judgments of interviewers (Joint Committee, 1988).

Procedures - specific steps used in the evaluation of instructors.

Process - a series of actions or operations used in achieving something.

Product Model - teacher evaluation model that utilizes various measures of student achievement as a measure of teacher competence (McGreal, 1983).

Productivity - the rate at which one can produce the desired results.

Propriety - the extent to which an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation as well as those affected by its results (Joint Committee, 1988).

Quality - a characteristic or attribute of something as it relates to excellence.

Qualitative information - facts and claims presented in narrative, not numerical, form (Joint Committee, 1988).

Quantitative information - facts and claims represented by numbers (Joint Committee, 1988).

Random sampling - drawing a number of items of any sort from a larger group or population, so that every individual item has the same (and independent) chance as any other to be chosen (Joint Committee, 1988).

Rating - a place assigned on a scale.

Reliability - the relative extent to which the measurement procedures assign the same value to a characteristic of an individual each time that it is measured under essentially the same circumstances (McCall, 1986).

Sample - a part of a population.

Score - any specific value in a range of possible values describing the assessment of an individual.

Self-report instrument - a device in which persons make and report judgments about their own performance (Joint Committee, 1988).

Shop Instructor - any individual assigned to teach students in a specific curriculum area where the primary teaching activity occurs in a school-room fitted with machinery and tools for instruction in the manual arts.

Significant difference (statistically) - an observed difference between two statistics that probably did not occur by chance (Joint Committee, 1988).

Standard - a principle commonly agreed to by experts in the conduct and use of evaluation by which to measure the value or quality of an evaluation (Joint Committee, 1988).

Standard score - a score that describes the location of a person's score within a set of scores in terms of distance from the mean in standard deviation units; it may include scores on certain derived scales (Joint Committee, 1988).

Statistic - a summary number used typically to describe a characteristic of a sample.

Subjective evaluation - an evaluation not open to verification by others not using public or communicable standards (Joint Committee, 1988).

Summative evaluation - at the end of a formative cycle, is comparative, judgmental and if the teacher is a subpar performer, it may be adjudicative (Manatt, 1988).

Teacher - any individual assigned to a specific grade, level and/or curriculum area with one or more regularly assigned student groups (Harris, 1986).

Teacher competence - refers to the repertoire of competencies a teacher possesses. Overall competence is a matter of the degree to which a teacher has mastered a set of individual competencies, some of which are more critical to a judgment of overall competence than others (Medley, 1979).

Teacher competency - refers to any single knowledge, skill, or professional value position, the possession of which is believed to be relevant to the successful practice of teaching. Competencies refer to specific things that teachers know, do, or believe but not to the effects of these attributes on others (Medley, 1979).

Teacher effectiveness - refers to the effect that the teacher's performance has on pupils. Teacher effectiveness depends not only on competence and performance, but also on the responses pupils make. Just as competence cannot predict performance under different situations, teacher performance cannot predict outcomes under different situations (Medley, 1979).

Teacher performance - refers to what the teacher does on the job rather than to what she or he can do (that is, how competent she or he is). Teacher performance is specific to the job situation; it depends on the competence of the teacher, the context in which the teacher works, and

the teacher's ability to apply his or her competencies at any given point in time (Medley, 1979).

Technical College - a regional two-year college that provides highly technical training to learners that are pursuing a technician level career. Some learners may transfer to four-year colleges for further training as technologists.

Utility - the extent to which an evaluation will serve the relevant information needs of evaluatees and other users (Joint Committee, 1988).

Validity - the extent to which the measurement procedures accurately reflect the variable being measured (McCall, 1986).

Variability - the extent to which scores in a distribution deviate from central tendency or from one another (McCall, 1986).

Variable - a characteristic that can take on different values.

Vocational Trade College - a localized post-secondary institution that provides trade related training, apprenticeship training, and basic skills education for those individuals primarily concerned with short term job readiness training.

Vocational-Technical College - an area-wide two-year college that is devoted primarily to occupational training of a practical, technical, and semi-technical nature. Little or no college transfer education takes place.

## CHAPTER II. REVIEW OF LITERATURE

*"Teaching is too important to too many to be conducted without a critical inquiry into its worth"*

Millman, 1981

## Introduction

There have been many words put to paper about the evaluation of those who teach. Many opinions exist as to the value of, need for, and purpose of faculty evaluation. Indeed, the casual observer would have difficulty identifying commonalities among evaluation systems. However, there does appear to be some central elements of the evaluation process. This study will briefly investigate the historical, legal, ethical, and applicational considerations of those processes.

## Background of the Problem

Historical Considerations

The development of the two-year colleges of this country has spanned a great number of years. Historians believe the roots of the two-year, postsecondary institutions to be entangled with our need to expand educational opportunities beyond the secondary level. During our period of rapid population growth, increased secondary school completion rates, emerging technological advancements, and a growing information base the universities could not accommodate the large and diverse population of new students. As a result, the two-year, postsecondary institutions were spawned and have grown to be a major contributor to the American system of higher education.



Since the roots of the two-year colleges are in the secondary school systems it follows that the faculty evaluation processes are also rooted there. Among those processes are administrative classroom observation, lesson plan analysis, methods review, and others. However, as the process of evolution continued, community college faculty started bringing in the model of peer evaluation from their colleagues in the four-year colleges and universities. In addition, many of them added student feedback to the process. Instructional improvement was a self-directed activity. Only those motivated to improve chose to act on the feedback of peers, administrators and students. Seldom, if ever, were poor performing instructors called to task. Generally, any evidence of classroom performance or student achievement satisfied evaluators. The reason for this was probably the uncertainty of criteria for evaluation, the lack of standards, or the absence of adequate monitoring and reporting schema. As the early use of these various processes of evaluation grew there was no consistency from college to college and sometimes no consistency from faculty member to faculty member within a specific college.

As partnerships between two-year, postsecondary institutions and business and industry became more commonplace, the administrators of educational institutions saw that the organized, systematic evaluation procedures used by their counterparts in the private sector were effective and valuable. Hence, the two-year, postsecondary institutions adopted and adapted various processes of the business and industry models. In fact, faculty evaluation was and still is considered a measure of organizational productivity. Ratcliff (1984) concluded that faculty evaluation can and should be a means for gathering information on the productivity of a community college.

Research on the current systems, processes and models of faculty evaluation is fragmented at best. A few case studies of isolated elements of various existing evaluation procedures have found their way into the journals but by and large no attempt has been made to collect data on a national level or to identify current practices throughout the network of two-year, postsecondary institutions.

Historically, there are three approaches to the evaluation of personnel: (1) the characteristics of the individual, (2) the products attributed to the individual, and (3) the processes used by the individual (Harris et al., 1979). When referring to the characteristics of the individual, Harris had in mind such relatively easy-to-measure qualities as knowledge of the subject or accepted professional practices, grade-point averages, college hours or degrees held, and years of professional experience. Some of these items, such as, grade-point averages and knowledge, are legitimate considerations at the time of initial employment but not in evaluating a person on the job. Evaluation of personnel on a product basis has a lot of appeal. In fact, it is the only direct way to evaluate an individual's performance because the other methods are based on the assumption of a reasonably high correlation with products. In the long run, products are what schooling is all about, so we must validate whatever we do by showing their relationships to outcomes.

The terms teacher and instructor have sometimes been used to denote the positional difference between public school faculty and two-year, postsecondary faculty. It is my contention that the activities of teachers and instructors are similar enough that their titles can be interchanged and this review of literature will not attempt to maintain a strict alignment with the traditional semantics.

The evaluation of teacher/instructor performance has been the topic of various books, the theme of numerous conferences, and the focus of countless articles. Certainly, the question of how to appraise the quality of teaching and teachers has been raised for as many years as there have been organized educational institutions. Popham (1975) attempts to put educational evaluation in perspective:

Most thinking people have always praised evaluation, at least in the abstract, as an intellectually defensible activity. Only the charlatan or the incompetent, it would seem has reason to fear the effects of evaluation. Through the centuries, therefore, our most capable scholars have recommended that human beings engage in evaluative operations.... Evaluation has historically been viewed, and quite properly so, as an integral activity of a rational approach to life.

Teacher performance evaluation in public schools is a relatively recent phenomenon. A recent review of the history of teacher evaluation noted that there was very little serious attention given to this topic until the 1970s and that the previous two decades were marked by self-evaluation, ceremonial congratulations, and neglect (Darling-Hammond, Wise, and Pease, 1983). In the years before the Eisenhower presidency, the teacher was seldom observed in the teaching process by an administrator. Likewise peer evaluation or other methods were basically not in use. Occasionally some zealous school board member would pay the classroom a visit, but probably had no training in teacher evaluation. From these humble beginnings teacher evaluation has evolved into various processes and systems. The current reform movements have placed additional demands

on improving the outcomes of schooling with teacher evaluation targeted as a focal point for these betterment efforts.

The intensive national scrutiny currently being performed and applied in the quest for better schools will likely provide an approach to teacher performance evaluation which maximizes accountability. However, there is much concern that this objective may not be reached jointly with the broader goal of school improvement. Researchers Darling-Hammond, Wise, and Pease (1983) explain the external pressures on schools and internal environmental organizational factors contribute to a school's receptivity and willingness to design a teacher performance evaluation system. They argue that "successful teacher evaluation requires consistent and shared views of the teaching-learning process and of the organizational context in which teacher evaluation takes place". The importance of similar "enabling conditions" were described by McLaughlin and Pfeifer (1986) when they hinged a successful evaluation process on a district's organizational environment. They believe the success of a teacher performance evaluation process depends upon the degree to which a district's organizational environment exhibits: "mutual trust between teachers and administrators, open channel communication, commitment to individual and instructional learning, visibility of evaluation activities and associated learning efforts".

The historical ties between public schools and postsecondary two-year colleges would indicate that this same press for accountability, for educational improvement, for mutual trust, and for consistent standards will be shared by both educational levels. This strong linkage requires that this review of literature consider not only the pertinent two-year college materials but also some selected public school literature.

Much of the significant research focused on teacher evaluation has been conducted at Iowa State University in the department of professional studies, educational administration section. This research has been centered around the four key questions of teacher evaluation cited earlier and the question of who should be evaluating whom. The research focused on the query about who should evaluate is directly or indirectly addressed by several researchers (Pinckney, 1982; Look, 1983; Darnell, 1984; Judkins, 1987; Mueller, 1987; Lueders, 1987; and Dzyacky, 1988). Some of the answers to the question of criteria can be found in a number of unpublished doctoral dissertations produced within the educational administration section of Iowa State University (Cameron, 1973; Hidlebaugh, 1973; Frudden, 1980; Allen, 1985; Noriega, 1987; and Uhl, 1988). Standards have long been a topic of study. At least six Iowa State University researchers sought to contribute knowledge in this area (Cameron, 1973; Garcia, 1980; Rucker, 1981; Westerberg, 1983; Harrington, 1984; and Peterson, 1988). How one monitors and reports progress has been the topic of another group of researchers (Edwards, 1985; Semones, 1987; Stevenson, 1987; Floden, 1987; and Wicks, 1988). Finally, a good deal of research has focused on attempting to answer the question of how to improve instructional performance (Faast, 1982; Rauhauser, 1983; Schycker, 1983; Mitchell, 1985; Spencer, 1985; Licklider, 1986; Nance, 1986; and Rice, 1986).

### Legal Considerations

Correctly performed instructor evaluation has successfully been upheld in the courts on numerous occasions (Lovain, 1984; Piele, 1979; Piele, 1981; Piele, 1983; Strike and Bull, 1981). Most of the legal actions have been centered around

various terminations of faculty members, both non-tenured and tenured. Academic administrators have gained support in the courts in recent years. Indeed, many of the legal decisions support administrative decisions to terminate incompetent tenured faculty. Such support for terminating tenured faculty should destroy the myth that tenure makes it impossible to dismiss faculty.

Strike and Bull (1981) summarize five indirect implications relative to teacher evaluation procedures which may lead to the termination of teachers:

First, evaluations should be undertaken on a regular basis. Second, permanent records of the results of evaluations should be maintained. Third, teachers should be informed of the evaluation results, including access to their personnel file. These policies allow evaluation to be used to establish the irremediable nature of the defects upon which a judgment of incompetence is based. Fourth, teachers should be given an opportunity to enter explanations and clarifications of or objections to particular evaluation findings into their personnel record at the time at which those findings are filed. Fifth, the evaluation records should be kept confidential (p. 339).

While the courts have acknowledged tenure to be a property interest of a faculty position -- with the assurance that dismissal procedures must conform to the requirements of due process -- they have also affirmed with consistency that colleges and public schools can terminate faculty for adequate cause. They have, however, affirmed such cases only where they have found that proper procedures were followed and evidence of a sound evaluation process.

The intent of faculty evaluation in two-year, postsecondary institutions has been to make instructors aware of their strengths and weaknesses, with the expectation that they would modify their behavior. The concept of faculty eval-

uation has evolved from being a threat to becoming a common and expected occurrence. Twenty years ago, measuring faculty performance varied from college to college and was even non-existent in some colleges. However, at a time when collective bargaining is popular; when questions of legality prevail; and when retention, accountability, and effectiveness are prevalent concerns, faculty evaluation has become a timely issue.

The art of faculty evaluation in community colleges appears to be no art at all. Cohen and Brawer (1982) point out how, as faculty gained more power as they broke off from the lower schools, they made evaluation plans more complex. In fact, they refer to evaluation procedures becoming more complicated and gaining "labyrinthine complexity".

In *Foleno v. Board of Education of the Township of Bedminster* (Piele, 1979) the court said "The board has the duty, in furnishing a thorough and efficient education, to evaluate the performances of its employees and to staff its classrooms with skillful and effective teachers" (p. 11). Even though the *Foleno v. Board of Education of the Township of Bedminster* decision was at the public school level its precedent reaches the postsecondary level in that the board has the duty to evaluate its employees and to staff its classrooms with competent, skilled teachers.

One of the many purposes of evaluation is to remove incompetent faculty members. The following is a summary of court cases that supported administrative movement to terminate faculty for reasons of incompetence. Such cases break down the stereotype that tenure protects incompetent faculty from dismissals. The Missouri Supreme Court considered the dismissal of a tenured junior college instructor for incompetency, inefficiency, and insubordination in

*Saunders v. Reorganized School District No. 2 of Osage County*. The court upheld the dismissal and held that the charge of inefficiency was supported by evidence of the plaintiff's manner of teaching (Lovain, 1984). In *Chung v. Park*, a tenured instructor was dismissed on grounds of intransigence in dealing with his supervisors, especially because of poor teaching. The federal district court held that the allegations were supported by substantial evidence (Lovain, 1984). Even at the four year university level poor teaching is grounds for dismissal. In *Jarva v. Fayetteville University*, a tenured professor was dismissed for, *inter alia*, poor teaching, being unprepared for classes, and poor relations with students. His civil rights challenge to the dismissal was rejected by the court (Lovain, 1984).

The above cases were summarized by Lovain, who points out that the cases show "the courts will defer to the expertise of academic administrators" when charges of incompetence in teaching "are supported by substantial and relevant evidence" (p. 423).

Piele has summarized a number of other cases relative to instructor performance appraisal. Excerpts of summary statements include:

...negotiation of evaluative criteria is against public policy because retention or promotion of teachers is a management prerogative (Piele, 1979).

...it is settled by law in most states that only the board hires and only the board fires (Piele, 1979).

...on the basis that substantial evidence of unfitness to teach, particularly evidence of lack of student progress, the court supported the administrative hearing record and upheld termination of a nineteen year tenured instructor (Piele, 1981).

...in reviewing the entire record, the high court found that substantial evidence supported four major teaching deficiencies: (1) exces-



sive use of worksheets, (2) lack of rapport, (3) lack of appropriate student discipline, and (4) lack of student progress. The latter basis, lack of student progress, was specifically related to express statutory grounds for discharge under Minnesota law (Piele, 1983).

...in affirming dismissal, the appellate court noted that the instructor's deficiencies were long standing and represented fundamental teaching inadequacies. The notice provided was appropriate and the period of remediation was reasonable for correction of the deficiencies. The court's reliance on classroom observation reports illustrates the considerable weight courts give to the evidentiary value of these records (Piele, 1983).

It now becomes extremely important that a well documented case leading to a board of trustees decision to fire a faculty or academic support person not be lost because of procedural mismanagement. If a state has a legislated tenure law spelling out the "Dismissal of Tenured Faculty Member for Cause," it should be followed step-by-step. For example, the Illinois Community College Tenure Law has such a provision and has the following requirements in it:

1. The board must first approve a motion by a majority vote of all its members.
2. The specific charges for dismissal shall be confidential but shall be issued to the faculty member upon request.
3. The board decision shall be final unless the tenured faculty member within ten days requests in writing to the board that a hearing be scheduled.
4. Such notice shall contain a bill of particulars.
5. All testimony at the hearing shall be taken under oath administered by the hearing officer.
6. The hearing officer shall, with reasonable dispatch, make a decision as to whether or not the tenured member shall be dismissed

and shall give a copy of the decision to both the tenured faculty member and the board.

7. The decision of the hearing officer shall be final and binding (Andrews, 1985).

There are other procedural statements in the law as to how to select a hearing officer, subpoena witnesses, etc.

### Ethical Considerations

Knowledge of ethical issues that include due process, discrimination, validity, reliability and inference will assist educational administrators in reworking evaluation systems to meet legal and ethical requirements (Peterson, 1983). According to Schurr (1982), "A code of ethics must legitimate the interests of academics by showing, at least, that good academic practice is auditable, serves society at large, and avoids conflicts of interest; is instructionally effective and evaluatively responsible; and is differentiated according to specializations, whose competencies are certified and relationships specified."

Ethics and, increasingly, the law require certain other steps when possibly unfavorable personnel action is contemplated. These include:

- (1) a chance to review the evidence and react to it
- (2) a chance to scrutinize the chain of argument from the evidence to the unfavorable conclusion
- (3) advance warning that provides time for improvement and a clear description of what degree of improvement will be satisfactory (This should not be taken to mean that the administration must provide a sure-fire remedy -- because that is not always possible -- but only a clear definition of what would constitute acceptable performance.)

- (4) the above events to be recorded as having occurred on specified dates, preferably with the log signed of by both parties (This is the "audit trail" requirement.)
- (5) since age will be disallowed as a criterion, we urgently need a sound evaluation system that can be used throughout the tenured years. Often a reduction in load, and salary, and a change of title is the kindest move, but it must be done using a system applied to everyone regardless of age, or it will be disallowed, and rightly so. "Applied" means enforced, pantomimed, on younger tenured teachers (Scriven, 1981).

A particularly important, but largely unstudied, problem of ethics as they pertain to the work of the educational evaluator is the consideration of the appropriate course of action to take when approached with inducements to bias, change or eliminate the results of an evaluation. There are all sorts of sticky decisions to be made by the evaluator. Some of these will unquestionably involve moral choices. Some of the moral implications of the decision alternatives to be encountered are so obvious that any person would know the difference between right and wrong. For instance, is it correct to suppress the results of a negative evaluation study because an instructor (who would perhaps be discharged as a consequence) offers the free summer use of a lake cottage in Canada? Or how about the meat from a deer he has hunted or new curtains for your office she has sewn? A moral person would reach a quick negative decision on each, some options of the evaluator will involve ethical choices that are not so transparent. Popham (1975) discusses the ethics of surreptitiously gathering measurements on people. "Some persons consider the invasion of privacy to be so intrinsically reprehensible that it is ethically repugnant under any circumstances. Some persons believe that the end (improvement of the quality of education) justifies the

means (surreptitious observation). Other people don't think that there is anything basically evil about collecting measurements surreptitiously."

Ideally, with such debatable ethical questions, the educational evaluator could consult documents in which the pros and cons of various positions have been examined by individuals well versed in the moral implications of such choices. Hopefully, in the future, we will see the appearance of such compilations. But for now, at least, evaluators are pretty much on their own in the ethical arena. The best idea, at least for the present, would be to have the evaluator engage in a fair amount of personal analysis of the moral suitability of various stances, then solicit the reactions of colleagues regarding such issues. By deliberately trying to think through the ethical considerations involved, alone and in consultation, the educational evaluator will probably come up with a more defensible ethical position than would have been adopted under immediate pressures for action.

### Applicational Considerations

It becomes incumbent upon educational evaluators to consider carefully the factors associated with a successful instructor performance appraisal process. A lesson can easily be taken from the experiences of public schools. Many factors have been at play in the public schools for, in some cases, longer than some two-year colleges have been in existence. From a practical viewpoint it is vitally important that training be provided to the instructor and the administrator.

A number of researchers have concluded that training in evaluation is critical to successful instructor performance appraisal (McGreal, 1983; Stiggins and Bridgeford, 1985; Weber, 1987). The emphasis on fair, rigorous, and relevant

evaluations points to the need for well-trained administrative evaluators to carry out such an important charge in the higher education institutions.

No less important a place should be assigned to the formal involvement in the evaluation design phase of what Manatt (1987b) calls stakeholders' committees. These committees are usually composed of teachers, administrators, board members, parents, patrons, and, sometimes, students and serve as the designers of the content and procedures of the teacher performance evaluation. These involvements provide much needed mutual trust and ownership between teacher and administrator -- a key ingredient for successful process design.

Additionally, a report published by the Rand Corporation reported several of the application based problems which have consistently surfaced in the literature. The most significant are summarized below:

1. Differences in teacher evaluation process design were substantial.
2. The researchers agree that certain organizational climate produces a greater readiness for the design of a teacher evaluation process, but, little agreement is available to suggest practice.
3. There is little consensus about what design process results in successful teacher evaluation.
4. Only a few districts had teacher evaluation processes that appeared to represent a well-developed system in which external and internal environmental demands were balanced to facilitate the attainment of both improvement and accountability.
5. The process of teacher evaluation design is an under conceptualized and underdeveloped activity (Wise et al., 1984).

The practical application considerations for two-year colleges are virtually the same as those for public schools. The search for literature relative to evaluation considerations for two-year colleges has unearthed only a very small amount of information. Some descriptions of various evaluation systems do, however, shed some light on what the essential elements of two-year college evaluation might be. In describing the full-time faculty evaluation system at North County Community College, Poole and Dellow (1983) use the following indices as measures of instructor effectiveness:

1. Motivating students toward superior achievement within his or her courses;
2. Generating an enthusiasm in and establishing rapport with students;
3. Presenting material in an orderly and preplanned method compatible with the stated objectives of the course. The level and intensity of the instruction should be compatible with course and curriculum objectives;
4. Making maximum use of library resources, audiovisual aids, laboratory equipment, and so on;
5. Using a variety of teaching techniques to achieve the desired objectives;
6. Evaluating student performance adequately and equitably within the framework of the defined grading policy of the college;
7. Keeping course materials, including textbook selection and reference reading lists, up to date;
8. Providing sufficient time to assist students on an individual basis and encouraging students to take advantage of such assistance;
9. Providing instruction in such a way that it is effective to the greatest possible number of people (Poole and Dellow, 1983)

Carl Sandburg College in Illinois uses a four-part, non-tenured faculty evaluation system. It doesn't spell out the weight to be given to the four main input factors of: (1) peer evaluation, (2) student evaluation, (3) faculty professional report, and (4) supervisory evaluation. Their flow chart would indicate that each of the four factors carry an equal weight (Carl Sandburg College, p. 9, in Andrews 1985).

The College of DuPage has not had a collective bargaining process in developing its procedures for evaluation. In its procedures on tenured faculty development, a faculty member's written statement (self-evaluation) and a student rating questionnaire provide the major sources of information to be used in the formal evaluation conference. DuPage's non-tenured evaluation procedures require both student and self-evaluation but make classroom visitations optional (Andrews, 1985).

The University of Florida sponsored a three-year National Faculty Evaluation Project for Community and Junior Colleges to assist eight community colleges to develop stronger faculty evaluation systems (Smith, 1983). The following is a summary of some of the elements which have been included in several of the evaluation systems that have been set up as a result of the University of Florida project (pp. 9-15):

1. Arapahoe Community College (Littleton, CO)
  - a. Annual self-evaluation
  - b. Annual student evaluations of teaching faculty
  - c. Peer evaluation every third year

2. Gateway Technical Institute (Racine, WI)

The instructors formally evaluated in this system will have three evaluations (Minimum):

- a. One by the instructor's coordinator
- b. Two others to be selected from student evaluations, self-evaluation, or a supervisor's evaluation.

3. Patrick Henry Community College (Martinsville, VA)

The following is a numerical/percentage breakout on Patrick Henry Community College's evaluation component of "teaching performance" which carries a weight of 70% in the overall evaluation process:

---

Min	Component	Student	Self	Supervisor
70%	1. Teaching Performance, or	65%	35%	0%
	2. Teaching, including classroom observation.	40%	30%	30%

---

This system works from the premise that all personnel holding faculty rank are presumed to possess the necessary professional qualifications to perform satisfactory in their positions, and that a "satisfactory" is expected of all faculty. It is interesting to note the large percentage weighted toward student input.

Smith (1983) points out that the faculty members under this system have the option of including classroom observation data from the supervisor in the evaluation plan but "most of the colleges in the University of Florida project ... have placed less emphasis on classroom observations as a form of faculty evalua-



tion." He suggests that the trend away from classroom visitations may be due to the research by Centra (1979) which supposedly was found to be highly unreliable.

The "Agreement" between the Los Angeles Community College District and its local AFT calls for regular employees (those who have completed probationary status) to be evaluated at least once every two academic years. Their procedures include: (1) peer evaluation, which uses student evaluations as part of its review process; and (2) administrative evaluation. Student evaluations are given more weight than any other part of their evaluation system which includes administrative in-class evaluation. In a system that gives more weight to unsophisticated student responses than it gives to its professional evaluators will, undoubtedly, fire very few incompetent faculty members.

Cohen and Brawer (1982) pointed out that, while some contracts mandated evaluation procedures for all faculty in a given institution (or state system), "faculty bargaining units leaned considerably more in the direction of protecting their members than toward enhancing professional performance."

In its summary on tenure and post-tenure evaluation, the National Commission on Higher Education Issues stated:

The Commission strongly affirms the continuing importance of faculty tenure as an essential instrument to protect academic freedom, and thereby ensure the highest quality in teaching and research.

But the time has come for campus administrations and faculties to review and, if needed, revise their procedures in order to assure themselves and the public that the procedures will produce fair, rigorous, and relevant evaluations (Fleming, 1982).

In discussing faculty development, Bevan (1980) says, "It is recognized that faculty evaluation can be an important aspect of faculty development." The reason he believes so is because, "such evaluation can reinforce personal growth and instructional improvement throughout a faculty member's career. "In one sense," he goes on to say, "the primary goals of faculty evaluation are identical to those of faculty development, i.e., the improvement of college teaching and the improvement of student learning" (p. 3).

The improvement of instruction in education has to be of prime importance. Governing Boards are in a position to make sure that it reaches such a level of high importance! Evaluation for persons who are to be considered for continued employment, tenure, salary advancements, promotion opportunities and merit recognition should begin very early. Such evaluation will assist the person to know if his or her level of preparation, organization, student expectations, etc., are at the level the individual college or secondary school expects of its faculty or academic support personnel. It should lead to both oral and written evaluation reports for improvement, encouragement, and recognition of good efforts.

Reaction to evaluation will help provide the true test of how a person wants to fit into the institution. Positive reaction to suggestions and criticism followed by some action to remediate any defects or concerns is a good indication that the individual is concerned, flexible, and capable of improving his or her performance. Negative and defensive reactions, followed by a "business-as-before" attitude also warns an administrator about some important personality traits with which one should be concerned.

Is there evidence that performance evaluation is not working in education at any level, that millions of dollars and precious human resources are not being used effectively? The review of literature for this study of over 100 articles and books on performance evaluation in education shows tremendous potential but a host of problems. There appears to be more positive elements on the public school's side of the ledger than there are on the two-year, postsecondary side. Public schools have not had to wrestle with the issue of academic freedom as often as their counterparts, nor have they had to be as responsive to classroom enrollments or the lack thereof. Further, most two-year, postsecondary school's have been built around an educational model that has incorporated a number of business and industry practices and philosophies, many of which were quality centered and product oriented and therefore tended to look more at the student outcome than the teachers performance, i.e., did the student get a job or did (s)he get accepted at a four year college without losing college credits.

#### McGreal's Models

In 1983 the Association for Supervision and Curriculum Development published a landmark book by Thomas L. McGreal entitled Successful Teacher Evaluation. McGreal identified a taxonomy of evaluation models by gleaning the literature and researching the common practices of teacher evaluators. He was able to demonstrate that five distinct models existed. He further demonstrated that virtually every evaluation process could be classified as one of the five models. The models are the Common Law Model, Goal-setting Model, Product Model, Clinical Supervision Model, and Artistic Model. Since this study seeks to classify the evaluation processes currently being utilized by two-year col-

leges in the United States of America according to these categories established by McGreal, each of the five models are discussed here-in-after.

#### Common Law Model

The model is characterized by high supervisor-low teacher involvement. Evaluation is seen as synonymous with observation. This model has similar procedures for tenured and non-tenured faculty members. The major emphasis is on summative evaluation. Formative evaluation is either non-existent or coincidental. Generally the Common Law Model has required instruments that force comparative judgments to be made between and among people. Consistently, this model is further characterized by the existence of standardized criteria, most of which are not research based but have evolved with the instrument over time (McGreal, 1983, pp. 10, 11). McGreal states: "the label 'common law' is used to describe certain evaluation systems since most districts who employ this form of evaluation have done so for so long that they have finally married it by formalizing the procedures" (p. 9).

#### Goal-setting Model

The model is characterized by an emphasis on an individualized approach to evaluation. Instructors and evaluators meet and confer to set and monitor goals. Generally, no checklist of criteria is used. Self-evaluation may be a component of this model.

The basic assumption underlying the goal-setting model is that when one formulates or participates in the formulation of performance goals that they will

have a clearer understanding of what is expected of them. This model is paper intensive and nearly impossible to use for ranking instructors (pp. 14-18).

### Product Model

The model is characterized by evaluation that is based upon the results or outcomes of student achievement tests or on competency - based evaluations, but not on methods, styles, or processes. Generally, the instruments for assessing student growth are norm - referenced tests and criterion referenced tests. This is a very controversial model because of the use of student performance as a method for assessing instructor performance (pp. 18-19).

Using student achievement as a measure of instructor competence rests on the assumption that enhancing student learning is an important function of teaching (Millman, 1981, p. 146). Certainly the methods used for measuring student learning gains are important to this model as are issues such as the use of tests and the influences of the classroom. There is, however, growing support for the use of student performance data for input in the formative evaluation of instructors. McGreal further suggests that those committed to improving instruction should include the use of student performance data as a component of summative evaluation.

### Clinical Supervision Model

The model is characterized by a close relationship between the instructor and the supervisor with emphasis on collegial rather than authoritarian orientation. It takes its principal data from the classroom and is designed to improve the instructor's performance. Since the focus includes formative as well as

summative evaluation the expectations of this model include greater teacher motivation and improvement (pp. 25-29).

There is general agreement (Goldhammer, 1969; Cogen, 1973; Acheson and Gall, 1980) that the sequence of clinical supervision contains five stages:

1. Pre-Observation conference
2. Observation of teaching
3. Analysis and strategy
4. Post-observation conference
5. Post-conference analysis

Manatt (1988) has enhanced this clinical supervision model by expanding the cycle to include more opportunities for feedback in the formative phase and by providing for various supporting and supplemental data to be included in an evaluation portfolio.

#### Artistic or Naturalistic Model

This model is characterized by a belief that teaching is an art; that the quality of the performance the instructor exhibits is likened to an aesthetic experience. The evaluation is more subjective and, perhaps, less precise. This approach for evaluation relies on the supervisor's appreciation for the artistic and natural abilities of the instructor's teaching performance. In such an approach to evaluation, the individual supervisor is the instrument through which what has gone on in the classroom is perceived. The major aim is to improve the quality of educational life in school.

This model assumes that the performance quality of the teaching act is directly proportional to the level of student learning. That is, the more

aesthetically pleasing the instructors performance, the better job the instructor is doing and the more the students are learning.

### The Personnel Evaluation Standards

Another landmark publication, The Personnel Evaluation Standards, was released in 1988 by the Joint Committee on Standards for Educational Evaluation, Daniel L. Stufflebeam Chair. This book is a guide for assessing or developing systems for evaluating education personnel. It presents and elaborates twenty-one standards by which to plan and assess systems for evaluating teachers, professors, administrators, counselors, and other educators. The twenty-one standards are divided into four general categories that correspond to four basic attributes of sound evaluation: propriety, utility, feasibility, and accuracy. Basically, the standards require that evaluations be proper, useful, feasible, and accurate.

The standards focus on systems used to evaluate the performance and qualifications of individuals. Moreover, they are intended to apply to a broad range of existing techniques (including observation, interview, applied performance tests, licensure tests, professional skills tests, development of a portfolio, supervisor assessment, peer assessment, and student assessment) and to new evaluation approaches as they are developed. The standards were developed by a Joint Committee with representatives from fourteen major professional associations concerned with education (Joint Committee, 1988). There was, however, no representative of the two-year colleges. Therefore, since the two-year colleges were left out this study will be useful in establishing the validity of the standards for two-year colleges.

The need for sound evaluation of education personnel is clear. Educational institutions must use evaluation to select, retain, and develop qualified personnel and to manage and facilitate their work. Clearly, the need for personnel evaluations in educational institutions is pervasive, important, and multifaceted.

According to Chairman Stufflebeam the function of the standards is to correct deficiencies in current practice and present educators and policy board members with a widely shared view of general principles for developing and assessing sound, acceptable personnel evaluation procedures and with practical advice for implementing them.

In its original report the Committee defined personnel evaluation as, "the systematic assessment of a person's performance and/or qualifications in relation to a professional role and some specified and defensible institutional purpose (Joint Committee, 1981, p. 12)." In this 1988 report they reaffirm the same definition (Joint Committee, 1988, pp. 8-9).

Five major assumptions guided the Committee's work. First, the fundamental purpose of personnel evaluation or any other education activity must be to provide effective services to students and society. Second, personnel evaluation practices should be constructive and free of unnecessarily threatening or demoralizing characteristics. Third, personnel evaluations are vital for planning sound professional development experiences. Fourth, disagreements about what constitutes good teaching, good administration, and good research may complicate personnel evaluation. Fifth, personnel evaluations vary in complexity and importance; consequently, applications of the standards may be crucial in some circumstances but out of place or even counterproductive in others (pp. 8-9).



The Committee also provides two points of clarification. First, the standards present criteria for judging evaluation plans, procedures, and reports. Second, the standards were developed for use in the United States.

The Propriety Standards are aimed at protecting the rights of persons affected by an evaluation, including student, instructors, counselors, administrators, and evaluators. Overall, the Propriety Standards require that evaluations be conducted legally, ethically, and with due regard for the welfare of students, other clients, and educators.

The five Utility Standards are intended to guide evaluations so that they will be informative, timely, and influential. The Utility Standards also require that evaluations be focused on predetermined uses, such as informing selection and tenure decisions or providing direction for staff development, and that they be conducted by persons with appropriate expertise and credibility. In general, these standards view personnel evaluation as an integral part of an institution's ongoing effort to recruit outstanding staff members, and, through timely and relevant evaluative feedback, to encourage and guide them to deliver high quality service.

The three Feasibility Standards promote evaluations that are efficient, easy to use, viable in the face of social, political, and governmental forces and constraints, and that will be adequately funded.

The eight Accuracy Standards aim at determining whether an evaluation has produced sound information about an educator's qualifications or performance. The overall rating of a personnel evaluation against the Accuracy Standards gives a good assessment of the evaluation's validity (pp. 10-13).

These standards were intended primarily to help four main groups to implement their evaluation responsibilities. This study was conducted as a member of the fourth audience listed by the Joint Committee. They state, "The fourth audience for the standards includes those who use The Personnel Evaluation Standards for research, development, or teaching. They can use the standards:

1. As a textbook for courses on personnel evaluation or personnel administration.
2. As criteria against which to evaluate alternative models for personnel evaluation.
3. As a framework for use in developing evaluation systems and better models for personnel evaluation.
4. As a logical structure for deriving and investigating questions and hypotheses about personnel evaluation (p. 15)."

Components two and four apply to this research.

Each standard is presented in a specific manner. Included are: a descriptor, the standard, an explanation, the rationale, some guidelines, common errors, selected illustrative cases, and supporting documentation. The Personnel Evaluation Standards is a well written and researched volume that should prove valuable to education evaluators for years to come. A further discussion of The Personnel Evaluation Standards and a complete summary of each of the twenty-one standards can be located in Appendix A, page 217.

### Approaches to Instructor Performance Appraisal

The public has come to believe that the key to educational improvement lies in upgrading the quality of instructors rather than other educational reforms

such as, emphasizing the basics, improving educational management, lowering class size, or updating the curriculum (Gallup, 1979; Gudridge, 1980; Vlaanderen, 1980). In response to these perceptions most states have legislated requirements for instructor evaluation (Beckham, 1981). Other policies have been put into place that pressure the educational institutions to utilize summative evaluation to make decisions about instructor dismissal, promotion, layoff, and even selection (Knapp, 1982; Peterson and Kauchak, 1982; Feldvebel, 1980).

The result of the various laws and policies appears to be tension between and among the various groups involved with evaluation. A parallel can be drawn from Knapp's (1982) articulation of stakeholders' perspectives in the K-12 school setting. Instructors have a stake in maintaining academic freedom, keeping their jobs, their sense of efficacy, and their collective, professional self-respect. Hence, instructors want evaluation systems that do not infringe on their rights, both academic freedom and the property rights of tenure. Further, instructors want evaluation systems that encourage professional growth without prescribing specific teaching models or methods, while at the same time showing appreciation for the complexity of the teaching act. Deans, division chairs, and department heads have a stake in maintaining stability in their organizations, allowing them to be responsive to student needs and to the bureaucratic concerns of the college while keeping the learning climate positive and staff morale intact.

#### Purposes of Instructor Evaluation Systems

The choice of an instructor evaluation system or process is associated with views of teaching work and of the college as an organization. However, many times these associations are made only implicitly in evaluation decision making.

The use to which evaluation results are put is a more definite factor in the choice as to which system, process, or model is employed. Instructor evaluation has four basic purposes which are divided equally between the areas of improvement (formative information) and accountability (summative information).

Formative evaluation can be utilized for individual professional growth, and school improvement. Summative evaluation can be employed in making individual personnel (job status) decisions and school status (for example certification) decisions. Many instructor evaluation systems are nominally intended to accomplish all four of these purposes, but different processes and methods are better suited to one or another of these objectives. In general, instructor evaluation systems most suited to improvement must provide a wide variety of descriptive data that highlights difficulties, deficiencies, and their sources as well as change possibilities that will lead to the desired improvement. Evaluation processes useful for accountability objectives must provide data that are objective, standardized, comparative, and adjudicative (Bolton, 1973; Costa, Garmston, and Lambert, 1988; Darling-Hammond, Wise, and Pease, 1983; Denham, 1987; Floden and Feiman, 1981; Harris, 1986; Joint Committee, 1988; Manatt, 1988; McGreal; 1983 and 1988; Popham, 1988; Redfern 1980).

#### Elements of Instructor Evaluation

There have been several recent reviews of instructor evaluation systems in which the authors identified a number of elements of instructor evaluation (Abrami, 1985; Arreola, 1986; Braskamp, Brandenburg, and Ory, 1984; Centra, 1977 and 1979; Darling-Hammond, Wise, and Pease, 1983; Doyle, 1983; Duke and Stiggins, 1986; Educational Research Service, 1979; Ellett, Capie and Johnson, 1980;

Haefele, 1980; Hawley, 1982; Lewis, 1982; Millman, 1981; Peterson and Kauchak, 1982; Seldin, 1980 and 1984). They reveal that the elements of instructor evaluation used seek to measure very different aspects of instructing and the instructor. Some seek to assess the quality of the instructor (instructor competence); others seek to assess the quality of instruction (instructor performance). Other elements claim to assess the instructor or their teaching by reference to student outcomes (instructor effectiveness). Following is a brief description of the most common elements of instructor evaluation used for evaluating two-year college personnel.

#### Self-evaluation

Self-evaluation is relatively common as a source of assessment data in instructor evaluation. The instructor can access information for self-evaluation from a number of easily developed data bases -- student or peer ratings, self assessment measures of student achievement, and so forth, to make judgements about their own teaching. In the rush to demonstrate accountability, the education community has either overlooked or cast aside this valuable evaluation tool. Self-ratings can be quite enlightening for both administrators and instructors. For example, during the various post-conferences (be they formative or summative) the administrators ratings can be compared with the instructors' and the differences between them discussed. This approach can be very useful in building open communication and trust. Administrators must be prepared to openly discuss the reasons why the ratings were given to the instructors. If there is a wide discrepancy in the ratings on some area of performance, this difference must be reconciled. This will require honesty, objectivity, and trust from both

parties. This technique also helps the administrator identify potential rater biases, instrument deficiencies, criteria and/or standards misunderstandings, and monitoring and reporting inadequacies. Although not useful for accountability purposes, the element of self-evaluation can be utilized in cooperation with other elements to determine individual professional development needs. In particular, both Manatt (1988) and Redfern (1980) consider self-evaluation an essential element of their evaluation systems.

In studies comparing supervisory and self-ratings, self-ratings have been found to be both more lenient (Kirchner, 1965; Parker et al., 1959) and harsher (Heneman, 1974) than the supervisor's rating. Thornton (1968) reviewed the literature on self-appraisal and concluded that self-ratings generally tend to show more leniency, less discriminate validity, less reliability, less agreement with other sources, and less help than others' ratings (supervisor, peers, and subordinates), although more recent reviews have found less definitive trends (Landy and Farr, 1980). Thus the question of whose ratings are more accurate is difficult to answer. It is possible, as Dunnette and Borman (1979) have pointed out, that ratings made from different perspectives are equally valid. This viewpoint has been supported by other work (Landy and Farr, 1980).

There are seven potential advantages of using self-ratings in the appraisal process.

1. It allows for the clarification of standards, criteria, goals and objectives.
2. It helps identify extremely biased or different ratings.
3. It can be continuous assessment.

4. It generates a larger data base on which personnel decisions can be made.
5. It is likely that instructors will act on data they collect on themselves.
6. It permits assessment of the reliability of ratings as well as the sensitivity of the rating format.
7. It allows participation by the teacher in the evaluation process that can foster commitment to the system.

### Peer Review

Another common element in instructor evaluation is peer review -- sometimes called an academic council. A number of authors have addressed the role of colleagues in evaluating instruction (Braskamp, Brandenburg, and Ory, 1984; Centra, 1975 and 1979; Cohen and McKeachie, 1980; Darling-Hammond, Wise, and Pease, 1983; French-Lazovik, 1981; Haefele, 1980; Lee, 1985; Peterson and Kauchak, 1982; Sechrest and Hoffman, 1982; Thomas, 1979; Weinbach and Randolph, 1984; Wilson, Dienst and Watson, 1973). This element covers a broad range of performance, encompassing not only instruction in the classroom, laboratory, or shop, but also what the instructor intends to have happen and other instructional behaviors as exhibited by grading, testing and assignment practices. This element is valuable for the purpose of improvement since the evaluators (normally a committee) will observe in the classroom and examine lesson plans, tests, graded assignments and examinations and provide the instructor with specific, practical suggestions for improvement. Peer evaluation requires a high degree of professional ethics and objectivity, as well as good observational and analysis skills. Because this element is not subject to direct administrative control and is more open to a variety of unregulated stan-

dards for assessing performance, it is not generally acceptable for fulfilling the purposes of accountability.

### Student Feedback

This element has long been employed in various instructor evaluation schemes. In effect it is viewed as another form of classroom observation, one that is from the student's point of view rather than the administrator's. Student ratings can be used for either formative or summative evaluation data. This element is inexpensive with a high degree of reliability (Judkins, 1987; Peterson and Kauchak, 1982). Several authors have discussed the student's role in evaluating instruction, many have indicated the student ratings should be considered an essential element of a quality evaluation system (Abrami, Leventhal and Dickens, 1981; Abrami, Leventhal and Perry, 1982; Aleamoni, 1981; Aubrecht, 1981; Berk, 1979; Braskamp, Brandenburg, and Ory, 1984; P.A. Cohen, 1980, 1981, and 1986; Cohen and Mays, 1981; Darling-Hammond, Wise, and Pease, 1983; Doyle, 1983; Feldman, 1976a, 1976b, 1977, 1978, 1979, 1983, and 1984; Gilmore, 1984; Gilmore, Kane and Naccarato, 1978; Haefele, 1980, Judkins, 1987; Marsh, 1984; McKeachie, 1979; McNeil and Popham, 1973; Menges and Brinko, 1986; Murray, 1980; Peterson and Kauchak, 1982). Several strengths of this element can be gleaned from the writings of those listed above. For example: (1) student feedback requires a minimum of resources, (2) instructors will change as a result of student feedback, (3) results correlate highly with other evaluation elements, (4) students observe the instructor on a daily basis, (5) student feedback is reliable and not affected by grades, and (6) several good instruments currently exist. This element can be, and has been, used for all four purposes of evaluation.



### Student Achievement

"Did the kids learn?" is perhaps the most important question we ask ourselves in education. Much has been written about using measures of student achievement on outcomes of significance as an element of instructional evaluation. This element has evaluative value because student attainment of objectives is a legitimate source of data on teacher performance and because tests of student performance measure the impact of teachers on students over a period of time. However, studies of the reliability of using test scores as a measure of instructional effectiveness consistently indicate that reliability is quite low, that there is some difficulty in designing appropriate tests, that gains on standardized tests often are inadequate to measure performance, and that no consideration is given for other important variables such as student intelligence, family background, test anxiety or school climate (Braskamp, Brandenburg, and Ory, 1984; Brophy, 1973; Centra and Potter, 1980; Darling-Hammond, Wise, and Pease, 1983; Doyle, 1983; Rosenshine, 1970; Seldin, 1980; Shavelson and Russo, 1977; Southern Regional Education Board, 1985; Veldman and Brophy, 1974). As a result, the primary purposes for using student achievement as an evaluation element appears to be to improve student learning and to identify instructors for merit recognition.

### Feedback Conference

This element is often times known as instructor interviews or conferencing and has been a prevalent evaluation element in the past and is a cornerstone of some of the more recent evaluation systems (Darling-Hammond, Wise, and

Pease, 1983). Haefele (1981, p. 49) first defined one of the purposes of teacher interviews (conferencing) to be communicating performance appraisals to practicing teachers.

Historically, the classroom observation and the subsequent post-observation conference has been the sum and substance of the evaluation process. However, recently conferencing has been viewed as an important skill and the conferences as essential elements of an effective evaluation system (Darling-Hammond, Wise, and Pease, 1983). Administrator evaluations of instruction have been the source of a great deal of recent literature, in particular, a preobservation conference has been recognized as useful in the involvement of teachers in their own professional development and for improved feedback from administrators (Ashbaugh and Kasten, 1987; Blackburn and Clark, 1975; Braskamp, Brandenburg, and Ory, 1984; Braunstien and Bentson, 1973; Centra, 1979; Doyle, 1983; Garawski, 1980; Gudridge, 1980; Manatt, 1988; McNeil and Popham, 1973; Redfern, 1980; Seldin, 1980).

### Supervisor Observation

This element is the cornerstone of most instructor evaluation systems. It involves administrators evaluating an instructor's performance through classroom observation, and a review of lesson design, student learning data, and feedback from students and peers. Generally, this element involves the use of a standardized instrument to describe an instructor's performance. Two of the most widely discussed evaluation systems are Manatt's (Manatt, Palmer and Hidlebaugh, 1976) 'Mutual Benefit Evaluation' more recently redefined as 'A Total Systems Approach to Teacher Performance Evaluation' (Manatt, 1988) and Red-

fern's (1980) 'Management by Objectives Evaluation'. Both Manatt (1988) and Redfren (1980) consider supervisor observation an essential element of their evaluation systems.

This element of evaluation has among its strengths the fact that the supervisor is familiar with the school goals, culture and climate; that (s)he often has additional information about instructor performance; and that (s)he can compare instructors within a school. Additionally, the element of supervisor observation requires a minimum of resources for observation, feedback, and follow-up. However, even advocates of supervisor observation recognize its shortcomings. The lack of inter-rater reliability, rater bias, limited occurrences of observation, and poorly constructed measurement instruments can put the reliability and validity of the results at risk (Andrews and Knight, 1987; Braskamp, Brandenburg, and Ory, 1984; Centra, 1979; Doyle, 1983; Evertson and Holley, 1981; Haefele, 1980; Lewis, 1982; Manatt, 1988; Medley, 1982; Peterson and Kauchak, 1982).

### Competency Tests

Competency tests will be divided into three general areas for this study. Those areas will be written examinations, simulated teaching, and assessment center exercises.

According to Shulman (1988), written examinations are regularly used to measure basic skills, knowledge of teaching content, and understanding of professional practice. He further asserts that they permit broad sampling of domains, are economical to use, and highly reliable. However, there are some draw backs associated with written examinations. Even though they excel at

measuring select pieces of knowledge, they fail to tap more integrated processes of judgement, decision making, and problem solving in more realistic contexts.

Occasionally administrators attempt to test for instructional competency by a means commonly called simulated teaching. The scenario for this element is generally that the instructor teaches a brief teaching unit on content not normally taught by the teacher to a special selected group of students. The students are given pre- and post-tests to measure their gains. The advantages of this element have been reported to include the facts that this method evaluates instructors in terms of student learning and that short-term feedback increases control over nonteacher variables assumed to influence students (Glass, 1974). However, some disadvantages are evident. First, this element does not allow for assessing student progress when exposed over time to the same instructor. Second, it is conducted in what has to be considered not a normal classroom situation. And third, the element would be quite expensive to conduct.

Another element that is used to determine the instructional competency of teachers is known as an assessment center. Assessment center exercises for instructors simulate real teaching episodes with real problems. The process of teaching is usually emphasized by instructors actually preparing lesson plans, laying out the lesson, teaching the lesson and evaluating students. Sometimes instructors are asked to critique video tapes of themselves and others teaching.

### Teacher Artifacts/Portfolios

This element of instructor evaluation appears to be growing in popularity. Shulman (1988) envisions a portfolio process that includes written tests of knowledge, systematic documentation of accomplishments, formal attestations

by colleagues and supervisors, and analysis of performance in assessment centers and the workplace. According to McGreal (1988) teaching artifacts include all instructional materials used to facilitate learning. This includes everything from textbooks, workbooks, and supplementary texts to learning kits, maps, audiovisual aids, films, dittoed material, study guides, question sheets, worksheets, problem sets, quizzes and tests. Manatt (1988) asserts that the practice of teacher evaluation should include the evaluation of the products of instruction as well as the process and the person. The teacher performance evaluation cycle advanced by Manatt has a specific step in which the evaluator adds "other" data before a summative evaluation report is written. These "other" data are a collection of teacher artifacts and relevant information.

#### Outside Evaluators

Both visiting teams of experts and individual outside evaluation specialists have been used to evaluate instructors. Generally these people observe instructor performance and/or review student learning data. Some of the advantages to this approach result from the fact that evaluators with special skills can be utilized. Also these people are external to the politics, problems and biases of the college. The primary disadvantages are the cost to bring visitors in for sufficient time and the college may tend to ignore data collected this way.

#### Indirect Measures

Indirect measures such as experience and educational background are traditionally linked to teacher salary and promotion opportunities. Other indirect measures such as attitudes, values, personality traits, professional commitment,

or outside interests have not been shown to discriminate between effective and ineffective instructors (Gage, 1963; King, 1981; Schalock, 1979). Despite this finding, it is arguable that indirect measures ought to be a supplementary data source for instructor evaluation. In fact, recent research in the area of staff development indicates that the more people engage in dialogues about teaching and learning the better they get at it (Griffin and Barnes, 1986).

#### Summary of the Review of Literature

The practice of evaluating instructors has been the target of extensive research. This review of literature relevant to instructor performance appraisal was undertaken to consider the value of, need for, and purpose of faculty evaluation. The review focused on identifying commonalities among evaluation systems. There does appear to be some central elements of the evaluation process. Additionally, this review of literature briefly investigated the historical, legal, ethical, and applicational considerations of instructor performance appraisal.

Most of the research on instructor performance appraisal is focused on instructor groups other than two-year college instructors. However, the research clearly points to the fact that instructor evaluation, regardless of instructional level, is a useful tool for instructional improvement. The common thread that runs through the research is that evaluation is necessary. Costa et al. (1988) best summarized the areas of agreement between the leading experts on instructor evaluation.

**Evaluation is a rigorous process, and evaluators must be skilled and trained in executing it. All affected parties must be involved in the**

process. The criteria for judgement must be defined, communicated, and understood by everyone. Supervision must be ongoing. Staff development is a necessary component. There is no substitute for strong instructional leadership and, when handled poorly, evaluation causes suffering for all involved – the teacher, the students, the administrator, the school board, and the superintendent (Costa et al., 1988).

To summarize then: two-year colleges have the need to perform instructor evaluation and the need to identify the best methods to do so. There appears to be no clear cut prescription to instructor evaluation, thus the need to study the current practices in two-year colleges and the efficacy of the established personnel evaluation standards for those two-year colleges.

## CHAPTER III. METHODOLOGY

### Introduction

This study sought to determine the current practices of instructor performance appraisal prevalent in two-year colleges across the nation. Two questionnaires were developed and administered.

Once the literature was reviewed, the problem identified, and a proposal approved the following procedures were employed for this investigation: (1) identification of population, (2) construction of questionnaires, (3) preparation of instructions and information to accompany questionnaires, (4) selection of a sample to respond to the initial questionnaire, (5) administration of the initial questionnaire and collection of data, (6) selection of the sample to respond to the primary data gathering questionnaire, (7) administration of the primary data gathering questionnaire and collection of data, (8) analysis of data, (9) summary, findings, and conclusions, and (10) recommendations of the study.

The questionnaires, subjects who participated, data collection procedures, and statistical analyses are reviewed in this chapter.

### Population

The population used in this study consisted of all 2250 two-year colleges in the United States and its territories as identified in the Chronicle Two-Year College Databook For 1988-89 School Year.



### Sample Selection

A ten percent sample of the population was randomly selected from the Chronicle Two-Year College Databook For 1988-89 School Year using random numbers generated and assigned on a MacIntosh Plus™ using a software application known as Stat Works©. This sample was used for the initial questionnaire. The Chronicle Two-Year College Databook For 1988-89 School Year contained the mailing addresses of all of the two-year colleges sampled in this phase of the study. The information gathered from the initial questionnaire then became the population for the primary data gathering questionnaire. The primary data gathering questionnaire went to a 40 percent random sample of those identified as the instructor evaluators as gathered in phase one of this study. The resulting sample size was three hundred and fifty-seven. Of the 357 instructor evaluators surveyed 256 or 71.7 percent responded.

### Methodology and Procedures

The steps involved in conducting this study are here-in-after presented in chronological order so that other skilled researchers could easily duplicate this study. After choosing a topic that was of interest and discussing it with the professors in charge of the major work a population to study was identified. A number of data sources were reviewed and the Chronicle Two-Year College Databook For 1988-89 School Year was selected as the most complete and current listing of two-year colleges. Next the review of literature was undertaken in order to determine what was currently known about the topic and to provide the researcher with sufficient background to make a significant contribution of new information to the existing research base. A series of objectives for the study

were developed. These objectives lead to the formulation of research questions and hypotheses. Since there currently existed no instrumentation that could answer the research questions or provide data from which to test the hypotheses, instrumentation was developed. Additionally, instructions and information to accompany the questionnaires were formulated. It was determined that two questionnaires would be needed in order to access the evaluators that were primarily responsible for instructor performance appraisal. As a result, the study was conducted in two phases.

During the first phase of the study, a random sample of leaders of two-year colleges were asked to respond to the first questionnaire. This phase of the study was designed to identify the key participants responsible for evaluating instructor performance in two-year colleges. Appendix B (p. 234) is a copy of the questionnaire used in this phase of the study. The responses requested were the names and titles of the supervisors responsible for instructor performance appraisal. The initial questionnaire was mailed to a ten percent random sample of the 2,250 two-year colleges listed in the Chronicle Two-Year College Databook For 1988-89 School Year. There were 225 colleges from 45 states and two American territories. Of the 225 two-year college leaders surveyed 187 or 83.11 percent responded. The respondents identified the type and size of their institution as well as those persons responsible for instructor performance appraisal. These respondents represented the two American territories and 37 states. The respondents identified 892 key participants in evaluating instructor performance in two-year colleges.

The second phase of the study was designed to gather the data about the current practices in instructor evaluation in two-year colleges. A random sample

of those identified as instructor performance evaluators by those surveyed with the first questionnaire were asked to respond to a lengthy data gathering questionnaire that was designed to collect information that would provide answers to the problems identified in this study. The primary data gathering questionnaire was the instrument used in data collection and can be found in Appendix C (p. 236). The primary data gathering questionnaire was mailed to a 40 percent random sample of those identified as the instructor evaluators as gathered in phase one of this study. A 40 percent sample was selected in order to provide sufficient data to conduct meaningful statistical analyses. The primary data gathering questionnaire was mailed directly to the identified instructor supervisor/evaluator. Appendix D (p. 241) is a copy of the accompanying information and instruction letter. A mailing label with a return address was also enclosed in the mailing. No return envelop or postage was supplied to the participants due to the request for policies, evaluation instruments, samples, etc.

Thirty days after the first mailing in phase two a second, follow-up mailing was completed. Sixty days after the first mailing in phase two a telephone follow-up was made to five non-respondents. Three of these five people agreed to complete and return the questionnaire and cited lack of time as a reason for non-response. Within a week after being telephoned two of three had indeed returned a completed questionnaire. One of the five telephoned non-respondents cited lack of return postage and envelop as a reason for non-response and further declined to respond. The final non-respondent had changed jobs within his college and didn't feel inclined to complete the questionnaire. Seventy-five days after the first mailing in phase two of the study the decision was made to end the data gathering activities and to process the data.

Two hundred and fifty-six people returned the questionnaire, 184 of them included various examples of their evaluation instruments, policies, faculty handbooks, and/or narrative descriptions of their evaluation process and procedures. These respondents represented 68 percent of the two-year colleges surveyed, 37 of the 45 states surveyed and two American territories.

### Data Analysis

Once the data were gathered the decision was made to use the Iowa State University Mainframe computer system and a statistical data analysis software known as SPSSx©. A MacIntosh Plus™ computer, a telecommunications software program known as MacSampson©, and an Apple™ 1200 Baud modem were used as the data entry vehicles. The SPSSx job control language, a coding scheme that was created for the data, and the data were entered into MacSampson© and down loaded onto the ISU Mainframe. For ease of viewing and editing on the MacIntosh™ monitor the data list records were divided into two rows, but still treated as a single data record. Once the data were resident in the ISU Mainframe, a program to execute statistical data analysis procedures was written. The procedures created were those necessary to test the hypotheses identified in Chapter One. The results of these analyses are discussed in Chapter IV.

### Instrumentation

The questionnaires were developed through a review of the literature and through consultation with a review panel made up of Professors Ebbers, Manatt, Sweeney and Stow. The initial questionnaire was mailed on January 25, 1989. The primary data gathering questionnaire was refined and validated by distribut-

ing it to a group of practicing educational administrators for their review. Responses, suggestions, and comments were received from the persons mentioned above and were incorporated into a finalized questionnaire which was distributed to the sample of key participants in instructor performance appraisal.

The initial questionnaire (Appendix B, p. 234) consisted of a request for the respondent to identify key participants in instructor performance appraisal at his or her particular two-year college. This initial questionnaire was distributed to the chief executive officers of the two-year colleges identified as the sample.

The primary data gathering questionnaire (Appendix C, p. 236) was designed to gather demographic data about the two-year college and professional and personal data about the respondent. Additionally, it asked the respondent to classify their two-year college's instructor performance appraisal system into one of McGreal's models. The questionnaire provided a brief but concise, description of each model. Next, the questionnaire requested that the respondent evaluate the efficacy of the model currently in use in his or her two-year college on each of the twenty-one personnel evaluation standards identified by the joint committee on standards for educational evaluation (Joint Committee, 1988). The questionnaire accurately restated each standard so that the respondent could rate each.

### Summary

The data gathered for this research were gathered from four sources -- a review of the literature, a collection of instructor evaluation documents voluntarily submitted from the two-year colleges surveyed, the two questionnaires distributed to the selected samples, and when necessary for clarification, follow-up

phone calls to the respondents. The data were statistically analyzed, profiles developed, conclusions drawn, and recommendations made.

## CHAPTER IV: ANALYSIS OF DATA

## Introduction

In this chapter, each of the research questions and research hypotheses presented in Chapter I will be restated and the results of the statistical tests used will then be displayed in table and/or graphic form. A narrative profile of the "typical" instructor evaluator will be presented. Instruments used in data collection can be found in Appendices C and D.

## Phase One Questionnaire

The initial questionnaire went to a ten percent random sample of the 2,250 two-year colleges listed in the 1988-1989 Chronicle Two-year Colleges Databook. Of the 225 two-year college leaders surveyed, 187 or 83.1 percent responded. The respondents identified the type and size of their institution as well as those persons responsible for instructor performance appraisal. The type of institution is shown in Table 1.

TABLE 1. Type of two-year college as identified by the leaders of those institutions responding to the initial questionnaire

Type	Frequency	Percent
Community College	125	66.8
Junior College	23	12.3
Technical College	18	9.6
Vocational Technical College	13	6.9
Community/Technical College	6	3.2
Other	2	1.0

It's interesting to note that nearly two thirds of the respondents identified their colleges as "community colleges".

The reported size of the institution is depicted by Figure 1. The categories represent the reported full time equivalent (F.T.E.) students enrolled at the college. Most of the respondents were from two-year colleges with less than 1501 F.T.E. In the category of two-year colleges that were reported as less than 1501 F.T.E., one college reported less than 100 F.T.E. In the category of more than 3501 F.T.E., three colleges reported F.T.E. in excess of 10,000 students.

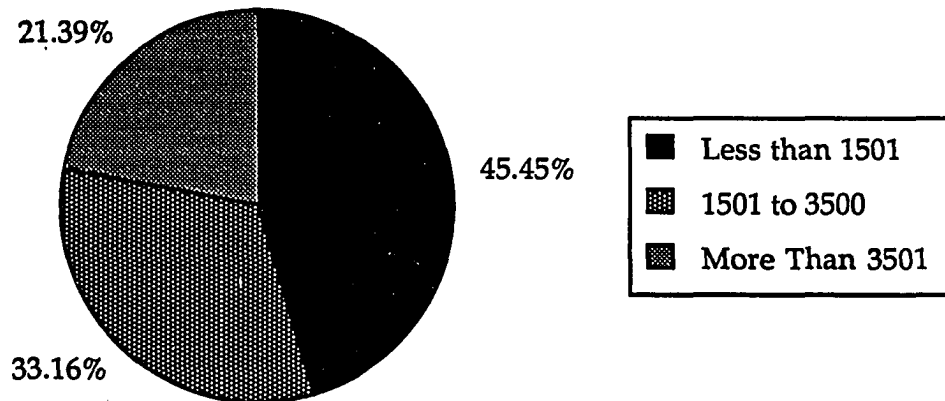


FIGURE 1. The size of the institution as reported by the leaders of those colleges responding to the initial questionnaire

The respondents identified 892 people as responsible for instructor performance appraisal. The dean of instruction or academic affairs was the job title most frequently identified as the position responsible for instructor performance



appraisal. Division chairs and department heads were also frequently identified as responsible for instructor performance appraisal. Traditionally, those persons in roles as division chairs and department heads have been members of the faculty with released time to conduct the business related to their respective roles. Table 2 shows the titles of those persons identified as being responsible for instructor performance appraisal.

TABLE 2: The titles of those persons identified as being responsible for instructor performance appraisal

Title	Frequency	Percent
President	10	1.1
Vice President	59	6.6
Dean/Director of Instruction	168	18.8
Dean of a Subject Area	84	9.4
Associate Dean	90	10.1
Division Chair	160	17.9
Department Head	139	15.6
Program Director/Manager	113	12.7
Coordinator/Supervisor	63	7.1
Missing	<u>6</u>	<u>0.6</u>
Total	892	99.9

#### Phase Two Questionnaire

The primary data gathering questionnaire went to a 40 percent random sample of those identified as the instructor evaluators as gathered in phase one

of this study. Of the 357 instructor evaluators surveyed, 256 or 71.7 percent responded. One hundred eighty-four of the respondents included various examples of their evaluation instruments, policies, faculty handbooks, and/or narrative descriptions of their evaluation process and procedures. These respondents represented 68 percent of the two-year colleges surveyed, 82.2 percent of the 45 states surveyed and two American territories. As in phase one of this study, the respondents identified the type and size of their institution. The respondents were given a list containing the types of two-year colleges to choose from and were asked to indicate the F.T.E. of their institution. The reported type of institution is shown in Table 3.

TABLE 3: Type of two-year college as identified by the participants in instructor evaluation responding to the primary data gathering questionnaire

Type	Frequency	Percent
Community College	170	66.4%
Junior College	33	12.9%
Technical College	25	9.8%
Vocational Technical College	18	7.0%
Community Technical College	9	3.5%
Other	<u>1</u>	<u>0.4%</u>
Total	256	100.0%

Figure 2 is a representation of the size of the two-year college as identified by the respondents to the primary data gathering questionnaire. No effort was

made to equalize the number of respondents representing the various sizes of the colleges.

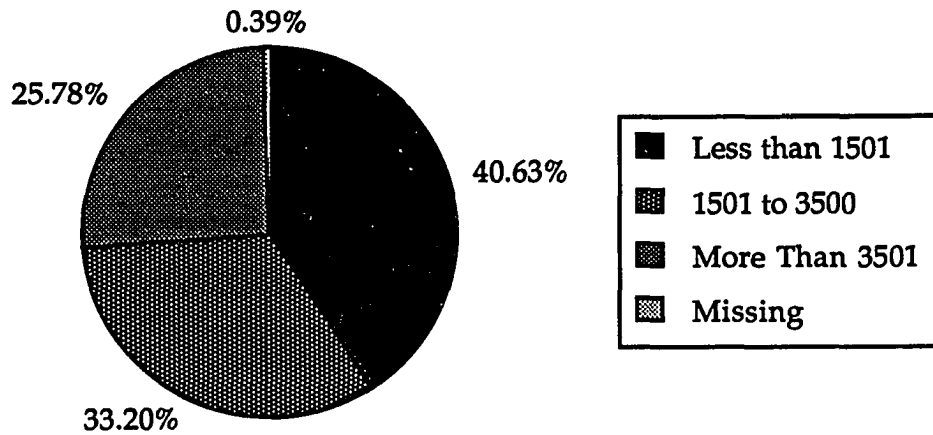


FIGURE 2: Size of the two-year college as identified by the respondents to the primary data gathering questionnaire

For simplicity in data reporting the research questions will be answered in chronological order.

#### The First Research Question

1. Who supervises whom and what? What is the span of control?

What are each supervisors duties?

- a. Specifically - Who are the instructional leaders?
- b. How many instructors do they supervise and/or evaluate?
- c. What other duties do these instructional leaders have responsibility for?

Of those responding there does not appear to be any one position or title that heavily dominates instructor evaluation at the two-year college level. Fig-

ure 3 is a listing, by title, of those responding to the primary data gathering questionnaire as instructor evaluators. The Dean of Instruction or Academic Affairs is most frequently identified as an instructor evaluator. The figure depicts the frequency of returns by job title.

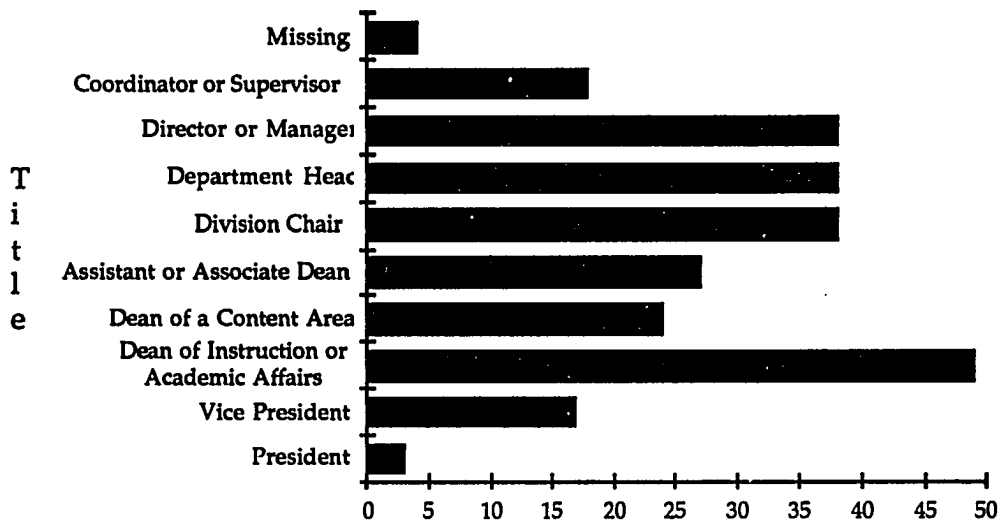


FIGURE 3. A listing, by title, of those responding to the primary data gathering questionnaire as instructor evaluators

However, when comparing the number of instructors evaluated to the respective positions it becomes apparent that the span of control varies by position. For example, those responding with the title of Vice-President evaluate over 133 instructors each while those identifying themselves as Department Heads averaged evaluating just under 20 instructors each. Table 4 is a comparison of the average number of instructors evaluated by title of the evaluator.

**TABLE 4: A comparison of the average number of instructors evaluated by title of the evaluator**

Title of Evaluator	No. of Respondents	Total No. of Instructors Evaluated	Ave. No. of Total Instructors Evaluated
President	03	61	20.3
Vice President	17	2274	133.8
Dean of Instruction	49	3062	62.5
Dean of Area	24	900	37.5
Ass't. or Assoc. Dean	27	1290	47.8
Division Chair	38	1276	33.6
Department Head	38	757	19.9
Director or Manager	38	956	25.2
Coordinator or Supervisor	18	569	31.6
Missing	<u>04</u>	<u>176</u>	<u>44.0</u>
Total or Overall Average	256	11,321	44.2

The overall average of 44.2 is greater than most of the averages for the various positions and is obviously heavily influenced by the unusually large average for the vice president. The figure that follows is a graphic representation of the data found in Table 4.

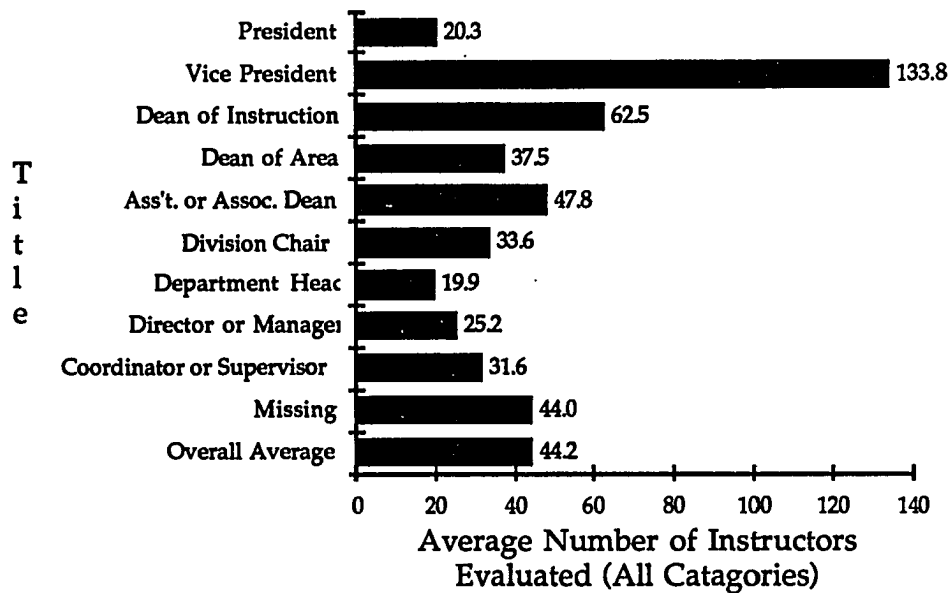


FIGURE 4: A comparison of the average number of instructors evaluated by title of the evaluator

Table 5 is intended to expand the comparison of the average number of instructors evaluated by the respondents to this survey. The data in this table are disaggregated by type of instructor being evaluated and represents the overall averages for each evaluator title. However, this table somewhat distorts the real span of control of two-year college evaluators because some evaluators don't evaluate instructors of each classification. Also, averaging doesn't allow for the variance between the different evaluator titles.

**TABLE 5: A comparison of the average number of instructors evaluated by title of the evaluator for each kind of instructor**

<b>Title of Evaluator</b>	<b>Ave. No. of Classroom Instructors Evaluated</b>	<b>Ave. No. of Clinical Instructors Evaluated</b>	<b>Ave. No. of Laboratory Instructors Evaluated</b>	<b>Ave. No. of Shop Instructors Evaluated</b>
President	20.3	0	0	0
Vice President	122.6	3.9	3.9	3.3
Dean of Instruction	53.8	2.5	3.6	2.7
Dean of Area	35.4	1.1	0.9	0.1
Ass't. or Assoc. Dean	37.0	1.4	5.3	4.0
Division Chair	30.7	0.5	0.9	1.4
Department Head	17.4	0.8	1.6	0.1
Director or Manager	22.2	1.3	0.9	0.7
Coordinator or Supervisor	24.5	1.4	3.9	1.8
Missing	<u>34.0</u>	<u>5.2</u>	<u>1.0</u>	<u>3.8</u>
Overall Average	38.6	1.6	2.4	1.7

A further comparison can be made by contrasting the overall average number of each kind of instructor being evaluated by all of the respondents with the same title, with the average number being evaluated by those evaluating one or more instructors of a particular kind.

Almost 76 per cent of the respondents indicated that they evaluate classroom instructors. The overall average number of classroom instructors being

evaluated per respondent is almost 10 points higher (48.5) if those with no responsibility to evaluate classroom instructors are not included.

**TABLE 6:** The average number of classroom instructors evaluated by all respondents compared with the average number evaluated by those evaluating at least one such instructor

Title of Evaluator	Ave. No. of Classroom Instructors Evaluated	Number of Respondents Actually Evaluating Classroom Instructors	Ave. No. of Classroom Instructors Actually Evaluated By Respondent
President	20.3	03	20.3
Vice President	122.6	13	160.4
Dean of Instruction	53.8	42	62.7
Dean of Area	35.4	22	38.6
Ass't. or Assoc. Dean	37.0	23	43.4
Division Chair	30.7	37	31.6
Department Head	17.4	38	30.7
Director or Manager	22.2	35	24.1
Coordinator/Supervisor	24.5	16	27.6
Missing	<u>34.0</u>	<u>03</u>	<u>45.3</u>
Overall Average	38.6	194	(75.8%) 48.5

The figure below is a graphic representation of the data found in Table 6. The overall average of 48.5 for those evaluating at least one classroom instructor is significantly greater than the overall average for all respondents. Thus the span of control is also greater than the overall average would initially indicate.



Also, it is interesting to note that the average for the department heads that have responsibility to evaluate classroom instructors is almost double the average for those who don't.

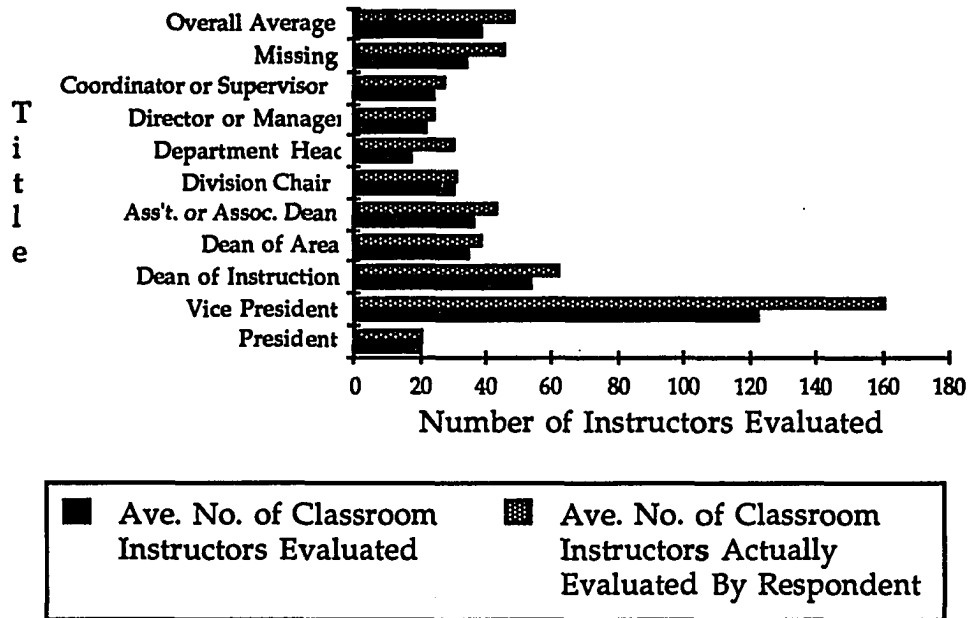


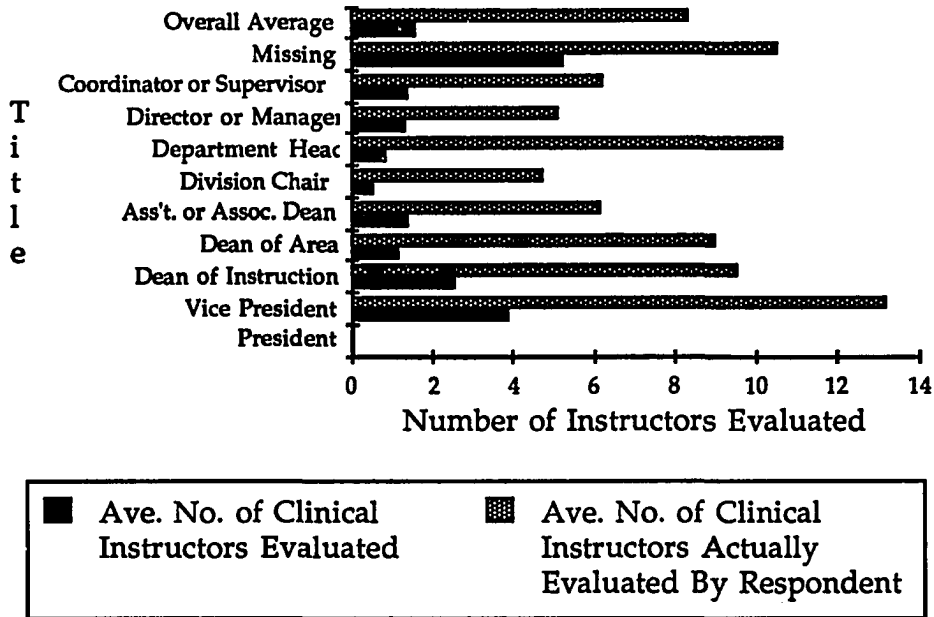
FIGURE 5: The average number of classroom instructors evaluated by all respondents compared with the number evaluated by those evaluating at least one such instructor

Almost 20 per cent of the respondents indicated that they evaluate clinical instructors. The overall average number of clinical instructors being evaluated per respondent is over 5 times higher when those with no responsibility to evaluate clinical instructors are not included. Table 7 contains a comparison of the average number of clinical instructors evaluated by the respondents to the survey.

TABLE 7: The average number of clinical instructors evaluated by all respondents compared with the average number evaluated by those evaluating at least one such instructor

Title of Evaluator	Ave. No. of Clinical Instructors Evaluated	Number of Respondents Actually Evaluating Clinical Instructors	Ave. No. of Clinical Instructors Actually Evaluated By Respondent
President	0	0	0
Vice President	3.9	5	13.2
Dean of Instruction	2.5	13	9.5
Dean of Area	1.1	3	9.0
Ass't. or Assoc. Dean	1.8	6	6.2
Division Chair	0.5	4	4.8
Department Head	0.8	3	10.7
Director or Manager	1.3	10	5.1
Coordinator/Supervisor	1.4	4	6.2
Missing	<u>5.2</u>	<u>2</u>	<u>10.5</u>
Overall Average	1.6	50 (19.5%)	8.4

Figure 6 is a graphic representation of the data found in Table 7. The overall average of 8.4 for those evaluating at least one clinical instructor is significantly greater than the overall average for all respondents.



**FIGURE 6:** The average number of clinical instructors evaluated by all respondents compared with the average number evaluated by those evaluating at least one such instructor

Nearly 28 per cent of the respondents indicated that they evaluate laboratory instructors. The overall average number of laboratory instructors being evaluated per respondent is over three times higher when those with no responsibility to evaluate laboratory instructors are not included. Table 8 contains a comparison of the average number of laboratory instructors evaluated by the survey respondents.

TABLE 8: The average number of laboratory instructors evaluated by all respondents with the average number evaluated by those evaluating at least one such instructor

Title of Evaluator	Ave. No. of Laboratory Instructors Evaluated	Number of Respondents Actually Evaluating Lab. Instructors	Ave. No. of Lab. Instructors Actually Evaluated By Respondent
President	0	0	0
Vice President	3.9	6	11.2
Dean of Instruction	3.6	18	9.7
Dean of Area	0.9	6	3.7
Ass't. or Assoc. Dean	5.3	10	14.8
Division Chair	0.9	8	4.5
Department Head	1.6	7	8.4
Director or Manager	0.9	9	4.0
Coordinator or Supervisor	3.9	6	11.7
Missing	<u>1.0</u>	<u>1</u>	<u>4.0</u>
Overall Average	2.4	71	(27.7%) 8.0

The figure that follows is a graphic representation of the data found in Table 8. The overall average of 8.0 for those evaluating at least one laboratory instructor is significantly greater than the overall average for all respondents.

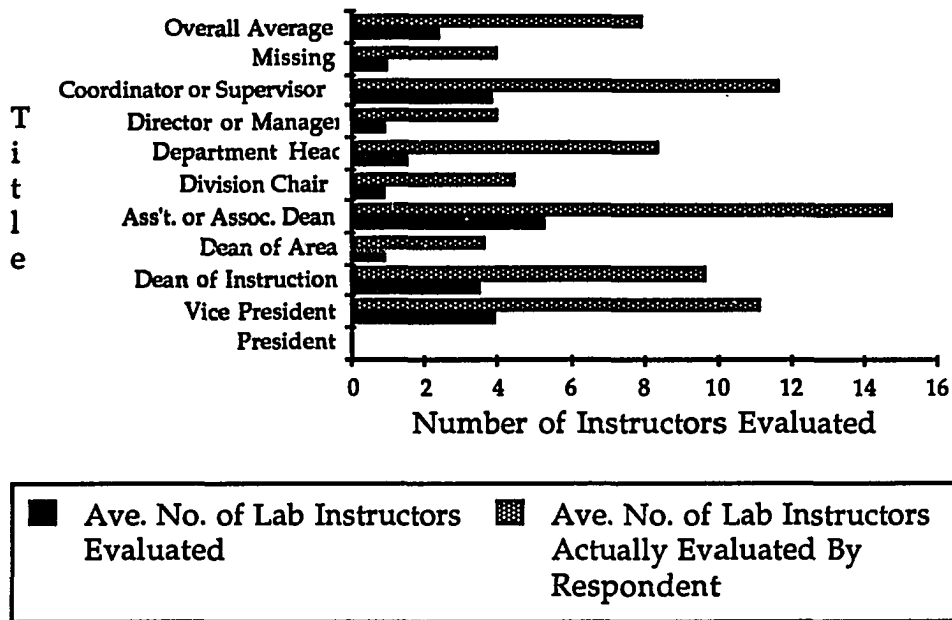


FIGURE 7: The average number of laboratory instructors evaluated by all respondents compared with the average number evaluated by those evaluating at least one such instructor

Less than 14 per cent of the respondents indicated that they evaluate shop instructors. However the overall average number of shop instructors being evaluated per respondent is over six and one half times higher when those with no responsibility to evaluate shop instructors are not included. A major difference is found at the division chair level evaluator. Division chairs are usually responsible for a specific group of instructors and therefore would most likely have responsibility for all or none of the shop instructors. Table 9 contains a comparison of the average number of shop instructors evaluated by the survey respondents.

TABLE 9: The average number of shop instructors evaluated by all respondents compared with the average number evaluated by those evaluating at least one such instructor.

Title of Evaluator	Ave. No. of Shop Instructors Evaluated	Number of Respondents Actually Evaluating Shop Instructors	Ave. No. of Shop Instructors Actually Evaluated By Respondent
President	0	0	0
Vice President	3.3	5	11.2
Dean of Instruction	2.7	7	18.6
Dean of Area	0.1	1	2.0
Ass't. or Assoc. Dean	4.0	7	15.3
Division Chair	1.4	2	26.5
Department Head	0.1	2	1.5
Director or Manager	0.7	6	4.5
Coordinator or Supervisor	1.8	3	11.0
Missing	<u>3.8</u>	<u>2</u>	<u>7.5</u>
Overall Average	1.7	35 (13.7%)	10.9

The following figure is a graphic representation of the data found in Table 9. The overall average of 10.9 for those evaluating at least one shop instructor is significantly greater than the overall average for all respondents.

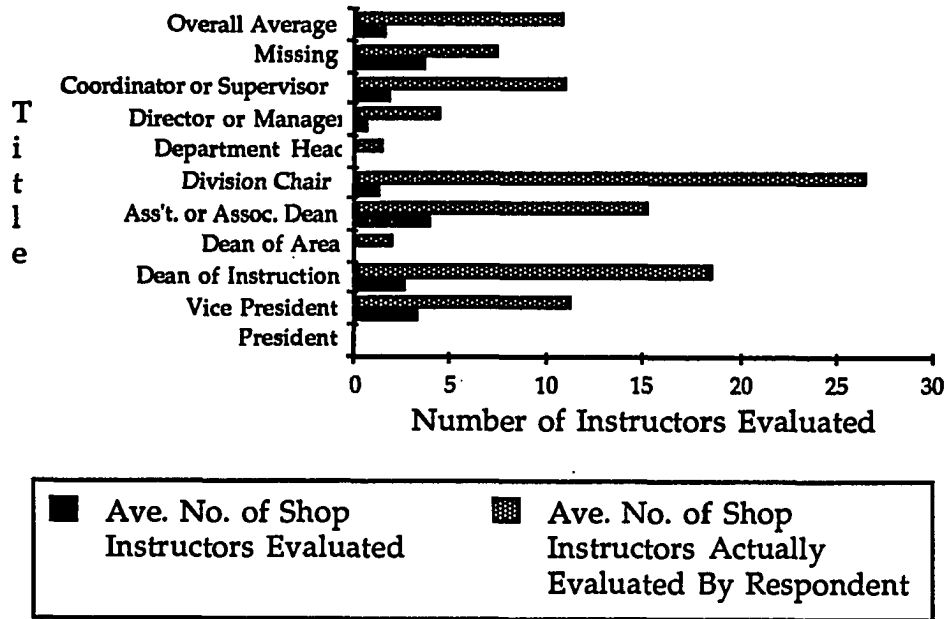


FIGURE 8: The average number of shop instructors evaluated by all respondents compared with the average number evaluated by those evaluating at least one such instructor

The respondents were given several categories of employees and asked to indicate the number of employees they were primarily responsible for evaluating. The respondents indicated that on average they evaluated over seven employees each that were not considered instructors. The next figure indicates the average number of employees evaluated in each employee category.

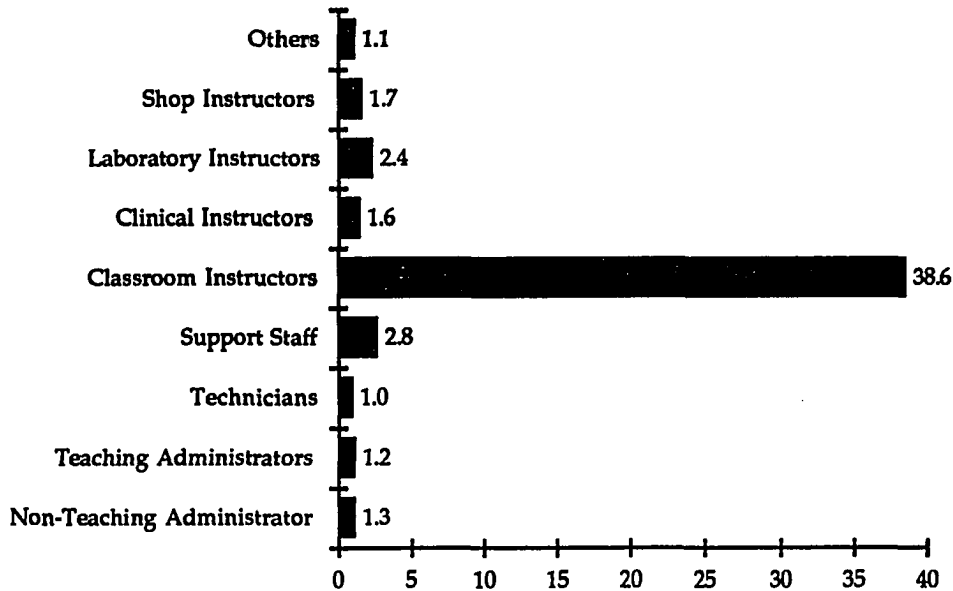


FIGURE 9. The average number of employees evaluated per instructional leader in two-year colleges as indicated by the respondents to the primary data gathering questionnaire

On average, 51.7 employees were evaluated by the instructional leaders responding to the primary data gathering questionnaire. Further, the responses to the survey query about the percentage of time spent on instructor evaluation resulted in an average of 8.9 per cent.

The instructional leaders surveyed indicated they were responsible for a wide variety of duties in addition to performance appraisal. Figure 10 graphically presents the average percentage of respondents indicating that they perform the selected duty.



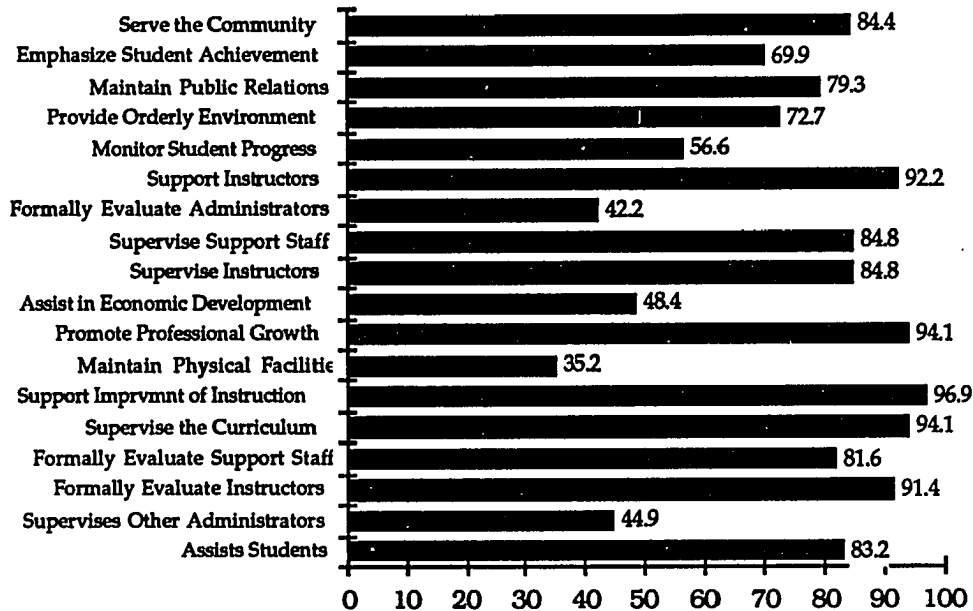


FIGURE 10: The average percentage of respondents indicating that they perform the selected duty

The highest ranked duties were; supporting improvement of instruction, supervising curriculum, promoting professional growth, providing support to instructors, and formally evaluating instructors.

### The Second Research Question

2. What are the models of evaluation being utilized in the two-year college?
  - a. Do the models used differ by type of institution or institutional size?
  - b. What percentage of two-year colleges are utilizing a "pay for performance", "merit pay", "career ladder", or other type of incentive program?

McGreal (1983) identified a taxonomy of evaluation models by gleaning the literature and researching the common practices of teacher evaluators. He was able to demonstrate that five distinct models existed. He further demonstrated that virtually every evaluation process could be classified as one of the five models. The models are the Common Law Model, Goal Setting Model, Product Model, Clinical Supervision Model, and Artistic Model. Those surveyed for this study were given a brief description of each of the models and asked to classify their instructor performance appraisal process accordingly. Table 10 is a compilation of those responses.

The relationships between the percentages in Table 10 can perhaps be more easily identified by viewing them graphically. Figure 11 is the graphic representation of the percentage of the models of evaluation utilized in the two-year colleges of the respondents to the primary data gathering questionnaire.

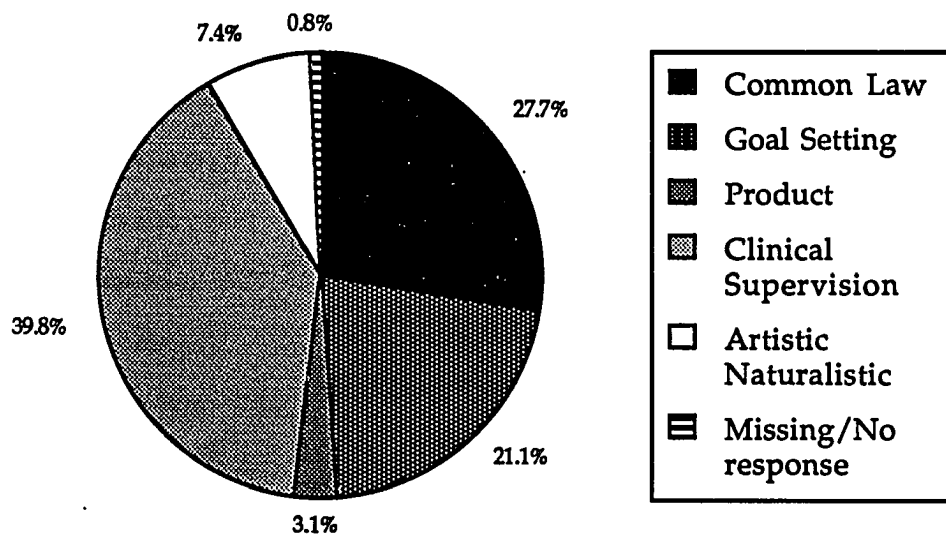


FIGURE 11: The percentage of the models of evaluation utilized in the two-year colleges of the respondents to the primary data gathering questionnaire

TABLE 10. The frequency and percentage of the models of evaluation utilized in the two-year colleges of the respondents to the primary data gathering questionnaire

Model	Frequency	Per cent
Common Law Model <sup>a</sup>	71	27.7
Goal Setting Model <sup>b</sup>	54	21.1
Product Model <sup>c</sup>	8	3.1
Clinical Supervision <sup>d</sup>	102	39.8
Artistic Naturalistic <sup>e</sup>	19	7.4
Missing/No response	2	<u>0.8</u>
Total	256	99.9

<sup>a</sup>This model is characterized by high supervisor - low teacher involvement; evaluation is synonymous with observation; major emphasis is on summative evaluation; standardized criteria; and comparative

judgements. This process usually relies on definitions, procedures, and processes that are traditional.

<sup>b</sup>This model is characterized by an emphasis on an individualized approach to evaluation. Instructors and evaluators meet and confer to set and monitor goals. Generally, no checklist of criteria is used. Self-evaluation may be a component of this model.

<sup>c</sup>This model is characterized by evaluation that is based upon the results or outcomes of student achievement tests or on competency-based evaluations, but not on methods, styles, or processes. Generally, the instruments for assessing student growth are norm-referenced tests and criterion-referenced tests.

<sup>d</sup>This model is characterized by a close relationship between the instructor and the supervisor with emphasis on collegial rather than authoritarian orientation. It takes its principal data from the classroom and is designed to improve the instructor's performance.

<sup>e</sup>This model is characterized by a belief that teaching is an art, that the quality of the performance the instructor exhibits is likened to an aesthetic experience. The evaluation is more subjective and, perhaps, less precise.

TABLE 11. The percentage of two-year colleges that have an institution or division-wide process or model

	Yes	No
College has an evaluation model	78.1%	21.9%

Additionally, 78.1 per cent of those surveyed responded that their college has an institution or division-wide evaluation process or model.

The models of evaluation being utilized in the two-year colleges are shown in Table 12. The most prevalent model in use is the Clinical Supervision Model.

TABLE 12. A comparison of the models of evaluation being utilized in two-year colleges

Model	Community College	Junior College	Vocational Technical College	Community Technical College	Technical College	Percentage All Colleges
Common Law	41	13	8	5	3	28.0
Goal Setting	34	6	6	5	3	21.3
Product	6	2	0	0	0	3.1
Clinical Supervision	73	9	10	8	2	40.2
Artistic-Naturalistic	<u>16</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>1</u>	<u>7.5</u>
Total of Colleges	170	31	25	18	9	100.0

A graphic representation of the data in Table 12 is found in Figure 12. The clinical supervision model is the model of choice in community colleges, voca-

tional technical colleges, and community technical colleges. The common law model is most prevalent in junior colleges. There appears to be no clear cut preference in technical colleges.

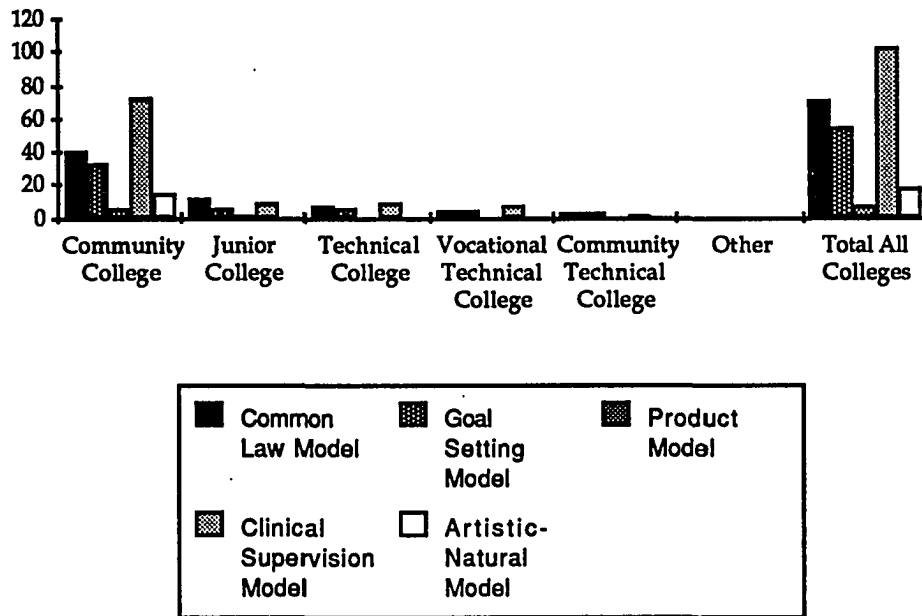


FIGURE 12. A comparison of the models of evaluation being utilized in two-year colleges

The models used do differ by type and size of institution. A series of three questions on the survey provide some insight into the differences. In response to the question about the number of formal evaluations conducted before a summative evaluation report is written, the responses ranged from none to eight. The average number of formal observations was 1.3, while the most frequent response was one. The number of formal observations made before the summative evaluation is written is reported in Table 13.

TABLE 13. The number of formal observations made before the summative evaluation is written

Number of Formal Observations	Frequency	Per cent
0	75	29.3
1	91	35.5
2	53	20.7
3	30	11.7
4	5	2.0
6	1	0.4
8	1	0.4

The second question in the series asked if there was a conference held prior to the formal observation. Surprisingly 183 of those surveyed, 71.5 per cent, indicated that no pre-observation conference was held. When asked if a conference was held after a formal observation, only 60.5 per cent, 155 respondents, stated that a post-observation conference was held. Table 14 is a representation of this data.

TABLE 14. The frequency and percentage of observation conferences used in the evaluation process of two-year colleges

Type of Conference	Yes	Per cent	No	Per cent
Pre-Observation	73	28.5	183	71.5
Post-Observation	155	60.5	101	39.5

Figures 13 through 17 depict the percentage of each type of two-year colleges using each of the various models of instructor evaluation.

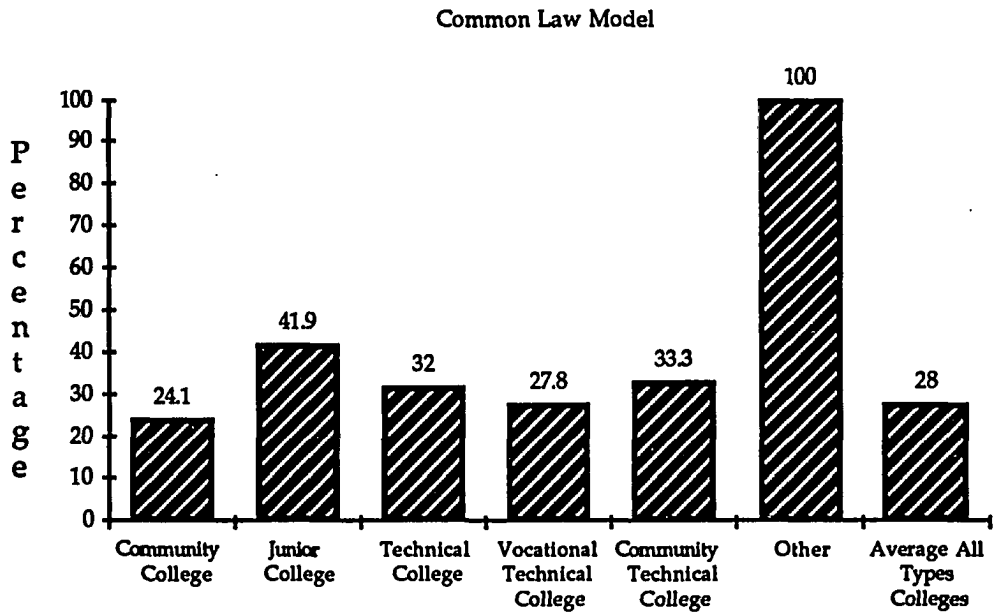


FIGURE 13. Percentage of each type of college using the common law model

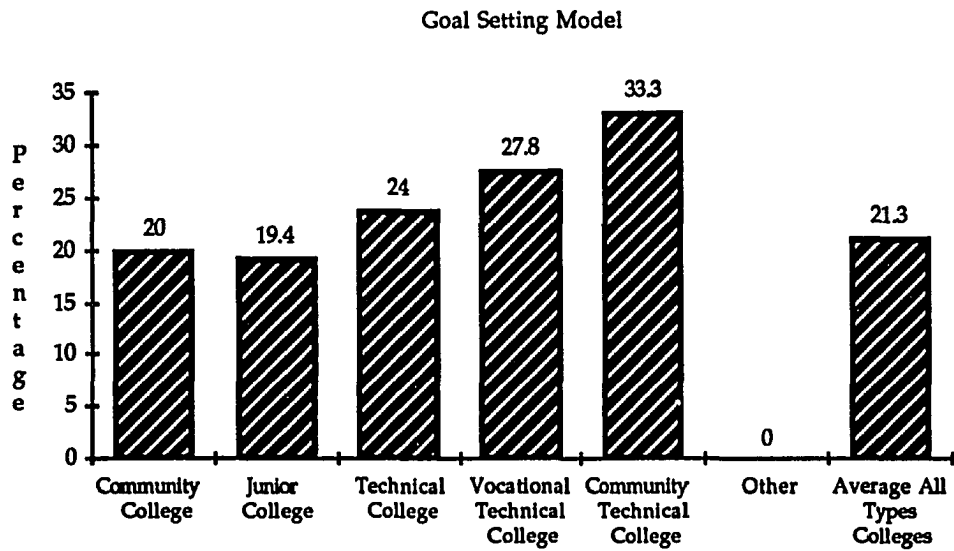


FIGURE 14. Percentage of each type of college using the goal setting model



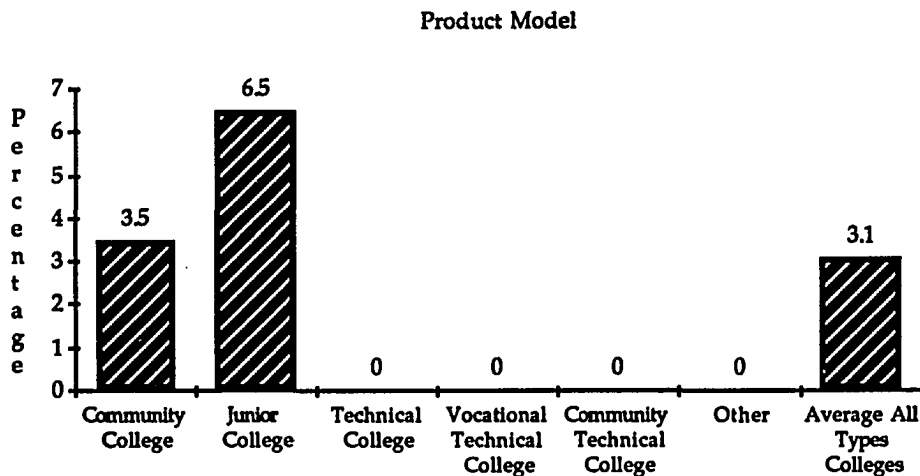


FIGURE 15. Percentage of each type of two-year college using the product model

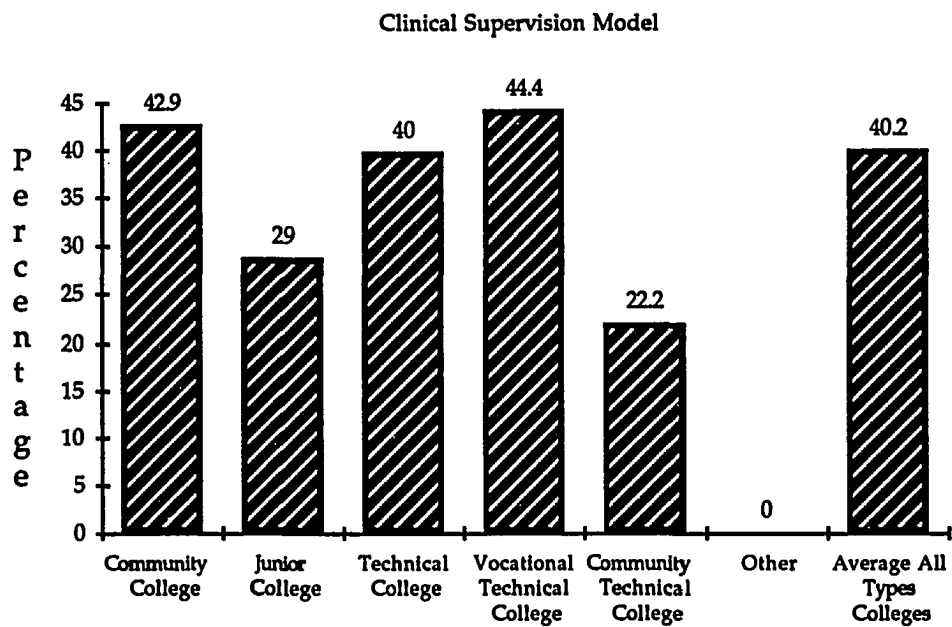


FIGURE 16. Percentage of each type of two-year college using the clinical supervision model

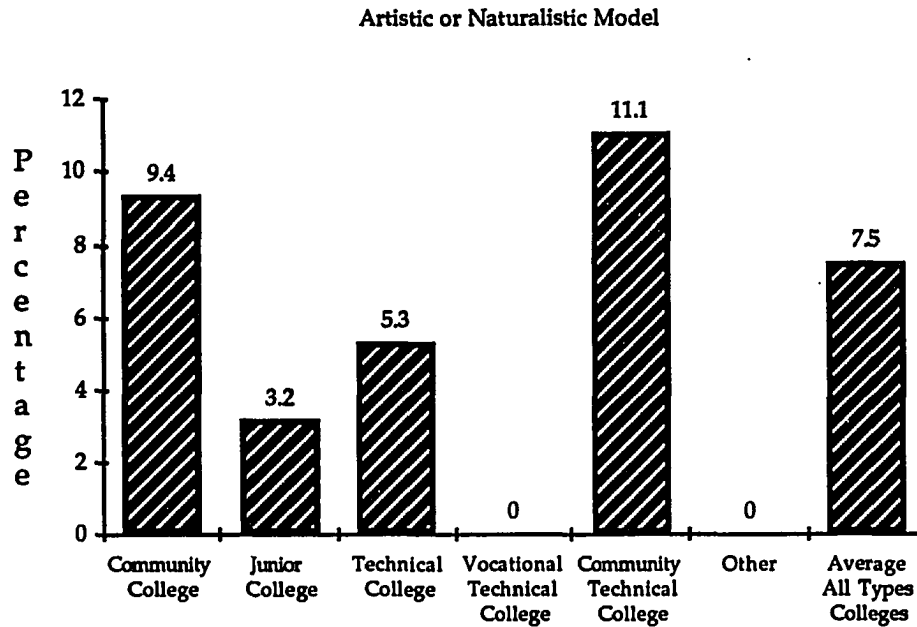


FIGURE 17. Percentage of each type of two-year college using the artistic-naturalistic model

It is interesting to note that there is evidence that the common law model, the goal setting model, and the clinical supervision model are used in all types of two-year colleges. The artistic or naturalistic model is used, although sparingly, in all types two-year colleges except vocational technical colleges. The product model is the least used and is not found in technical colleges, vocational technical colleges, or community technical colleges. This finding is curious in that, technical colleges, vocational technical colleges, and community technical colleges are normally associated with competency based education and are product centered.

The proportion relative to the amount of usage of each type of evaluation model is also consistent across the various sizes of two-year colleges. The most frequently used evaluation model in all sizes of two-year colleges is the clinical supervision model. The common law model and the goal setting model, respectively, are the next most frequently used. Figure 18 is a comparison of the type of evaluation model in use in various size two-year colleges.

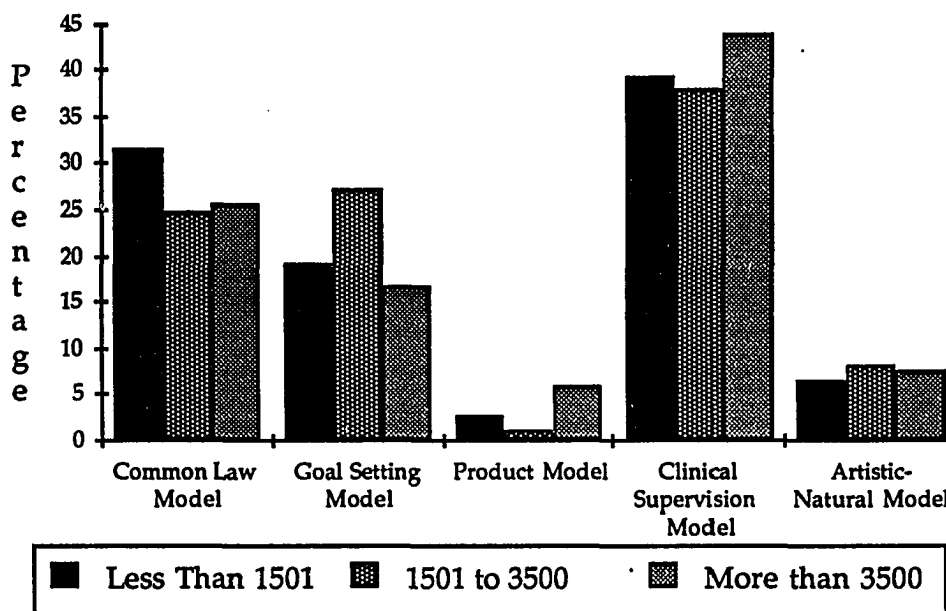


FIGURE 18. A comparison of the type of evaluation model in use in various size two-year colleges

Figures 19 through 21 depict the percentage of each size of two-year colleges using each of the various models of instructor evaluation. The most prevalent type of evaluation model in use for each size of two-year college is the clinical supervision model. The least utilized type of evaluation model is the product model.

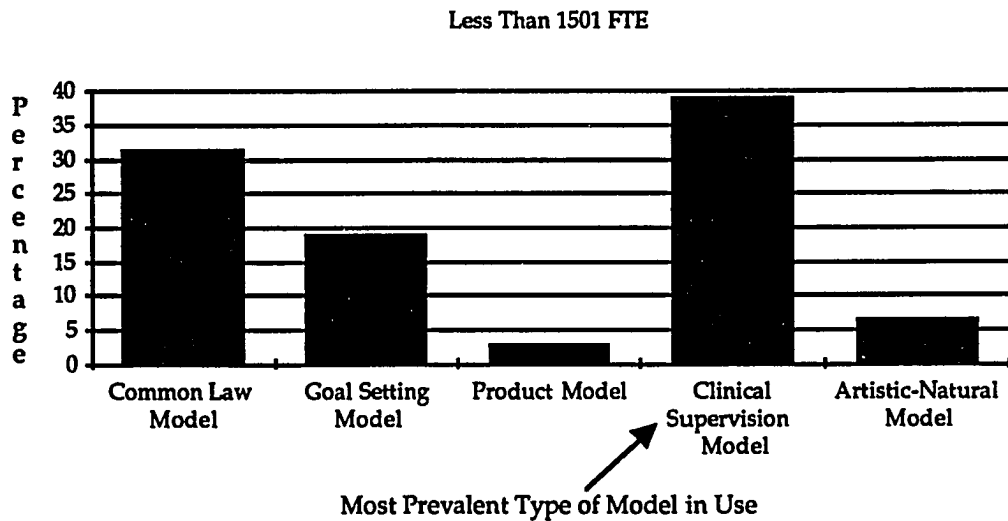


FIGURE 19. The percentage of two-year colleges with less than 1501 FTE using the various evaluation models

In the smaller colleges, those with less than 1501 full time equivalent students, the clinical supervision model was challenged in usage by the common law model. Approximately 39 percent of the respondents from small two-year colleges indicated their college uses the clinical supervision model, while nearly 32 percent of the respondents from the same size colleges indicated their model in use was the common law model. The seven percent difference between the two most prevalent models in use in small colleges is the most narrow margin of difference among the various sizes of two-year colleges.

Between 1501 and 3500 FTE

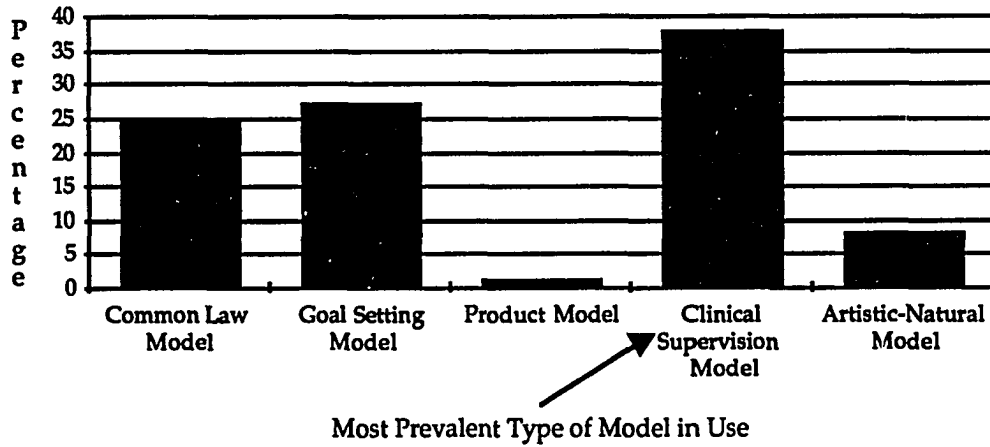


FIGURE 20. The percentage of two-year colleges with enrollments between 1501 and 3500 FTE using the various evaluation models

The mid-sized colleges, those with 1501 to 3500 full time equivalent students, use the clinical supervision model of evaluation more frequently than any of the other models. The second most frequently used model of evaluation is the goal setting model, while the common law model is a close third. Only one of the 84 respondents for mid-sized two-year colleges indicated their college used the product model. A further review of the data indicated the FTE of the college was listed as 3500, one short of the large two-year college category. The artistic or naturalistic model appeared to be more popular in the mid-sized two-year college than either of the other size categories.

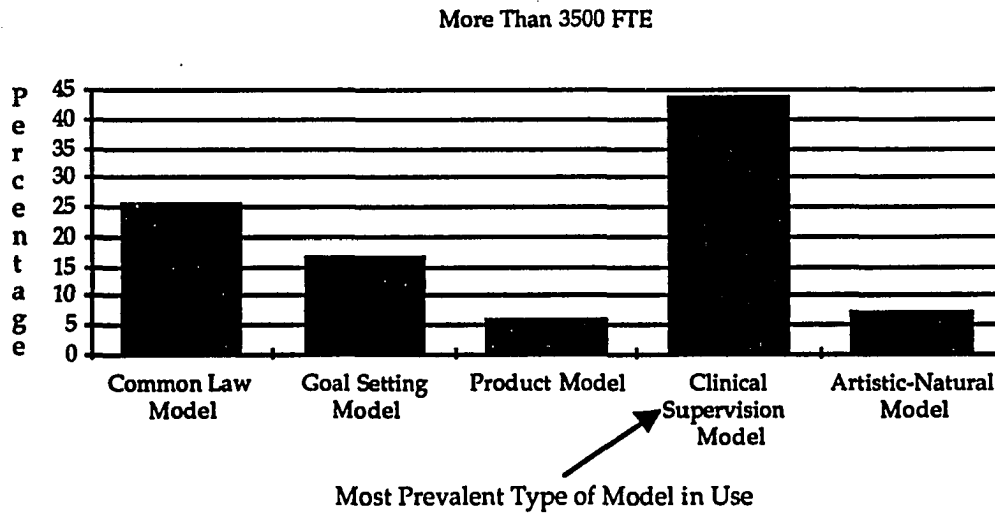


FIGURE 21. The percentage of two-year colleges with more than 3500 FTE using the various evaluation models

In the larger two-year colleges, those having enrollments of over 3500 full time equivalent students, the clinical supervision model is even more dominant for the choice of evaluation models. Nearly 44 percent of the 66 respondents for large sized two-year colleges indicated the model in use to be the clinical supervision model, while only 25.8 percent of the respondents named the second place finisher, the common law model.

According to those responding to the survey only 18 per cent of the two-year colleges they represent are utilizing a "pay for performance", "merit pay", "career ladder", or other type of incentive program. Table 15 displays the responses of the evaluators.

TABLE 15. The per cent of the two-year colleges that are utilizing a "pay for performance", "merit pay", "career ladder", or other type of incentive program

Incentive Program	Frequency	Percent
Yes	46	18
No	210	82

The following is a graphic representation of the data in Table 15.

Incentive Programs Are Utilized

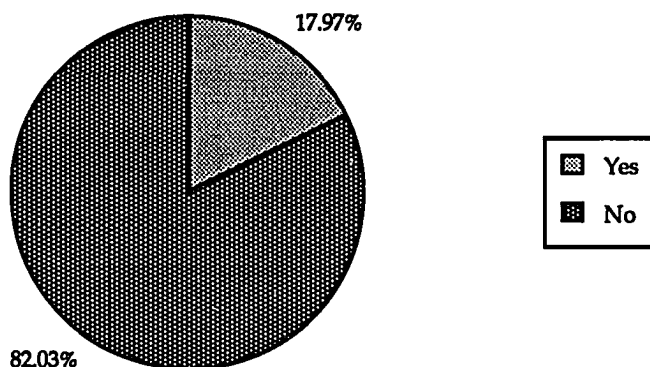


FIGURE 22. The per cent of the two-year colleges that are utilizing a "pay for performance", "merit pay", "career ladder", or other type of incentive program

Several respondents wrote comments about this question on the survey. Most of those comments indicated that in the near future their college was planning to begin an incentive plan of some sort.

A further analysis can be made by comparing the use of incentive plans in the various sizes of two-year colleges. A review of Table 16 reveals that the larger the two-year college is, the less likely an incentive plan is in use. There is a statistically significant difference at the .05 level between each of the sizes of two-year colleges as it relates to the average usage of incentive plans in instructor evaluation.

TABLE 16. A comparison, by college size, of the per cent of the two-year colleges that are and are not utilizing some type of incentive program

Incentive Programs are used	% NO	% YES
Less Than 1501 FTE	73.1	26.9
1501 to 3500 FTE	83.5	16.5
More than 3500 FTE	93.9	6.1

An additional analysis can be made by comparing the use of incentive plans in the various types of two-year colleges. A review of Table 17 reveals no clear pattern among the two-year colleges. Technical colleges are using incentive plans most frequently. On the other hand, the type of two-year college that is using incentive plans the least is the Junior College. In fact, there is a statistical difference at the .05 level between Junior colleges and all other types of two-year colleges relative to the usage of incentive plans in instructor evaluation.



TABLE 17. A comparison, by college type, of the per cent of the two-year colleges that are and are not utilizing some type of incentive program

Type of College	Incentive programs used	
	% NO	% YES
Vocational Technical	77.8	22.2
Community Technical	88.9	11.1
Technical	76.0	24.0
Community	81.2	18.2
Junior	90.6	9.4

### The Third Research Question

3. What are the elements of evaluation being used in instructor performance appraisal in two-year colleges?
  - a. What percentage of states mandate two-year college instructor evaluation?
  - b. What percentage of two-year colleges require instructor evaluation?
  - c. Is the same instrument utilized for all instructors?
  - d. To what extent is peer evaluation being utilized?
  - e. To what extent is an academic council being utilized in two-year college evaluation?
  - f. To what extent is student feedback being used for evaluation?
  - g. To what extent are student achievement data being used for evaluation?
  - h. Are outside evaluators being used?

The elements of evaluation appear to be used in a variety of evaluation systems and to varying degrees within the two-year colleges that employ them. The results of this investigation will be presented by first depicting the combined responses for all the respondents to this survey, then by providing a summary of the the results when considered by sorting the responses by size and type of two-year college.

Almost 40 per cent of the respondents to the survey indicated that the state in which their two-year college is located mandates instructor evaluation. Table 18 is a representation of the frequency and respective percentages of the responses to the question asking if the state mandates instructor performance appraisal.

TABLE 18. The per cent of the two-year colleges, represented by the respondents to this survey, that are in states where instructor evaluation is mandated

State Mandated Evaluation	Frequency	Percent
Yes	101	39.5
No	155	60.5

The figure that follows is a graphic representation of the per cent of states that mandate two-year college instructor evaluation.

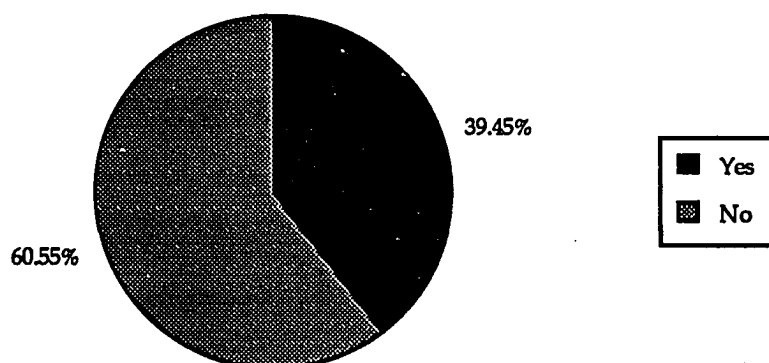


FIGURE 23. The per cent of the states that mandate instructor evaluation for two-year colleges

Although no correlation is known to exist, the smaller a two-year college is the more likely it is to be located in a state that mandates instructor performance appraisal.

TABLE 19. A comparison, by enrollment size, of the per cent of the two-year colleges in states that mandate instructor evaluation

State mandates instructor evaluation	% NO	% YES
Less Than 1501 FTE	66.3	33.7
1501 to 3500 FTE	62.4	37.6
More than 3500 FTE	50.0	50.0

When comparing two-year colleges, there are less technical colleges located in states that mandate instructor performance appraisal than other types of two-year colleges.

TABLE 20. A comparison, by type, of the per cent of the two-year colleges in states that mandate instructor evaluation

Type of College	State mandates instructor evaluation	
	% NO	% YES
Vocational Technical	61.1	38.9
Community Technical	55.6	44.4
Technical	80.0	20.0
Community	57.6	42.4
Junior	62.5	37.5

The per cent of the two-year colleges that require instructor evaluation is much greater. Over 95 per cent of the colleges represented by the respondents to this survey require instructor evaluation. The figure that follows is a graphic representation of the per cent of the two-year colleges that require instructor evaluation.

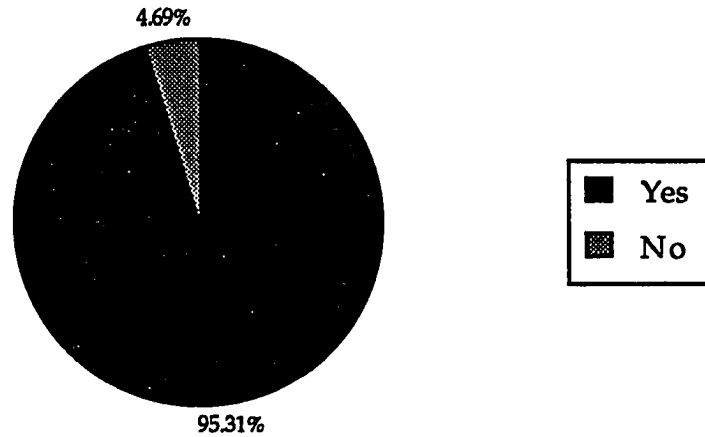


FIGURE 24. The per cent of the two-year colleges that require instructor evaluation

Another comparison can be made by considering the enrollment of the two-year colleges. This comparison reveals that all sizes of colleges are very similar in the requirement to evaluate instructors.

TABLE 21. A comparison, by enrollment size, of the per cent of the two-year colleges that require instructor evaluation

Colleges requires instructor evaluation	% NO	% YES
Less Than 1501 FTE	4.8	95.2
1501 to 3500 FTE	4.7	95.3
More than 3500 FTE	4.5	95.5

Community Technical colleges appear to be the least likely type of two-year college to require instructor performance appraisal. On the other hand, Junior colleges are most likely to require instructor evaluation.

TABLE 22. A comparison, by type, of the per cent of the two-year colleges that require instructor evaluation

Type of College	Colleges requires instructor evaluation	
	% NO	% YES
Vocational Technical	5.6	94.4
Community Technical	11.1	88.9
Technical	4.0	96.0
Community	5.3	94.7
Junior	0.0	100.0

There is no significant difference at the .05 level between the percentage of colleges that require instructor evaluation.

Even though there are a number of different kinds of instructors found in the two-year college, most of the colleges use the same instrument to evaluate all instructors. Two hundred and fifteen of the 256 respondents indicated that the same instrument was used for all instructors. Table 23 and Figure 25 are representations of the data collected.

TABLE 23. The per cent of the two-year colleges that use the same instrument for all instructors

Same Instrument for All	Frequency	Percent
Yes	215	84
No	41	16

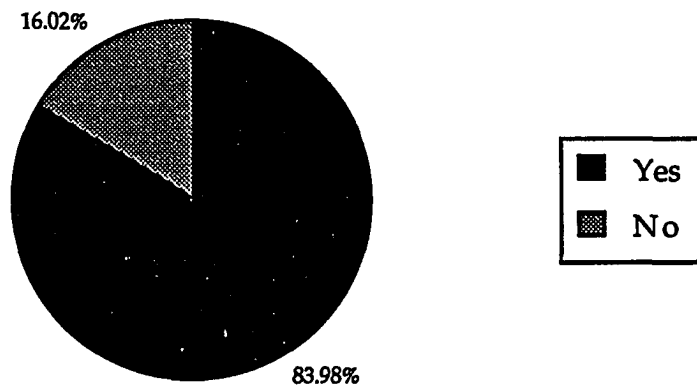


FIGURE 25. The per cent of the two-year colleges that use the same instrument for all instructors

The next table is a comparison of the per cent of the two-year colleges that use the same evaluation instrument for all instructors when sorted by the F.T.E. of the institution. The larger the two-year college is -- the less likely the same instrument is used for all instructors.

TABLE 24. A comparison, by enrollment size, of the per cent of the two-year colleges that use the same evaluation instrument for all instructors

Same instrument used for all instructors	% NO	% YES
Less Than 1501 FTE	10.6	89.4
1501 to 3500 FTE	16.5	83.5
More than 3500 FTE	24.2	75.8

When a comparison is made based on the type of two-year college, the Community Technical colleges use the same instrument for all instructors. Junior colleges are most likely to use a different instrument for evaluating different kinds of instructors.

TABLE 25. A comparison, by type, of the per cent of the two-year colleges that use the same evaluation instrument for all instructors

Type of College	Same instrument used for all instructors	
	% NO	% YES
Vocational Technical	16.7	83.3
Community Technical	0.0	100.0
Technical	16.0	84.0
Community	16.5	83.5
Junior	18.8	81.2



Another common element of evaluation involves the use of peers to participate in instructor evaluation. Over 42 per cent of the respondents to the primary data gathering questionnaire indicated that peer feedback was an element of their colleges' evaluation system. Table 26 lists the frequencies and percentages of positive and negative responses relative to the use of peer feedback.

TABLE 26. The per cent of the two-year colleges that use peer feedback as a part of instructor evaluation

Peer Feedback Used	Frequency	Percent
Yes	108	42.2
No	148	57.8

This graphic is a depiction of the data contained in Table 26.

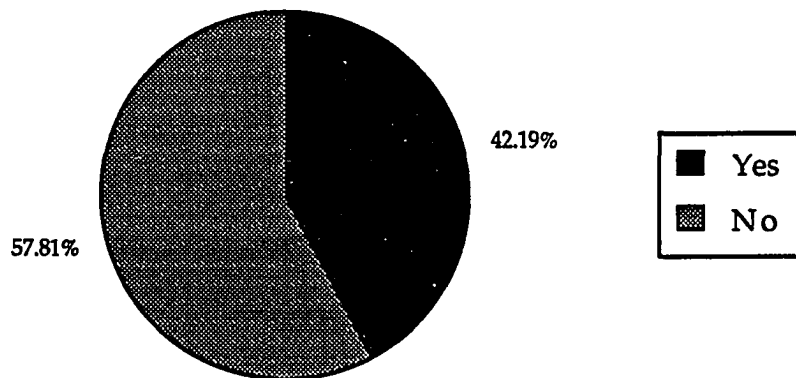


FIGURE 26. The per cent of the two-year colleges that use peer feedback as a part of instructor evaluation

A comparison, by enrollment size, reveals that peer feedback is more likely to be an element of evaluation in two-year colleges with enrollments over 3500 FTE.

TABLE 27. A comparison, by enrollment size, of the per cent of the two-year colleges that use peer feedback in instructor evaluation

Peer feedback is a part of evaluation	% NO	% YES
Less Than 1501 FTE	62.5	37.5
1501 to 3500 FTE	61.2	38.8
More than 3500 FTE	47.0	53.0

A review of the data, by type, of the per cent of the two-year colleges that use peer feedback in instructor evaluation indicates that peer feedback is more likely to be a part of evaluation in Community, Junior, or Technical colleges than it is in Vocational Technical or Community Technical colleges. In fact, Vocational Technical or Community Technical colleges seldom use peer feedback. A further review of the data led to the discovery that over 90 per cent of the Vocational Technical or Community Technical colleges that are using peer feedback categorized themselves as large (more than 3500 F.T.E.) two-year colleges. Table 28 contains the comparison data of the per cent of the different types of two-year colleges that use peer feedback in instructor evaluation.

TABLE 28. A comparison, by type, of the per cent of the two-year colleges that use peer feedback in instructor evaluation

Type of College	Peer feedback is a part of evaluation	
	% NO	% YES
Vocational Technical	77.8	22.2
Community Technical	88.9	11.1
Technical	44.0	56.0
Community	58.2	41.8
Junior	50.0	50.0

The per cent of the two-year colleges that use an academic council as a part of instructor evaluation is very small. This element of evaluation is more closely associated with four-year colleges. More than 90 per cent of the respondents to this survey stated that their two-year college did not use an academic council approach as a part of the evaluation system in their colleges. Table 29 contains the data relative to the responses about the use of academic councils in two-year colleges.

TABLE 29. The per cent of the two-year colleges that use an academic council as a part of instructor evaluation

Academic Council Used	Frequency	Percent
Yes	25	9.8
No	231	90.2

This graphic is a depiction of the data contained in Table 29.

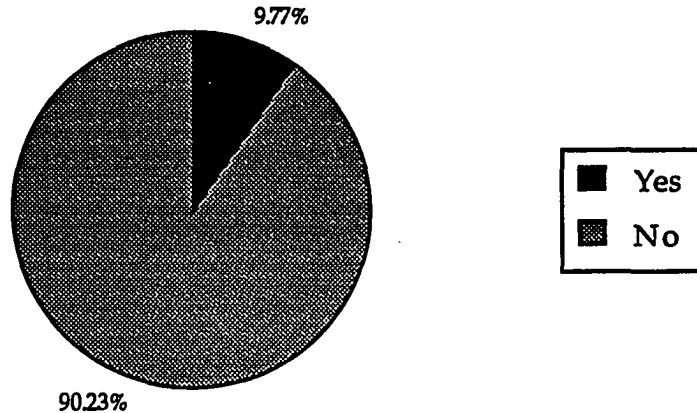


FIGURE 27. The per cent of the two-year colleges that use an academic council as a part of instructor evaluation

Although seldom used, an academic council is more likely to be an element of evaluation in two-year colleges with enrollments of less than 1501 FTE.

TABLE 30. A comparison, by enrollment size, of the per cent of the two-year colleges that use an academic council in instructor evaluation

An academic council is used	% NO	% YES
Less Than 1501 FTE	88.5	11.5
1501 to 3500 FTE	91.8	8.2
More than 3500 FTE	92.4	7.6

Academic councils are used more often in Technical colleges than any other type of college.

TABLE 31. A comparison, by type, of the per cent of the two-year colleges that use an academic council in instructor evaluation

Type of College	An academic council is used	
	% NO	% YES
Vocational Technical	94.4	5.6
Community Technical	100.0	0.0
Technical	84.0	16.0
Community	90.6	9.4
Junior	90.6	9.4

An amazingly large percentage of the respondents reported that student feedback is employed as an element of the instructor performance appraisal system in their two-year college. The following Table and its corresponding figure reveal the results of the survey.

TABLE 32. The per cent of the two-year colleges that use student feedback as a part of instructor evaluation

Student Feedback Used	Frequency	Percent
Yes	234	91.4
No	22	8.6

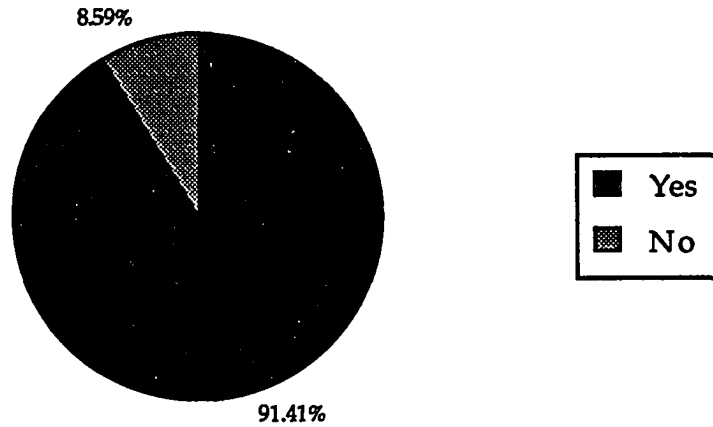


FIGURE 28. The per cent of the two-year colleges that use student feedback as a part of instructor evaluation

Student feedback is used in evaluation less often in two-year colleges with enrollments between 1500 and 3500 FTE. Student feedback appears to be used an equally amount of time in both large and small two-year colleges.

TABLE 33. A comparison, by enrollment size, of the per cent of the two-year colleges use student feedback in instructor evaluation

Student feedback is used in evaluation	% NO	% YES
Less Than 1501 FTE	6.7	93.3
1501 to 3500 FTE	11.8	88.2
More than 3500 FTE	6.1	93.9

TABLE 34. A comparison, by type, of the per cent of the two-year colleges use student feedback in instructor evaluation

Type of College	Student feedback is used in evaluation	
	% NO	% YES
Vocational Technical	11.1	88.9
Community Technical	22.2	77.8
Technical	4.0	96.0
Community	8.2	91.8
Junior	6.2	93.8

The per cent of the two-year colleges that use student achievement as a part of instructor evaluation is considerably less than those that use student feedback. Almost 22 per cent of the respondents reported that student achievement is an element of their respective two-year college's evaluation system. Table 35 contains the analysis of the data relative to the per cent of the two-year colleges that use student achievement as a part of instructor evaluation.

TABLE 35. The per cent of the two-year colleges that use student achievement as a part of instructor evaluation

Student Achievement Used	Frequency	Percent
Yes	56	21.9
No	200	78.1

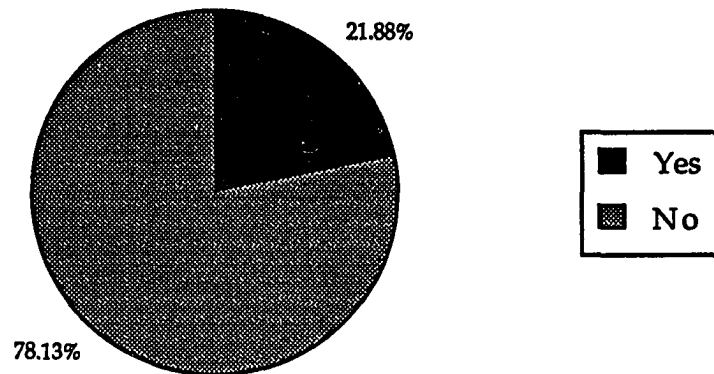


FIGURE 29. The per cent of the two-year colleges that use student achievement as a part of instructor evaluation

When viewed by enrollment size, the smaller two-year colleges appear to be more likely to employ student achievement as an element of instructor evaluation. Table 36 is a comparison, by enrollment size, of the per cent of the two-year colleges that use student achievement in instructor evaluation.

TABLE 36. A comparison, by enrollment size, of the per cent of the two-year colleges that use student achievement in instructor evaluation

Student achievement is used in evaluation	% NO	% YES
Less Than 1501 FTE	75.0	25.0
1501 to 3500 FTE	80.0	20.0
More than 3500 FTE	81.8	18.2



The next table is a comparison, by type, of the per cent of the two-year colleges that use student achievement in instructor evaluation. Student achievement is used less often in the evaluation of instructors in Community Technical colleges.

TABLE 37. A comparison, by type, of the per cent of the two-year colleges use student achievement in instructor evaluation

Type of College	Student achievement is used in evaluation	
	% NO	% YES
Vocational Technical	77.8	22.2
Community Technical	100.0	0.0
Technical	72.0	28.0
Community	78.8	21.2
Junior	75.0	25.0

The per cent of the two-year colleges represented by the respondents to this survey that use outside evaluators as a part of instructor evaluation is relatively small. Only 8.2 percent of the key participants in two-year college instructor evaluation indicated that the use of outside evaluators was an element of the evaluation system in their college. Table 38 and Figure 30 depict the two-year colleges that use outside evaluators as a part of instructor evaluation.

TABLE 38. The per cent of the two-year colleges that use outside evaluators as a part of instructor evaluation

Outside Evaluators Used	Frequency	Percent
Yes	21	8.2
No	235	91.8

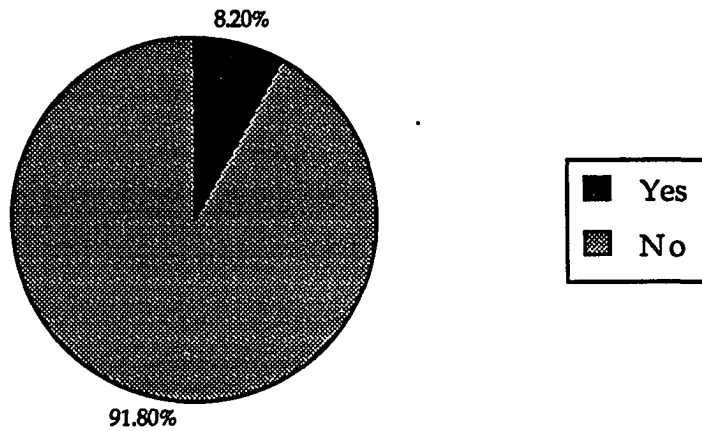


FIGURE 30. The two-year colleges that use outside evaluators as a part of instructor evaluation

TABLE 39. A comparison, by enrollment size, of the per cent of the two-year colleges that use outside evaluators in instructor evaluation

Outside evaluators used in evaluation	% NO	% YES
Less Than 1501 FTE	93.3	6.7
1501 to 3500 FTE	91.8	8.2
More than 3500 FTE	89.4	10.6

The larger the college the more likely that the evaluation process utilizes the element of outside evaluators.

TABLE 40. A comparison, by type, of the per cent of the two-year colleges that use outside evaluators in instructor evaluation

Type of College	Outside evaluators used in evaluation	
	% NO	% YES
Vocational Technical	94.4	5.6
Community Technical	100.0	0.0
Technical	80.0	20.0
Community	92.4	7.6
Junior	96.9	3.1

Outside evaluators are used more often in Technical colleges than any other type of college. Responses from representatives of Community Technical colleges indicated that outside evaluators are not used in their institutions.

#### The Fourth Research Question

4. What are the educational and technical backgrounds of those doing the supervision of instructors?
  - a. What percentage of the evaluators have had previous training in pedagogy?
  - b. What are the highest degrees held by those doing instructor evaluation?

A large majority of the respondents to this study indicate that they have previous training in pedagogy. Over 70 per cent of all the two-year college instructor evaluators responding to this survey stated that have completed a teacher preparation program. The representative of the respondents answers can be viewed in Table 41.

TABLE 41. The per cent of the two-year college instructor evaluators responding to this survey that have completed a teacher preparation program

Completed Teacher Preparation Program	Frequency	Percent
Yes	180	70.3
No	76	29.7

A comparison can also be made by the enrollment size of the two-year colleges. No pattern can be observed in the data analysis. However, the evaluators in larger two-year colleges have significantly more pedagogical training, at the .05 level, than the evaluators in mid-sized two-year colleges as it pertains to the percentage of all of them that have completed a teacher preparation program. Table 42 depicts this comparison.

TABLE 42. A comparison, by enrollment size, of the per cent of the two-year college instructor evaluators responding to this survey that have completed a teacher preparation program

Completed teacher preparation program	% NO	% YES
Less Than 1501 FTE	29.8	69.2
1501 to 3500 FTE	24.7	75.3
More than 3500 FTE	34.8	65.2

It is interesting to note that instructors are more likely to have completed a teacher preparation program in two-year colleges where the type of college is described by the word technical appearing in the title. Additionally, the instructor evaluators in Community Technical colleges indicated that they all had completed a teacher preparation program. Community colleges and Junior colleges are very similar when compared by the per cent of instructor evaluators that have completed a teacher preparation program. Table 43 is the tabular comparison of the responses of the instructor evaluators sorted by type of college.

TABLE 43. The per cent of the two-year college instructor evaluators that have completed a teacher preparation program

Type of College	Completed teacher preparation program	
	% NO	% YES
Vocational Technical	11.1	88.9
Community Technical	0.0	100.0
Technical	12.0	88.0
Community	34.1	65.9
Junior	34.1	65.6

The educational backgrounds of the respondents can be viewed in another way by reviewing the highest degree earned by the instructor evaluator. To make this comparison more quantitative, a number was assigned to the reported degrees so that a contrived mean could be compared. The highest number, nine, was given to those reporting that a post doctoral degree was the highest earned. The next highest number, eight, was given to those reporting a Ph.D. degree, and so on down. This arbitrary assignment of a number was for convenience of comparison and should not be considered a value judgement. Table 44 is a reporting of the highest degree earned by the respondents to this survey.

TABLE 44. The highest degree earned by all the two-year college instructor evaluators responding to the survey

Highest Degree Earned	Frequency	Percent
Post Doctorate	1	0.4
Ph.D.	52	20.3
Ed.D.	34	13.3
Ed. Spec.	8	3.1
Masters	144	56.3
Bachelors	5	2.0
Associates	2	0.8
Other	7	2.7
Missing	3	1.2
Contrived Mean		4.6

It is interesting to note that over 93 per cent reported their highest earned degree to be a master's degree or higher. Supposedly this percentage is larger than the the one that would be reported by a random sampling of K-12 teachers and smaller than one that would result from a random sampling of four-year college and university instructors.

The following figure is a graphic representation of the highest degree earned by all the two-year college instructor evaluators responding to the survey.

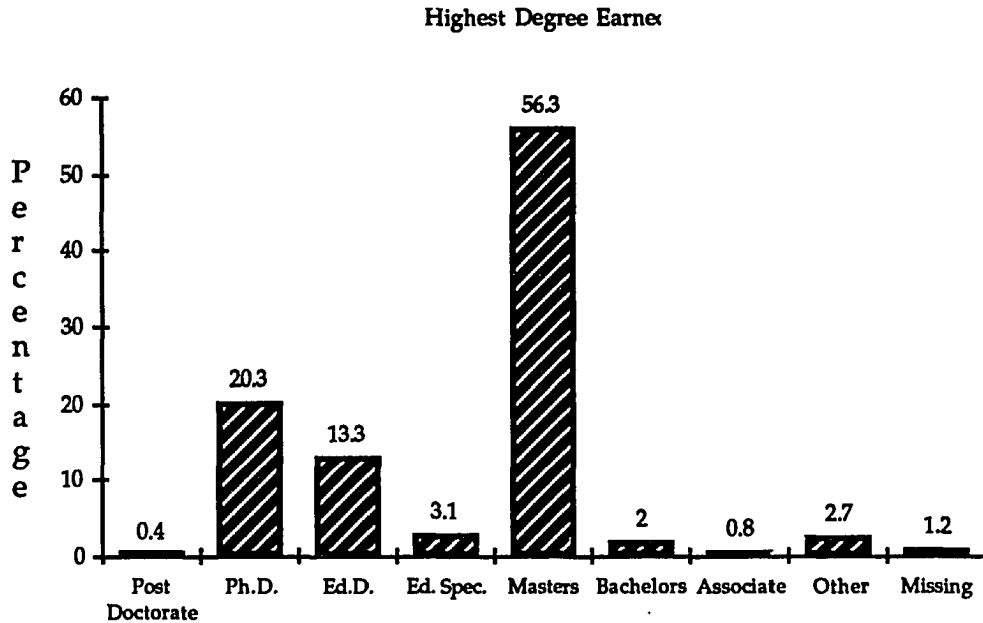


FIGURE 31. A graphic representation of the highest degree earned by all the two-year college instructor evaluators responding to the survey

Additional analysis can be made by the enrollment size of the two-year colleges. No pattern can be observed in the data analysis. However, the evaluators in larger two-year colleges a slightly larger contrived mean than the evaluators in other sized two-year colleges. The largest percentage of Ph.D. degrees are in the larger sized two-year colleges, while the largest number of Ed.D. degrees are in the smaller sized two-year colleges. The largest percentage of master's degrees are at the mid-sized colleges. Table 45 is a comparison of the highest degree earned by the two-year college instructor evaluators responding to the survey when sorted by the enrollment size of the colleges they represent.



TABLE 45. The highest degree earned by the instructor evaluators sorted by the enrollment size of the colleges they represent

Highest degree earned	Less Than 1501 FTE	1501 to 3500 FTE	More than 3500 FTE
Post Doctorate	0	0	1.5%
Ph.D.	19.2	11.8	33.3
Ed.D.	14.4	12.9	12.1
Ed. Spec.	2.9	4.7	1.5
Masters	58.7	60.0	48.5
Bachelors	1.9	3.5	0
Associates	1.9	0	0
Other	1.0	4.7	1.5
Missing	0	2.4	1.5
Contrived Mean	4.6	4.4	4.9

When sorted by the type of two-year college the respondent is associated with the contrived means are not significantly different and no pattern can be identified. However, it is noteworthy that Vocational Technical colleges have the lowest contrived mean, while the Community Technical colleges have the highest contrived mean. Table 46 contains the compared data.

**TABLE 46.** A comparison of the highest degree earned by all the two-year college instructor evaluators responding to the survey when sorted by the type of the colleges they represent

Highest degree earned	Type of College				
	Vocational Technical	Community Technical	Technical	Community	Junior
Post Doctorate	0.0%	0.0%	0.0%	0.6%	0.0%
Ph.D.	11.1	22.2	16.0	20.6	25.0
Ed.D.	0	0	12.0	15.3	15.6
Ed. Spec.	0	11.1	8.0	2.9	46.9
Masters	83.3	55.6	60.0	55.3	0
Bachelors	0	11.1	4.0	1.8	0
Associates	0	0	0	0.6	3.1
Other	0	0	0	0.6	6.3
Missing	5.6	0	0	0.6	3.1
Mean	4.2	4.8	4.7	4.6	4.4

This data were analyzed as they relate to the type of two year college. A chart was constructed to compare the percentages of the respondents reporting educational training beyond the master's degree with those reporting educational training at or below the master's degree. The graphic that follows is the chart showing the various relationships.

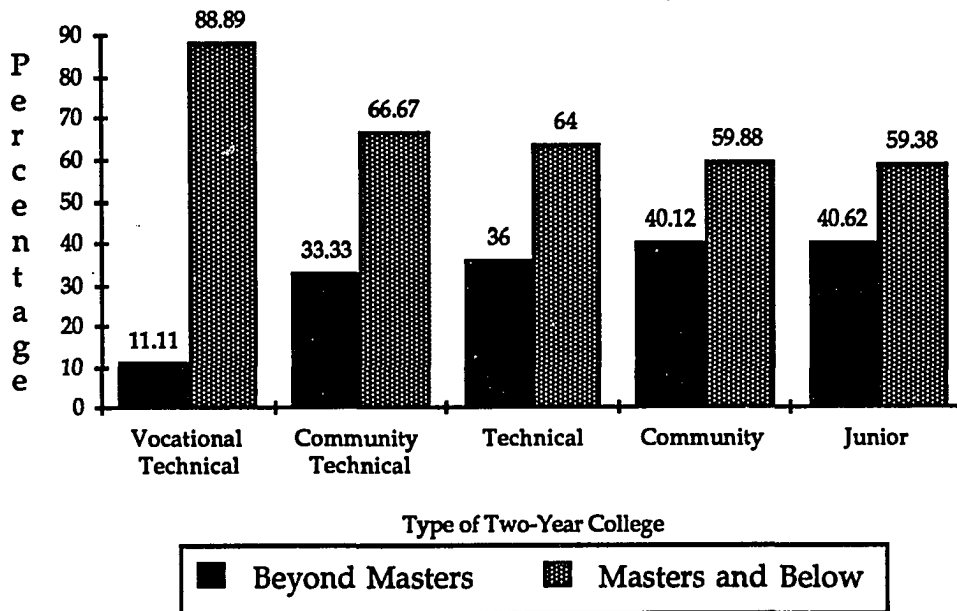


FIGURE 32. The highest degree earned by the evaluators sorted by the type of college and grouped by those with earned degrees above the master's degree level and those with master's degrees or below

The percentage of those with earned degrees beyond the master's degree is considerably smaller in the Vocational Technical colleges than in the other types of colleges.

#### The Fifth Research Question

5. What training does the evaluator have in evaluation?
  - a. How hours of evaluator training has the average evaluator had?
  - b. What differences, if any, are there between the size and types of institutions in regard to the amount of training evaluators have?

- c. Is certification in evaluation is held by the majority of evaluators?
- d. What is the perceived need for additional training in evaluation?

The amount of training that evaluators have experienced varies greatly. Over 41 per cent of the respondents indicated they have never had any training as an evaluator. On the other hand, over 8 per cent of the respondents have had more than 75 contact hours of training in evaluation. Table 47 is a comparison of the number of hours of evaluator training completed by the two-year college instructor evaluators responding to the survey.

TABLE 47. The number of hours of evaluator training completed by the instructor evaluators responding to the survey

Contact Hours of Evaluator Training	Frequency	Percent
None or Blank	106	41.4
1 to 25 Hours	70	27.3
26 to 50 Hours	42	16.4
51 to 75 Hours	17	6.6
More Than 75 Hours	21	8.2

Figure 33 is a graphic representation of the number of hours of evaluator training completed by the two-year college instructor evaluators responding to

the survey. Over two-thirds of those evaluating instructors in two-year colleges have less than 25 contact hours of training in evaluation.

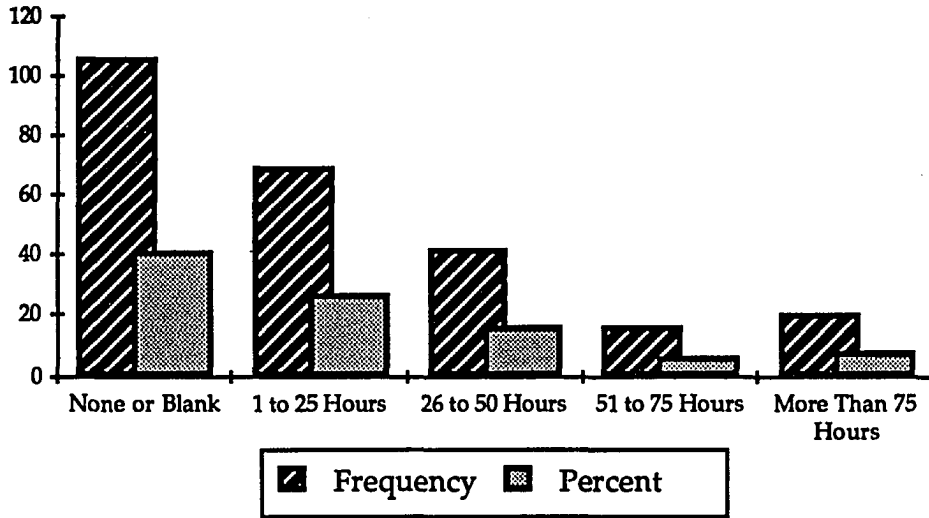


FIGURE 33. The number of hours of evaluator training completed by the two-year college instructor evaluators

A comparison, by college enrollment size, of the number of hours of evaluator training completed by the two-year college instructor evaluators responding to the survey is found in Table 48. There are no patterns from which to predict the characteristics of a particular size of two-year college.

TABLE 48. A comparison, by college enrollment size, of the number of hours of evaluator training completed by instructor evaluators

Hours of evaluator training	Less Than 1501 FTE	1501 to 3500 FTE	More than 3500 FTE
None or Blank	44.2	35.3	40.9
1 to 25 contact hours	25.0	29.4	28.8
26 to 50 contact hours	17.3	18.8	12.1
51 to 75 contact hours	2.9	10.6	7.6
More than 75 hours	8.7	5.9	9.1

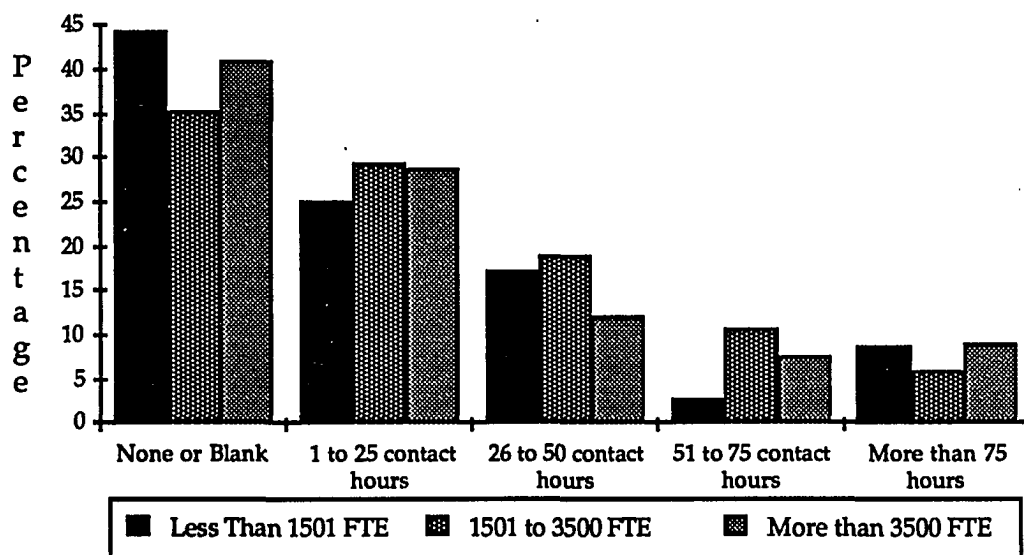


FIGURE 34. A graphic comparison, by college enrollment size, of the number of hours of evaluator training completed by the two-year college instructor evaluators

Another comparison can be made by sorting the responses by college type. The Community Technical college instructor evaluators appear to have more training in evaluation than instructor evaluators in other types of two-year colleges. The Community college instructor evaluators appear to have less training in evaluation than instructor evaluators in other types of two-year colleges.

TABLE 49. A comparison, by college type, of the number of hours of evaluator training completed by the two-year college instructor evaluators

Hours of Evaluator Training	Type of College				
	Vocational Technical	Community Technical	Technical	Community	Junior
None or Blank	38.9	11.1	28.0	44.1	46.9
1 to 25 contact hours	22.2	33.3	24.0	29.4	21.9
26 to 50 contact hours	22.2	44.4	28.0	12.9	15.6
51 to 75 contact hours	5.6	11.1	8.0	6.5	6.3
More than 75 hours	11.1	0.0	8.0	7.1	9.4

Figure 35 is a graphic representation of the comparison by college type of the number of hours of evaluator training completed by the two-year college instructor evaluators responding to the survey.

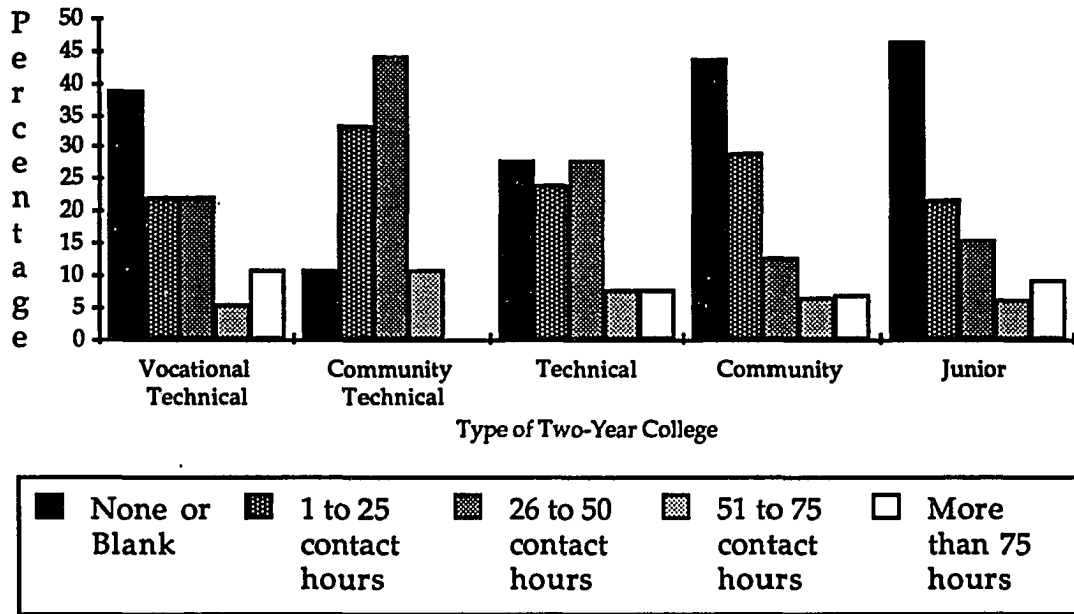


FIGURE 35. A graphic representation by type of college of the number of hours of evaluator training completed by the two-year college instructor evaluators responding to the survey

A comparison of the number and percentage of the two-year college instructor evaluators responding to the survey that are certified as instructor evaluators is shown in Table 50.

It is interesting to note that less than 4 per cent of the respondents are certified as instructor evaluators. A further sorting of those certified revealed no pattern from state to state.



TABLE 50. A comparison of the number and percentage of the two-year college instructor evaluators that are certified as instructor evaluators

Certified Evaluator	Frequency	Percent
Yes	10	3.9
No	246	96.1

A graphic representation of the number and percentage of the two-year college instructor evaluators responding to the survey that are certified as instructor evaluators is depicted by Figure 36.

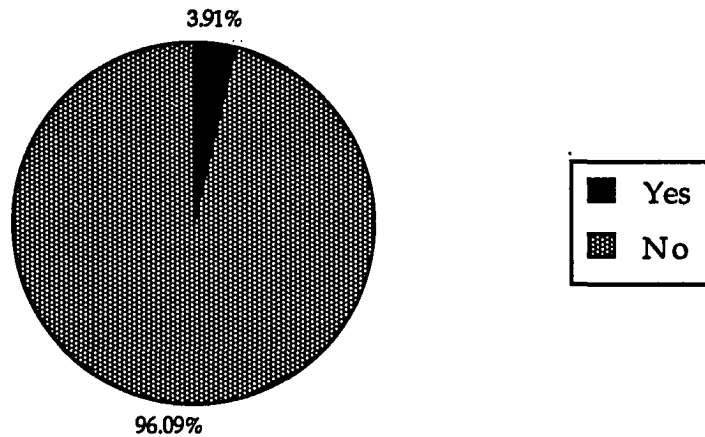


FIGURE 36. A graphic representation of the number and percentage of the two-year college instructor evaluators responding to the survey that are certified as instructor evaluators

A further comparison was made by sorting the data by enrollment size. There were no significant results. Table 51 contains the comparison of the two-year college instructor evaluators responding to the survey that are certified as instructor evaluators when sorted by college enrollments.

TABLE 51. A comparison, by college enrollment size, of the two-year college instructor evaluators that are certified as instructor evaluators

Is certification held by evaluator	Less Than		More than
	1501 FTE	1501 to 3500 FTE	3500 FTE
Percent responding "Yes"	3.8	5.9	0.0

Additional comparisons were made by two-year college type. Only two type of two-year colleges were found to have certified instructor evaluators, the Community Technical and Technical colleges. A full one-third of the respondents from Technical colleges reported that they were certified as instructor evaluators. Table 52 is a comparison, by college type, of the two-year college instructor evaluators responding to the survey that are certified as instructor evaluators.

TABLE 52. A comparison, by college type, of the two-year college instructor evaluators that are certified as instructor evaluators

Is certification held by evaluator	Type of College				
	Vocational Technical	Community Technical	Technical	Community	Junior
Percent responding "Yes"	0.0%	5.6%	33.3%	0.0%	0.0%

All but 17 of the 256 respondents indicated at least some need for additional training in instructor evaluation. The average perceived need is 5.02, the median and the mode are both 5 and the standard deviation is 2.65. It is also interesting to note that the dispersion of the responses is fairly consistent across the possible answers. A comparison of the respondents' perception of their need for more training in instructor evaluation is pictured in Table 53.

TABLE 53. A comparison of the respondents' perception of their need for more training in instructor evaluation

Needs More Training		Frequency	Percent
Least need	0	17	6.6
	1	15	5.9
	2	25	9.8
	3	20	7.8
	4	14	5.5
	5	49	19.1
	6	28	10.9
	7	36	14.1
	8	29	11.3
Greatest Need	9	23	9.0
Mean			5.02

A graphic representation of the comparison of the respondents' perception of their need for more training in instructor evaluation is shown in Figure 37. The mid-point of the scale is at 4.5 and almost two-thirds of the respondents indicate a perceived need above that level.

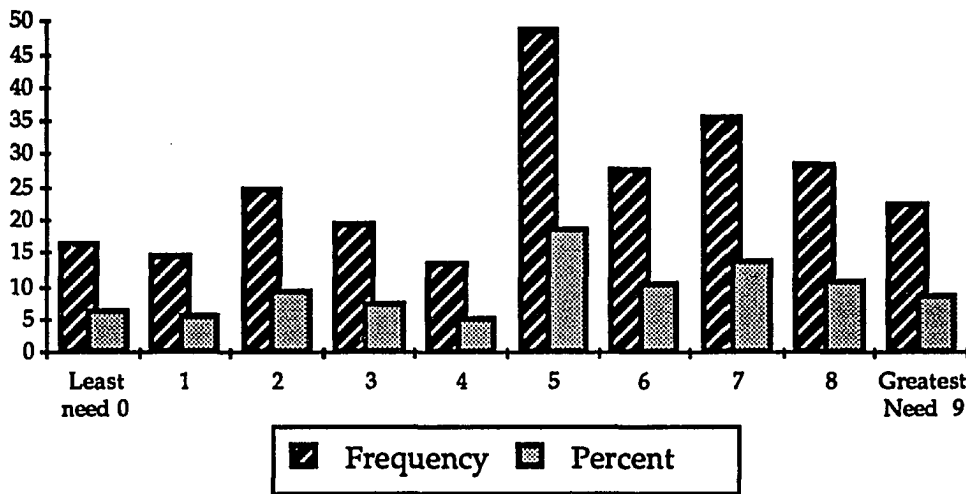


FIGURE 37. A graphic representation of the comparison of the respondents' perception of their need for more training in instructor evaluation

When compared by the enrollment sizes of the two-year colleges, the average perceived need for additional training in instructor evaluation by instructor evaluators responding to the survey reveals a pattern. The smaller the two-year college the higher the perceived need for additional training in instructor evaluation. This pattern is in contrast to the pattern of the average number of instructors evaluated in various sizes two-year colleges where the instructor evaluators in larger colleges have significantly larger average numbers of instructors to evaluate. A graphic representation, by college enrollment size, of the

average number of instructors evaluated by the typical instructor evaluator is depicted by Figure 38.

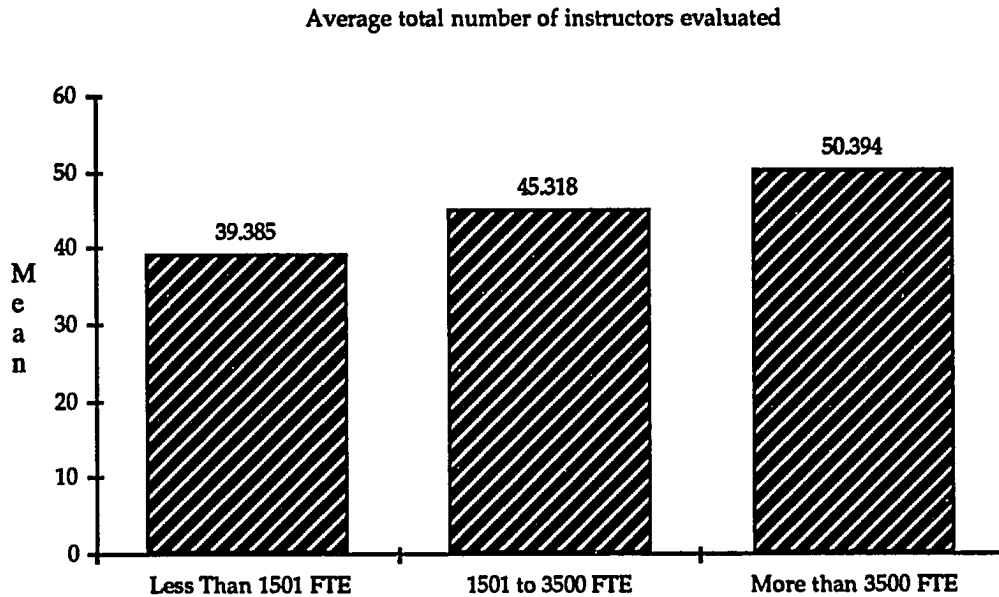


FIGURE 38. A graphic representation, by college enrollment size, of the average number of instructors evaluated by the typical instructor evaluator

Table 54 is a comparison, by college enrollment size, of the percentage two-year college instructor evaluators responding to the survey as to their perceived need for additional training in instructor evaluation. Most instructor evaluators perceive a moderate to strong need for more training in instructor evaluation.

TABLE 54. A comparison, by college enrollment size, of the perceived need for additional training in instructor evaluation

Perceived need for additional training		Less Than		More than
		1501 FTE	1501 to 3500 FTE	3500 FTE
Least Need	0	3.8	5.9	10.6
	1	3.8	3.5	12.1
	2	11.5	9.4	7.6
	3	8.7	7.1	7.6
	4	3.8	9.4	3.0
	5	17.3	20.0	21.2
	6	8.7	15.3	9.1
	7	13.5	11.8	18.2
	8	14.4	11.8	6.1
Most need	9	14.4	5.9	4.5
Mean		5.5	5.0	4.4

A further review can be made by comparing the mean of the responses of instructor evaluators from various size two-year colleges. Their perceived need for additional training in instructor evaluation is the highest among those from Community Technical colleges and Technical colleges. The lowest perceived need for additional training in instructor evaluation is among those from Community colleges and Junior colleges.

TABLE 55. A comparison, by college type, of the perceived need for additional training in instructor evaluation

Perceived need for additional training		Type of College				
		Vocational Technical	Community Technical	Technical	Community	Junior
Least Need	0	5.6	0	0	7.6	6.3
	1	5.6	0	4.0	5.9	9.4
	2	5.6	0	0	10.6	15.6
	3	11.1	11.1	12.0	7.6	3.1
	4	5.6	11.1	8.0	4.1	9.4
	5	27.8	0	16.0	20.6	15.6
	6	11.1	11.1	16.0	10.0	12.5
	7	11.1	11.1	24.0	14.1	9.4
	8	5.6	44.4	16.0	9.4	12.5
Most need	9	11.1	11.1	4.0	10.0	6.3
Mean		5.0	6.8	5.8	4.9	4.6

The perceived need for additional instructor evaluation training by the instructor evaluators responding to this survey when sorted by the type of college is graphically represented in the figure that follows.

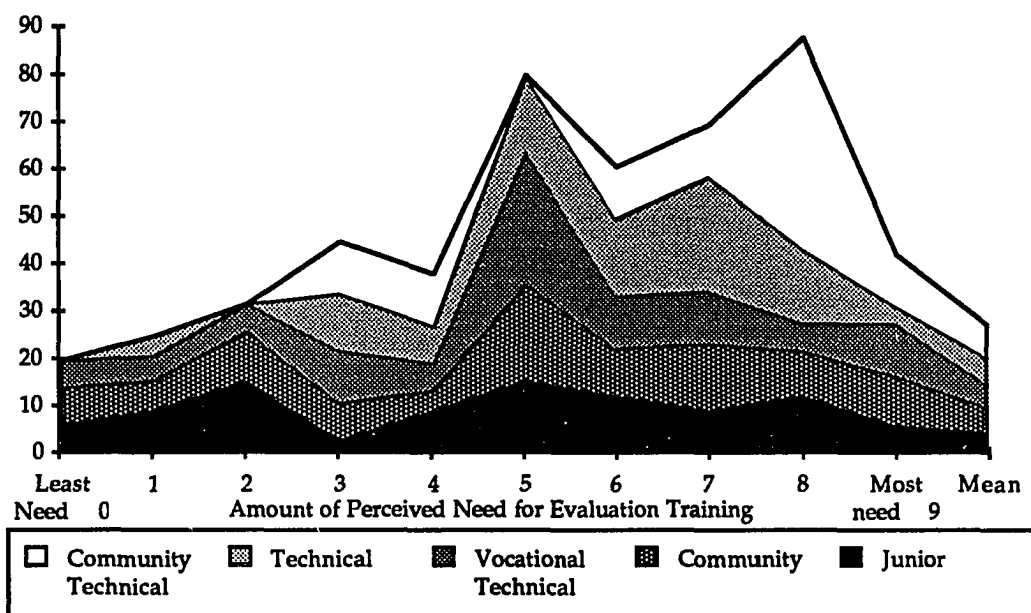


FIGURE 39. A graphic representation, by college type, of the perceived need for additional instructor evaluation by the instructor evaluators responding to this survey

#### The Sixth Research Question

6. Is the frequency of occurrence of the evaluation models in use, as classified by McGreal's taxonomy, dependent on the type of the two-year college?

This research question resulted in the formulation of the first null hypothesis for this study: "The distribution of models in use among the two-year colleges is independent of the five types of two-year colleges." In order to answer the research question and test the hypothesis the chi square ( $\chi^2$ ) test of independence was used. The  $\chi^2$ -distributions comprise a family of distributions, each determined by a single degrees-of-freedom value. The  $\chi^2$ -value itself does not



indicate where statistical significance lies in the case of a statistically significant test. However, the cell frequencies can be inspected for those contributing large  $(O-E)^2/E$  values to the  $\chi^2$ -value.

A 5 X 6 contingency table was constructed from the data for models by types of college. The calculated value of  $\chi^2$  is 15.50867, the critical value of the test statistic is 31.410 with the alpha set at .05. Since the calculated value of  $\chi^2$  does not exceed the critical value, the null hypothesis is retained. The result of the null hypothesis being retained leads to the conclusion that, the use of various evaluation models is independent of the type of two-year college. Hence, the conclusion that there is no difference between the types of two-year colleges regarding their use of different models of evaluation. There is, however, a caveat. In the 5 X 6 contingency table for this statistical test over 20 per cent of the expected cell frequencies are small. As a result the *Yates Correction for Continuity* was calculated as an adjustment. The results did not change enough to cause a rejection of the null hypothesis.

Additionally, the data were regrouped for further study. A 2 X 5 contingency table was constructed by combining the Community college and Junior college data together as one column and combining all the other colleges' data into a second column. This was a logical combination due to the fact that all the other colleges had within their names the word, technical. The new  $\chi^2$ -value was computed to be 4.969, the critical value of the test statistic is 9.488. Once again, the calculated value of  $\chi^2$  does not exceed the critical value therefore the null hypothesis is retained. The conclusion remains, that there is no difference

between the types of two-year colleges regarding their use of different models of evaluation.

#### The Seventh Research Question

7. Is the frequency of occurrence of the evaluation models in use, as classified by McGreal's taxonomy, dependent on the size of the two-year college?

This research question resulted in the formulation of the second null hypothesis for this study: "The distribution of models in use among the two-year colleges is independent of the size of two-year colleges." In order to answer the research question and test the hypothesis the chi square ( $\chi^2$ ) *test of independence* was used.

A 3 X 5 contingency table was constructed from the data for models by types of college. The calculated value of  $\chi^2$  is 6.56475, the critical value of the test statistic is 15.507 with the alpha set at .05. Since the calculated value of  $\chi^2$  does not exceed the critical value, the null hypothesis is retained. The result of the null hypothesis being retained leads to the conclusion that, the use of various evaluation models is independent of the size of two-year college. Hence, the conclusion that there is no difference between the sizes of two-year colleges regarding their use of different models of evaluation.

#### The Eighth Research Question

8. Is there a relationship between the model in use and the type or size of two-year college?

This research question resulted in the formulation of the third null hypothesis for this study: "There is no relationship between the model in use and the type or size of two-year college." In order to answer this research question and to test the hypothesis contingency coefficients were determined for both the type of two-year college and the size of two-year college.

The contingency coefficient is the correlation coefficient that is the appropriate procedure to determine the relationship between two-variables measured on the discrete-nominal scale that are not discrete dichotomies. This statistic is computed directly from the  $\chi^2$  statistic. As a result, this index of relationship can be used for the 5 X 6 and 3 X 5 contingency tables constructed in this study to test the first two hypotheses. As with any correlation coefficient, the greater the value of the coefficient, the stronger the relationship between the two variables of the contingency table. The maximum value of the contingency coefficient is not 1.0. A maximum value must be determined for each given contingency table.

When considering the data relative to the relationship between the type of two-year colleges and the evaluation model in use for the 5 X 6 contingency table a contingency coefficient of 0.23988 is computed. The maximum value of the 5 X 6 contingency table is computed by determining the square root of the smallest number of categories (5) minus 1 divided by the smallest number of categories. That computation results in 0.89443 as the maximum value for this particular contingency coefficient. Thus the magnitude of the relationship between the model in use and the type of two-year college is interpreted as being low.

When considering the data relative to the relationship between the size of two-year colleges and the evaluation model in use for the 3 X 5 contingency table a contingency coefficient of 0.15873 is computed. The maximum value of the 3 X 5 contingency table is computed by determining the square root of the smallest number of categories (3) minus 1 divided by the smallest number of categories. That computation results in 0.81650 as the maximum value for this particular contingency coefficient. Thus the magnitude of the relationship between the model in use and the size of two-year college is interpreted as being very low.

#### The Ninth Research Question

9. What are the efficacy ratings for each personnel standard?

Each of the 21 personnel standards advanced by the Joint Committee (1988) were included on the primary data gathering questionnaire. The respondents were asked to rate the quality of the model or process their institution uses on each of the standards. The responses were converted to numeric values using the following scheme: strongly agree = 4; agree = 3; disagree = 2; strongly disagree = 1; and not applicable or missing = 0. Zero values were not included in the data tabulation. A comparison of the statistics relative to the responses, about the Personnel Standards, of the two-year college instructor evaluators participating in this survey is contained in Table 56.

TABLE 56. The statistics relative to the responses about the Personnel Standards

Personnel Standard	Mean	Median	Std. Dev	Variance
Service Orientation	3.185	3	0.707	0.500
Formal Evaluation Guidelines	3.188	3	0.619	0.383
Conflict of Interest	3.090	3	0.625	0.390
Access To Evaluation Reports	3.601	4	0.559	0.312
Interactions With Employees	3.475	4	0.607	0.368
Constructive Orientation	3.262	3	0.670	0.449
Defined Uses	3.174	3	0.685	0.470
Evaluator Credibility	3.123	3	0.661	0.437
Functional Reporting	2.941	3	0.771	0.595
Follow-up Impact	3.053	3	0.677	0.458
Practical Procedures	3.069	3	0.623	0.388
Political Viability	2.972	3	0.719	0.518
Fiscal Viability	2.733	3	0.776	0.602
Defined Role	2.946	3	0.702	0.493
Work Environment	2.914	3	0.625	0.391
Doc. of Procedures	3.046	3	0.623	0.389
Valid Measurement	2.911	3	0.688	0.473
Reliable Measurement	2.741	3	0.708	0.501
Systematic Data Control	3.270	3	0.626	0.392
Bias Control	2.844	3	0.711	0.506
Monitoring Eval. System	2.931	3	0.799	0.639

For the purposes of this study, the mean response of the participants is called the efficacy rating. The efficacy rating for each of the personnel standards is contained in the following figure.



FIGURE 40. The efficacy ratings for the Personnel Standards advanced by the Joint Committee (1988)

The highest efficacy rating is for Standard P4: Access to Personnel Evaluation Reports. The standard (Joint Committee, 1988) states, "Access to reports of personnel evaluation should be limited to those individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured." The lowest efficacy rating is for Standard FS 13: Fiscal Viability. The standard (Joint Committee, 1988) states, "Adequate time and resources should be

provided for personnel evaluation activities, so that evaluation plans can be effectively and efficiently implemented."

The Joint Committee (1988) subdivided the standards into four categories that correspond to four basic attributes of sound evaluation: propriety, utility, feasibility, and accuracy. Basically, the standards require that evaluations be proper, useful, feasible, and accurate. The next series of figures depict the efficacy ratings within each of the four categories.

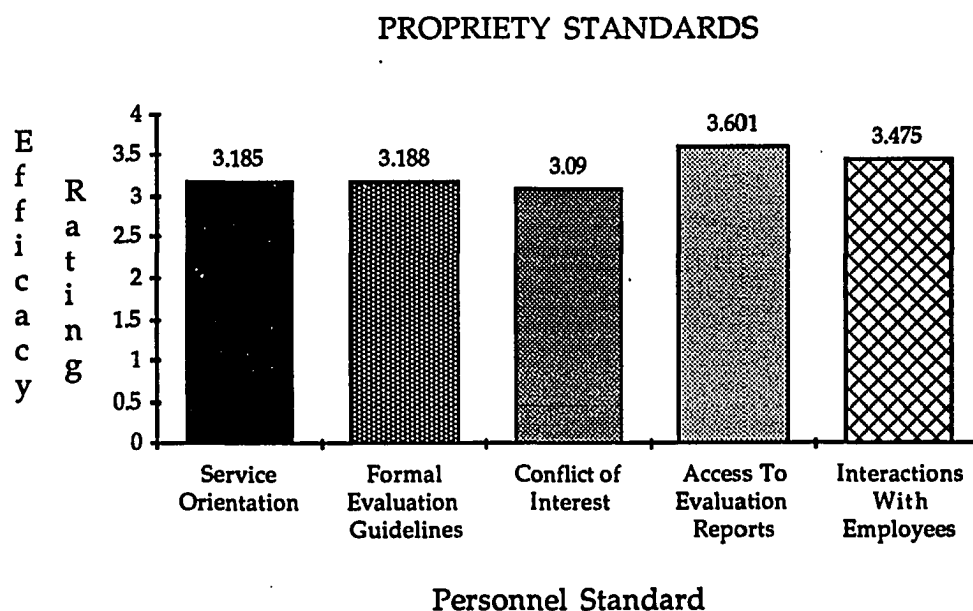


FIGURE 41. The efficacy ratings for the propriety standards advanced by the Joint Committee (1988)

UTILITY STANDARDS

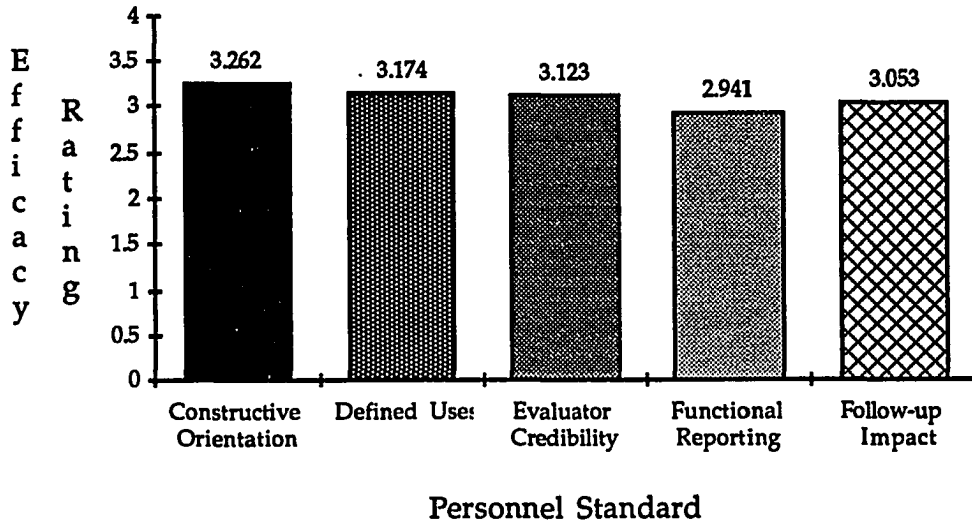


FIGURE 42. The efficacy ratings for the utility standards advanced by the Joint Committee (1988)

FEASIBILITY STANDARDS

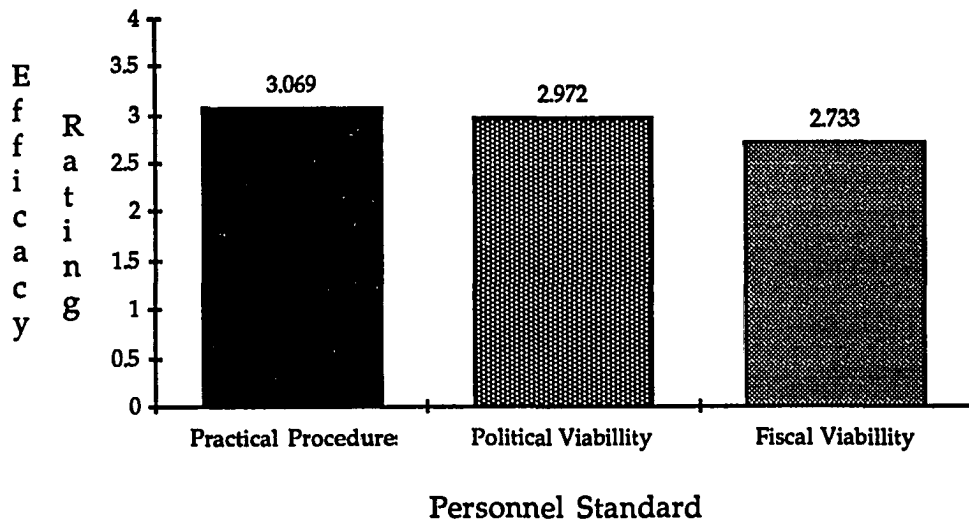


FIGURE 43. The efficacy ratings for the feasibility standards advanced by the Joint Committee (1988)



## ACCURACY STANDARDS

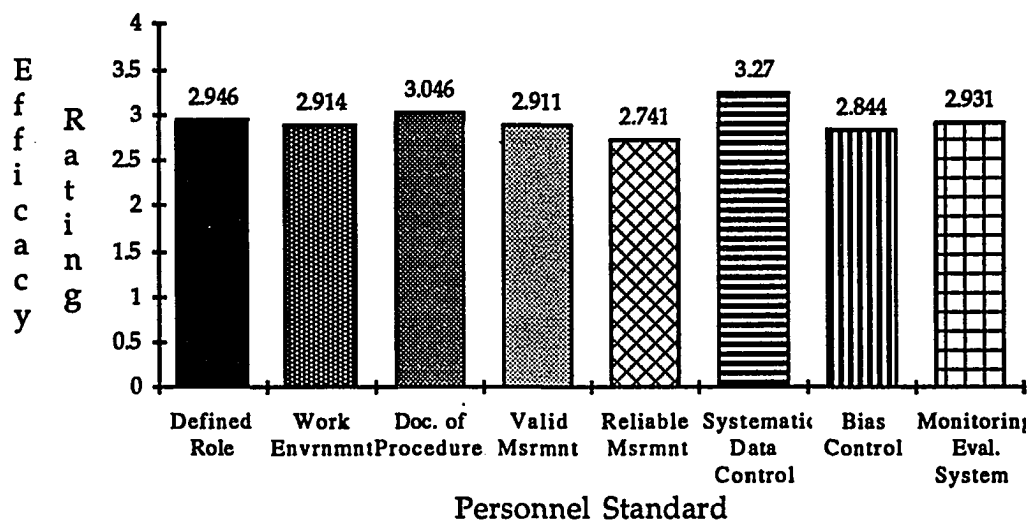


FIGURE 44. The efficacy ratings for the accuracy standards advanced by the Joint Committee (1988)

It is of interest to compare the average efficacy ratings for the four categories of personnel standards. This comparison reveals that the respondents felt their evaluation systems were adequate in almost all areas.

TABLE 57. A comparison of the efficacy ratings for the four categories of the Personnel Standards

Category of Personnel Standard	Average Efficacy Rating
Propriety Standards	3.308
Utility Standards	3.111
Feasibility Standards	2.925
Accuracy Standards	3.002

The lowest average efficacy rating is in the category of feasibility standards. The highest average efficacy rating is in the category of propriety standards. However, the differences in the averages are not statistically significant. A graphic representation of the average efficacy ratings for the four categories of the personnel standards advanced by the Joint Committee (1988) is contained in Figure 44.

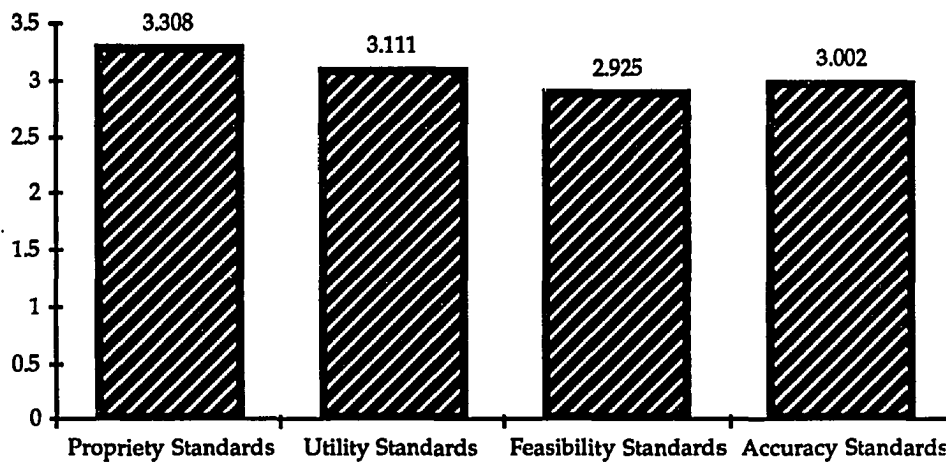


FIGURE 45. The average efficacy ratings for the four categories of the personnel standards advanced by the Joint Committee (1988)

#### The Tenth Research Question

10. Are differences in the efficacy ratings on each personnel standard attributable to the model in use, or the type or size of two-year college?

This research question resulted in the formulation of the rest of the null hypotheses for this study. Each hypothesis, 4 through 6, will be reprinted here-

in-after as their test results are reported. The fourth hypothesis stated: "There is no significant difference in the efficacy rating for each personnel standard between the evaluation models in use." In order to answer this research question and to test this hypothesis, an analysis of variance (ANOVA) was computed to determine whether the observed differences can be reasonably attributed to chance or whether there is reason to suspect true differences between the five models of instructor evaluation.

The ANOVA is used to test the hypothesis by calculating the F statistic. The observed significance level is obtained by comparing the calculated F to values of the F distribution with  $k-1$  and  $N-k$  degrees of freedom. The observed significance level is the probability of obtaining an F statistic at least as large as the one calculated when all population means are equal. If this probability is small enough, the hypothesis is rejected.

A significant F statistic indicates only that the population means are probably unequal. It does not pinpoint where the differences are. A post hoc multiple comparison test is used to decide which pairs or combinations of means are not equal. The Scheffé method is recommended when there is a significant F-ratio in the ANOVA, and when the group sample sizes are unequal (Hinkle et al., 1979). The Scheffé method is conservative for pairwise comparisons of means. It requires larger differences between means for significance than most of the other comparison methods.

The SPSSx procedure ONEWAY was used to produce the one-way analysis of variance. Output includes sums of squares, degrees of freedom, mean squares, and the F ratio and its significance. Additionally, the SPSSx procedure RANGES=SCHEFFÉ was used to generate the post hoc test results.

Output includes either the statement that no two groups are significantly different at the 0.05 level or a matrix denoting the pairs or groups that are significant at the 0.05 level. The results of the ANOVA is presented in Table 59.

The results of the ANOVA revealed significant differences in the efficacy ratings between the five models of instructor evaluation for each of seven of the personnel standards. However, when the Scheffé post hoc multiple comparison test was conducted, only personnel standard US9, FUNCTIONAL REPORTING, produced pinpointed differences between the five models. The clinical supervision model has a significantly higher efficacy rating than the goal setting model at the 0.05 level for the FUNCTIONAL REPORTING (US9) personnel standard. The analysis of variance of personnel standard US9: Functional Reporting by model of evaluation is depicted by Table 58.

TABLE 58. Analysis of variance of personnel standard US9: Functional Reporting by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	6.1713	1.5428	2.649	0.0344*
Within Groups	213	124.0535	0.5824		
Total	217	130.2248			

\*Significant at .05 level.

TABLE 59. The results of an ANOVA of the efficacy ratings for each personnel standard between the five models of evaluation

Standard	Observed F Ratio	Calculated F Probability	Retain or Reject Hypothesis
PS1	2.5205*	0.0418	REJECT
PS2	0.6108	0.6553	RETAIN
PS3	0.6673	0.6154	RETAIN
PS4	0.4035	0.8060	RETAIN
PS5	3.2456*	0.0128	REJECT
US6	2.6819*	0.0322	REJECT
US7	1.2095	0.3074	RETAIN
US8	1.7003	0.1507	RETAIN
US9	2.6490*	0.0344	REJECT
US10	1.9216	0.1076	RETAIN
FS11	3.0921*	0.0165	REJECT
FS12	3.0687*	0.0172	REJECT
FS13	2.2389	0.0656	RETAIN
AS14	2.0472	0.0886	RETAIN
AS15	1.5301	0.1943	RETAIN
AS16	1.6810	0.1552	RETAIN
AS17	2.1553	0.0749	RETAIN
AS18	1.2146	0.3055	RETAIN
AS19	3.0912*	0.0166	REJECT
AS20	1.7398	0.1422	RETAIN
AS21	0.7442	0.5628	RETAIN

\*Significant at .05 level.

Because the Scheffé method keeps the experiment-wise Type I error rate at the a priori alpha level, it failed to pinpoint which pairs or combinations of means are significantly different in six of the seven cases. As a result, further statistical analysis was necessary. First, a much less conservative post hoc test was performed. *Duncan's multiple range test* was used to produce multiple comparisons between all groups. Once potential differences were identified, orthogonal contrasts were planned and conducted. The hypotheses reflected by the contrasts are independent and therefore the experiment-wise Type I error rate is maintained at the 0.05 alpha level. This procedure is statistically more powerful than the Scheffé method.

The use of the orthogonal contrasts produced some robust statistics and pinpointed which pairs of means are significantly different for each case. The results of the statistical analysis have been tabled separately and will be discussed in the following pages.

Three significant differences were identified between the means of the efficacy ratings for the personnel standard PS1: Service Orientation on the models of evaluation in use. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the SERVICE ORIENTATION (PS1) personnel standard. Also, the efficacy rating of the artistic-naturalistic model is significantly higher than the goal setting model. Table 60 contains the data for the analysis of variance of personnel standard PS1: Service Orientation by model of evaluation.

TABLE 60. Analysis of variance of personnel standard PS1: Service Orientation by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	4.7463	1.1866	2.5205	0.0418*
Within Groups	242	113.9257	0.4708		
Total	246	118.6721			

Contrast Coefficient Matrix					
	Common Law Model	Goal Setting Model	Product Model	Clinical Supervision Model	Artistic-Naturalistic Model
Contrast 1	-1	0	0	0	1
Contrast 2	0	-1	0	0	1
Contrast 3	-1	0	0	1	0
Contrast 4	0	-1	0	1	0

Pooled Variance Estimate					
	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	0.3019	0.1816	1.663	242	0.098
Contrast 2	0.3700	0.1872	1.977	242	0.049*
Contrast 3	0.2330	0.1074	2.170	242	0.031*
Contrast 4	0.3011	0.1166	2.583	242	0.010*

\*Significant at .05 level.

Two significant differences were identified between the means of the efficacy ratings for the personnel standard PS5: Interactions with Evaluatees on the models of evaluation in use. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the INTERACTIONS WITH EMPLOYEES (PS5) personnel standard. Table 61 contains the data for the analysis of variance of per-

sonnel standard PS5: Interactions with Evaluatees by model of evaluation

TABLE 61. Analysis of variance of personnel standard PS5: Interactions with Evaluatees by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	4.6276	1.1569	3.2456	0.0128*
Within Groups	248	88.4001	0.3565		
Total	252	93.0277			

Contrast Coefficient Matrix					
	Common Law Model	Goal Setting Model	Product Model	Clinical Supervision Model	Artistic-Naturalistic Model
Contrast 1	-1	0	1	0	0
Contrast 2	-1	2	2	1	0
Contrast 3	0	-1	0	1	0
Contrast 4	0	0	1	0	-1

Pooled Variance Estimate					
	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	0.4261	0.2227	1.913	248	0.057
Contrast 2	0.2839	0.0923	3.077	248	0.002*
Contrast 3	0.2004	0.1005	1.995	248	0.047*
Contrast 4	0.4167	0.2537	1.642	248	0.102

\*Significant at .05 level.

Two significant differences were identified between the means of the efficacy ratings for the personnel standard US6: Constructive Orientation on the models of evaluation in use. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the CONSTRUCTION ORIENTATION (US6) person-



nel standard. Table 62 contains the data for the analysis of variance of personnel standard US6: Constructive Orientation by model of evaluation

TABLE 62. Analysis of variance of personnel standard US6: Constructive Orientation by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	4.7025	1.1756	2.6819	0.0322*
Within Groups	245	107.3975	0.4384		
Total	249	112.1000			

Contrast Coefficient Matrix					
	Common Law Model	Goal Setting Model	Product Model	Clinical Supervision Model	Artistic-Naturalistic Model
Contrast 1	-1	0	1	0	0
Contrast 2	0	-1	1	0	0
Contrast 3	-1	0	0	1	0
Contrast 4	0	-1	0	1	0

Pooled Variance Estimate					
	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	0.3000	0.2625	1.143	245	0.254
Contrast 2	0.3154	0.2663	1.184	245	0.237
Contrast 3	0.2734	0.1028	2.660	245	0.008*
Contrast 4	0.2888	0.1121	2.576	245	0.011*

\*Significant at .05 level.

The means of the efficacy ratings for the personnel standard FS11: Practical Procedures have two significant differences between the models of evaluation in use. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for

the PRACTICAL PROCEDURES (FS11) personnel standard. Table 63 contains the data.

TABLE 63. Analysis of variance of personnel standard FS11: Practical Procedures by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	4.6777	1.1694	3.0921	0.0165*
Within Groups	241	91.1475	0.3782		
Total	245	95.8252			

Contrast Coefficient Matrix					
	Common Law Model	Goal Setting Model	Product Model	Clinical Supervision Model	Artistic-Naturalistic Model
Contrast 1	-1	0	0	1	0
Contrast 2	0	-1	1	0	0
Contrast 3	0	-1	0	1	0
Contrast 4	0	0	1	0	-1

Pooled Variance Estimate					
	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	0.2325	0.0965	2.410	241	0.017*
Contrast 2	0.3480	0.2339	1.488	241	0.138
Contrast 3	0.3159	0.1056	2.990	241	0.003*
Contrast 4	0.3056	0.2613	1.169	241	0.243

\*Significant at .05 level.

Two significant differences were identified between the means of the efficacy ratings for the personnel standard FS12: Political Viability on the models of evaluation in use. The clinical supervision model has a significantly higher efficacy rating than either the artistic-naturalistic model or the goal setting model at

the 0.05 level for the POLITICAL VIABILITY (FS12) personnel standard. Table 64 contains the data.

TABLE 64. Analysis of variance of personnel standard FS12: Political Viability by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	6.1942	1.5485	3.0687	0.0172*
Within Groups	239	120.6050	0.5046		
Total	243	126.7992			

Contrast Coefficient Matrix					
	Common Law Model	Goal Setting Model	Product Model	Clinical Supervision Model	Artistic-Naturalistic Model
Contrast 1	-1	0	1	0	0
Contrast 2	0	0	0	1	-1
Contrast 3	-1	0	0	1	0
Contrast 4	0	0	1	0	-1

Pooled Variance Estimate					
	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	0.4412	0.2655	1.662	239	0.098
Contrast 2	0.4091	0.1820	2.247	239	0.026*
Contrast 3	0.3225	0.1119	2.882	239	0.004*
Contrast 4	0.5278	0.3018	1.748	239	0.082

\*Significant at .05 level.

Both the clinical supervision model and the product model have a significantly higher efficacy rating than the artistic-naturalistic model at the 0.05 level for the SYSTEMATIC DATA CONTROL (AS19) personnel standard. Table 63 contains the data.

TABLE 65. Analysis of variance of personnel standard AS19: Systematic Data Control by model of evaluation

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	4	4.7218	1.1804	3.0912	0.0166*
Within Groups	241	92.0302	0.3819		
Total	245	96.7520			

Contrast Coefficient Matrix					
	Common Law Model	Goal Setting Model	Product Model	Clinical Supervision Model	Artistic-Naturalistic Model
Contrast 1	-1	0	1	0	0
Contrast 2	0	-1	1	0	0
Contrast 3	0	0	0	1	-1
Contrast 4	0	0	1	0	-1

Pooled Variance Estimate					
	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	0.3862	0.2312	1.671	241	0.096
Contrast 2	0.4552	0.2344	1.942	241	0.053
Contrast 3	0.4356	0.1582	2.753	241	0.006*
Contrast 4	0.6806	0.2626	2.592	241	0.010*

\*Significant at .05 level.

This tenth research question also resulted in the formulation of the fifth hypothesis, it stated: "There is no significant difference in the efficacy rating for each personnel standard between the types of two-year colleges." In order to answer this research question and to test this hypothesis, an analysis of variance (ANOVA) was computed to determine whether the observed differences can be reasonably attributed to chance or whether there is reason to suspect true differences between the six types of two-year colleges. The results

of the ANOVA is presented in Table 66. There were no significant differences.

Additionally, this tenth research question also resulted in the formulation of the sixth hypothesis, it stated: "There is no significant difference in the efficacy rating for each personnel standard between the sizes of two-year colleges." In order to answer this research question and to test this hypothesis, an analysis of variance (ANOVA) was computed to determine whether the observed differences can be reasonably attributed to chance or whether there is reason to suspect true differences between the three sizes of two-year colleges. The results of the ANOVA is presented in Table 67. There were no significant differences.

#### Summary

A narrative profile of the "typical" instructor evaluator can be constructed by compiling the average responses to the data gathering questionnaire. This "typical" instructor evaluator is most likely to be found in a Community College with less than 1501 FTE. Even though the state probably doesn't have a law mandating instructor evaluation his/her college is likely to require it. He or she is most often called a Dean of Instruction or Academic Affairs and is responsible to evaluate at least 44 instructors and 8 non-instructors. The types of the 44 instructors usually include 38 classroom instructors, 2 clinical instructors, 2 laboratory instructors, and 2 shop instructors. In addition to spending an average of almost 9 percent of their time on instructor evaluation they also perform at least 14 other duties. These other duties include supporting improvement of instruction, supervising curriculum, promoting professional growth, providing support to instructors, assisting students, serving the community, supervising support

staff and others. The evaluator holds at least a master's degree, has completed a teacher preparation program, but holds no certificate as an evaluator and has not had any training as an evaluator. The "typical" instructor evaluator perceives a moderate to strong need for more training in instructor evaluation.

The Clinical Supervision model of instructor evaluation is most likely to be utilized by the "typical" evaluator. Most of the time he or she will use the same instrument to evaluate all instructors and make only one formal observation of the instructor. No preobservation conference is likely to be held. However, a post-observation conference probably will be held. Pay for performance, merit pay, or a career ladder is probably not available as an incentive program for the evaluator. Normally student feedback will be considered in the evaluation of the instructor, but student achievement, outside evaluators, an academic council, or peer feedback will not normally be considered.

Overall, the efficacy ratings were positive. Finally, there are very few differences in the efficacy ratings on each personnel standard that are attributable to the model in use, or the type or size of two-year college.

TABLE 66. The results of an ANOVA of the efficacy ratings for each personnel standard between the six types of colleges

Standard	Observed F Ratio	Calculated F Probability	Retain or Reject Hypothesis
PS1	0.4705	0.7573	RETAIN
PS2	1.3010	0.2703	RETAIN
PS3	1.3650	0.2471	RETAIN
PS4	0.7766	0.5413	RETAIN
PS5	0.6805	0.6060	RETAIN
US6	0.8815	0.4756	RETAIN
US7	1.6177	0.1704	RETAIN
US8	0.6825	0.6047	RETAIN
US9	0.2505	0.9091	RETAIN
US10	0.4484	0.7735	RETAIN
FS11	0.6765	0.6089	RETAIN
FS12	0.5416	0.7053	RETAIN
FS13	0.8307	0.5068	RETAIN
AS14	1.1758	0.3221	RETAIN
AS15	1.9452	0.1039	RETAIN
AS16	1.2481	0.2914	RETAIN
AS17	1.8593	0.1185	RETAIN
AS18	1.6452	0.1639	RETAIN
AS19	0.1117	0.9783	RETAIN
AS20	1.4461	0.2197	RETAIN
AS21	0.6779	0.6079	RETAIN

TABLE 67. The results of an ANOVA of the efficacy ratings for each personnel standard between the three sizes of colleges

Standard	Observed F Ratio	Calculated F Probability	Retain or Reject Hypothesis
PS1	1.3299	0.2664	RETAIN
PS2	0.3926	0.6757	RETAIN
PS3	0.6436	0.5264	RETAIN
PS4	0.1195	0.8874	RETAIN
PS5	0.2378	0.7886	RETAIN
US6	1.0796	0.3413	RETAIN
US7	0.0137	0.9864	RETAIN
US8	0.3771	0.6862	RETAIN
US9	0.7947	0.4530	RETAIN
US10	0.0634	0.9386	RETAIN
FS11	0.9811	0.3764	RETAIN
FS12	0.2055	0.8144	RETAIN
FS13	2.2749	0.1050	RETAIN
AS14	0.6387	0.5289	RETAIN
AS15	0.1941	0.8237	RETAIN
AS16	0.8853	0.4136	RETAIN
AS17	0.0031	0.9969	RETAIN
AS18	0.3427	0.7102	RETAIN
AS19	0.5279	0.5905	RETAIN
AS20	1.0245	0.3606	RETAIN
AS21	0.2324	0.7928	RETAIN



## CHAPTER V: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

### Summary

The basic problems for this study were to identify the current practices of two-year colleges in the area of instructor performance appraisal; to classify the models of evaluation according to the McGreal (1983) taxonomy; to establish the efficacy of each model as it relates to the twenty-one personnel evaluation standards recently identified by the Joint Committee (1988); and to develop a profile of the "typical" supervisor conducting instructor performance appraisal. To accomplish these goals, two questionnaires were used to collect data from two-year college leaders, instructional supervisors, and other administrators responsible for evaluating instructors. Subjects were members of two-year college organizations and were randomly selected to respond to the questionnaires.

The study was conducted in two phases. During the first phase of the study, a 10 per cent random sample of 2250 leaders of two-year colleges were asked to respond to the first questionnaire. Over 83 per cent (187) of those surveyed responded. The responses requested were the names and titles of the supervisors responsible for instructor performance appraisal. In the second phase of the study, a 40 per cent sample of those identified as instructor performance evaluators by those surveyed with the first questionnaire, were asked to respond to a lengthy data gathering questionnaire that was designed to collect information that would provide answers to the problems identified in this study. Over 71 per cent (256) of those surveyed responded. All the data collected were treated statistically.

The first problem, to identify the current practices of two-year colleges in the area of instructor performance appraisal, required a comprehensive questionnaire. The questionnaire obligated the respondent to answer 80 questions. The responses were first tabulated and classified by practice and then compared by type and size of two-year college using analysis of variance as the statistical procedure of comparison.

The second problem, to classify the models of evaluation according to the McGreal (1983) taxonomy, required the respondent to select the model of instructor evaluation that best matched their current evaluation process. The respondents were asked to make their selections from the description of the models that was provided within the questionnaire. Immediately following the classification section were questions about the number of formal observations and the use of pre-observation and post-observation conferences. Statistical analysis involved producing frequency histograms, conducting the chi square ( $\chi^2$ ) test of independence, and determining contingency coefficients of correlation.

The third problem, to establish the efficacy of each model as it relates to the twenty-one personnel evaluation standards recently identified by the Joint Committee (1988), required the respondent to rate the efficacy of their particular model using a four point Likert type scale. The standard was stated and the respondent was asked to strongly agree, agree, disagree, or strongly disagree that the evaluation model in their institution did precisely what the standard cited. The statistical treatment of these data included determining means, medians, standard deviations, and variances as well as making additional comparisons by use of multiple analyses of variance.

The fourth problem, to develop a profile of the "typical" supervisor conducting instructor performance appraisal, required the researcher to compile demographic and other pertinent data. The respondents supplied the needed data as a part of their responses to the primary data gathering questionnaire. The average or most frequently occurring response, whichever was most appropriate, was used to produce the profile of a "typical" two-year college instructor performance evaluator.

### Findings

Within the limitations of the study the following findings were established. The research questions and hypotheses will be answered in chronological order.

#### The first research question:

Who supervises whom and what? What is the span of control? What are each supervisors duties?

1. The job title most frequently identified as the instructional leader responsible for instructor evaluation was Dean of Instruction or Academic Affairs.

2. The Dean of Instruction or Academic Affairs was responsible for the evaluation of an average of 62.5 instructors.

3. The 62.5 instructors evaluated by the Dean of Instruction or Academic Affairs include an average of 53.8 classroom instructors, 2.5 clinical instructors, 3.5 laboratory instructors and 2.7 shop instructors.

4. The Dean of Instruction or Academic Affairs also is responsible for the evaluation of at least 7 non-instructors; i.e., 1.3 non-teaching administrators,

1.2 teaching administrators, 1.0 technicians, 2.8 support staff personnel, and 1.1 others.

5. The span of control for the Dean of Instruction or Academic Affairs was an average of 70 people.

6. Overall the instructional leaders in two-year colleges are responsible for the evaluation of an average of 44.2 instructors and 7.5 non-instructors.

a. Instructional leaders in colleges with more than 3500 F.T.E.

have an average of 50.4 instructors to evaluate.

b. Instructional leaders in colleges with between 1501 and

3500 F.T.E. have an average of 45.3 instructors to evaluate.

c. Instructional leaders in colleges with less than 1501 F.T.E.

have an average of 39.4 instructors to evaluate.

7. The instructional leaders in two-year colleges spend 8.9 per cent of their time evaluating instructors.

8. Instructional leaders reported to have a number of other duties:

a. 96.9 per cent support the improvement of instruction,

b. 94.1 per cent supervise the curriculum,

c. 94.1 per cent promote professional growth,

d. 92.2 per cent support instructors,

e. 91.4 per cent formally evaluate instructors,

f. 84.8 per cent supervise support staff,

g. 84.8 per cent supervise instructors,

h. 84.4 per cent serve in the community,

i. 83.2 per cent assist students,

j. 81.6 per cent formally evaluate support staff,

- k. 79.3 per cent maintain public relations,
- l. 72.7 per cent provide orderly environment,
- m. 69.9 per cent emphasize student achievement,
- n. 56.6 per cent monitor student progress,
- o. 48.4 per cent assist in economic development for their area,
- p. 44.9 per cent supervise other administrators,
- q. 42.2 per cent formally evaluate other administrators, and
- r. 35.2 per cent maintain physical facilities.

The second research question:

What are the models of evaluation being utilized in the two-year college?

9. The clinical supervision model was identified as the most commonly used model of evaluation. Of those responding:

- a. 39.8 per cent identified the clinical supervision model as the one employed in their institution,
- b. 27.7 per cent identified the common law model as the one employed in their institution,
- c. 21.1 per cent identified the goal setting model as the one employed in their institution,
- d. 7.4 per cent identified the artistic or naturalistic model (the belief that teaching is an art) as the one employed in their institution, and
- e. 3.1 per cent identified the product model as the one employed in their institution.

10. The clinical supervision model is used more often in community colleges, vocational technical colleges, and community technical colleges. The common law model is the most prevalent in junior colleges.

11. The models used do not differ by institutional size. The clinical supervision model is the most prevalent model in all sizes of two-colleges.

12. Only 28.5 per cent of the respondents indicated that a pre-observation conference is held as a part of instructor evaluation. Post-observation conferences were reported to be held in 60.5 per cent of the instructor evaluation processes.

13. Even though 39.8 per cent of two-year college instructional leaders claim to be using the clinical supervision model, only 26.6 per cent use the elements of a pre-observation and a post-observation conference.

14. The average number of formal evaluations conducted before a summative evaluation report is written was 1.3 per instructor evaluated.

15. Eighteen per cent of the two-year colleges, represented by the respondents to this survey, are utilizing a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.

16. The smaller the two-year college the more likely a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program is in use. Of those responding:

- a. 26.9 per cent of the colleges with an enrollment of less than 1501 F.T.E. use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.

- b. 16.5 per cent of the colleges with an enrollment between 1501 and 3500 F.T.E. use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.
- c. 6.1 per cent of the colleges with an enrollment of more than 3500 F.T.E. use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.

17. Technical colleges are the most likely to use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program. Of those responding:

- a. 24.0 per cent of the Technical colleges use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.
- b. 22.2 per cent of the Vocational Technical colleges use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.
- c. 18.2 per cent of the Community colleges use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.
- d. 11.1 per cent of the Community Technical colleges use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.
- e. 9.4 per cent of the Junior colleges use a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program.

The third research question:

What are the elements of evaluation being used in instructor performance evaluation?

18. Thirty-nine and one-half per cent of the respondents indicated that the state in which their two-year college is located mandates instructor evaluation.

19. Over 95 per cent of the colleges represented by the respondents to this survey require instructor evaluation.

20. The same evaluation instrument is used for all instructors in 84 per cent of the two-year colleges.

21. Peer feedback was an element of 42.2 per cent two-year college evaluation systems.

22. Peer feedback is more likely to be an element of evaluation in two-year colleges with enrollments over 3500 F.T.E. Of those responding:

- a. 53.0 per cent of two-year colleges with enrollments over 3500 F.T.E. use peer feedback as a part of instructor evaluation.
- b. 38.8 per cent of two-year colleges with enrollments between 1501 and 3500 F.T.E. use peer feedback as a part of instructor evaluation.
- c. 37.5 per cent of two-year colleges with enrollments under 1501 F.T.E. use peer feedback as a part of instructor evaluation.



23. Peer feedback is more likely to be a part of evaluation in Technical, Junior, and Community colleges than it is in Vocational Technical and Community Technical colleges. Of those responding:

- a. 56.0 per cent of the Technical colleges use peer feedback as a part of instructor evaluation.
- b. 50.0 per cent of the Junior colleges use peer feedback as a part of instructor evaluation.
- c. 41.8 per cent of the Community colleges use peer feedback as a part of instructor evaluation.
- d. 22.2 per cent of the Vocational Technical colleges use peer feedback as a part of instructor evaluation.
- e. 11.1 per cent of the Community Technical colleges use peer feedback as a part of instructor evaluation.

24. Very few two-year colleges use an academic council as a part of instructor evaluation. Overall, only 9.8 per cent use an academic council as a part of instructor evaluation.

25. A large percentage (91.4) of the respondents reported that student feedback is employed as an element of the instructor performance appraisal in their two-year college.

26. Student feedback is most likely to be an element of evaluation in two-year colleges with enrollments over 3500 F.T.E. Of those responding:

- a. 93.9 per cent of two-year colleges with enrollments over 3500 F.T.E. use student feedback as a part of instructor evaluation.

- b. 93.3 per cent of two-year colleges with enrollments under 1501 F.T.E. use student feedback as a part of instructor evaluation.
- c. 88.2 per cent of two-year colleges with enrollments between 1501 and 3500 F.T.E. use student feedback as a part of instructor evaluation.

27. Student feedback is more likely to be a part of evaluation in Technical, Junior, and Community colleges than it is in Vocational Technical and Community Technical colleges. Of those responding:

- a. 96.0 per cent of the Technical colleges use student feedback as a part of instructor evaluation.
- b. 93.8 per cent of the Junior colleges use student feedback as a part of instructor evaluation.
- c. 91.8 per cent of the Community colleges use student feedback as a part of instructor evaluation.
- d. 88.9 per cent of the Vocational Technical colleges use student feedback as a part of instructor evaluation.
- e. 77.8 per cent of the Community Technical colleges use student feedback as a part of instructor evaluation.

28. Student achievement is an element of instructor evaluation in 21.9 per cent of the colleges represented by the respondents to this survey. The smaller the two-year college the more likely student achievement is an element of instructor evaluation. None of the respondents from Community Technical colleges indicated that student achievement is an element of instructor evaluation in their colleges.

29. Only 8.2 per cent of the key participants in two-year college instructor evaluation indicated that the use of outside evaluators was an element of the evaluation system in their college.

The fourth research question:

What are the educational and technical backgrounds of those doing the supervision of instructors?

30. A large majority (70.3) of the respondents to this study indicate that they have previous training in pedagogy. Evaluators are more likely to have completed an teacher preparation program in Community Technical, Vocational Technical, and Technical colleges. Of those responding:

- a. 100.0 per cent of the instructor evaluators in Community Technical colleges have completed a teacher preparation program.
- b. 88.9 per cent of the instructor evaluators in Vocational Technical colleges have completed a teacher preparation program.
- c. 88.0 per cent of the instructor evaluators in Technical colleges have completed a teacher preparation program.
- d. 65.9 per cent of the instructor evaluators in Community colleges have completed a teacher preparation program.
- e. 65.6 per cent of the instructor evaluators in Junior colleges have completed a teacher preparation program.

31. Over 93 per cent of the instructor evaluators reported their highest earned degree to be a master's degree or higher.

32. The percentage of those with earned degrees beyond the master's degree is considerably smaller in the Vocational Technical colleges than in the other types of colleges. Of those responding:

- a. 11.1 per cent of the instructor evaluators in Vocational Technical colleges hold an earned degree beyond the master's level.
- b. 33.3 per cent of the instructor evaluators in Community Technical colleges hold an earned degree beyond the master's level.
- c. 36.0 per cent of the instructor evaluators in Technical colleges hold an earned degree beyond the master's level.
- d. 40.1 per cent of the instructor evaluators in Community colleges hold an earned degree beyond the master's level.
- e. 40.6 per cent of the instructor evaluators in Junior colleges hold an earned degree beyond the master's level.

The fifth research question:

What training does the evaluator have in evaluation?

33. Over 41 per cent of the respondents indicated they have never had any training as an evaluator. On the other hand, over 8 per cent of the respondents have had more than 75 contact hours of training in evaluation. Over two-thirds of those evaluating instructors in two-year colleges have less than 25 contact hours of training in evaluation.

34. A comparison, by college enrollment size, of the number of hours of evaluator training reveal that there are no patterns from which to predict the characteristics of a particular size of two-year college.

35. The Community Technical college instructor evaluators appear to have more training in evaluation than instructor evaluators in other types of two-year colleges. The Community college instructor evaluators appear to have less training in evaluation than instructor evaluators in other types of two-year colleges.

36. Less than 4 per cent of the respondents are certified as instructor evaluators. A further sorting of those certified revealed no pattern from state to state.

37. All but 17 of the 256 respondents indicated at least some need for additional training in instructor evaluation. On a scale from 0 to 9, the average perceived need is 5.02, the median and the mode are both 5 and the standard deviation is 2.65. The mid-point of the scale is at 4.5 and almost two-thirds of the respondents indicate a perceived need above that level.

38. The smaller the two-year college the higher the perceived need for additional training in instructor evaluation. This pattern is in contrast to the pattern of the average number of instructors evaluated in various sizes two-year colleges where the instructor evaluators in larger colleges have significantly larger average numbers of instructors to evaluate.

39. The perceived need for additional training in instructor evaluation is the highest among those from Community Technical colleges and Technical colleges. The lowest perceived need for additional training in instructor evaluation is among those from Community colleges and Junior colleges.

The sixth research question.

Is the frequency of occurrence of the evaluation models in use dependent on the type of the two-year college? The first hypothesis: The distribution of models in use among the two-year colleges is independent of the five types of two-year colleges.

40. There is no difference between the types of two-year colleges regarding their use of different models of evaluation. Hence, the distribution of models in use among the two-year colleges is independent of the five types of two-year colleges.

The seventh research question.

Is the frequency of occurrence of the evaluation models in use dependent on the size of the two-year college? The second hypothesis: The distribution of models in use among the two-year colleges is independent of the size of two-year colleges.

41. There is no difference between the sizes of two-year colleges regarding their use of different models of evaluation. Hence, the distribution of models in use among the two-year colleges is independent of the size of two-year colleges.

The eighth research question.

Is there a relationship between the model in use and the type or size of two-year college? The third hypothesis: There is no relationship between the model in use and the type or size of two-year college.

42. There is little or no relationship between the model in use and the type or size of two-year college. When considering the data relative to the relationship between the type of two-year colleges and the evaluation model in use a contingency coefficient of 0.23988 is computed. Thus the magnitude of the relationship between the model in use and the type of two-year college is interpreted as being low. When considering the data relative to the relationship between the size of two-year colleges and the evaluation model in use a contingency coefficient of 0.15873 is computed. Thus the magnitude of the relationship between the model in use and the size of two-year college is interpreted as being very low.

The ninth research question.

What are the efficacy ratings (the power or capacity to produce the desired effect) for each personnel standard?

43. The efficacy ratings for each of the 21 personnel standards advanced by the Joint Committee (1988) were included on the primary data gathering questionnaire. The respondents were asked to rate the quality of the model or process their institution uses on each of the standards. The responses were converted to numeric values ranging from a high of four (4) to a low of one (1). Responses of "not applicable" or those missing responses were not included in the data tabulation. The highest efficacy rating was for Standard P4: Access to Personnel

Evaluation Reports. The lowest efficacy rating is for Standard FS 13: Fiscal Viability. A listing of the responses follows:

- a. 3.601 Standard P4: Access to Personnel Evaluation Reports -- Access to reports of personnel evaluation should be limited to those individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured.
- b. 3.475 Standard P5: Interactions with Evaluatees -- The evaluation should address evaluatees in a professional, considerate, and courteous manner, so that their self-esteem, motivation, professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged.
- c. 3.270 Standard A6: Systematic Data Control -- The information used in the evaluation should be kept secure, and carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.
- d. 3.262 Standard U1: Constructive Orientation -- Evaluations should be constructive, so that they help institutions to develop human resources and encourage and assist those evaluated to provide excellent service.
- e. 3.188 Standard P2: Formal Evaluation Guidelines -- Guidelines for personnel evaluations should be recorded in statements of policy, negotiated agreements, and/or personnel evaluation manuals, so that evaluations are consistent, equitable, and in accordance with pertinent laws and ethical codes.



- f. 3.185 Standard P1: Service Orientation -- Evaluations of educators should promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society are met.
- g. 3.174 Standard U2: Defined Uses -- The users and the intended uses of a personnel evaluation should be identified, so that the evaluation can address appropriate questions.
- h. 3.123 Standard U3: Evaluator Credibility -- The evaluation system should be managed and executed by persons with the necessary qualifications, skills, and authority, and evaluators should conduct themselves professionally, so that evaluation reports are respected and used.
- i. 3.090 Standard P3: Conflict of Interest -- Conflicts of interest should be identified and dealt with openly and honestly, so that they do not compromise the evaluation process and results.
- j. 3.069 Standard F1: Practical Procedures -- Personnel evaluation procedures should be planned and conducted so that they produce needed information while minimizing disruption and cost.
- k. 3.053 Standard U5: Follow-Up and Impact -- Evaluations should be followed up, so that users and evaluatees are aided to understand the results and take appropriate actions.
- l. 3.046 Standard A3: Documentation of Procedures -- The evaluation procedures actually followed should be documented, so that the

evaluatees and other users can assess the actual, in relation to intended, procedures.

- m. 2.972 Standard F2: Political Viability -- The personnel evaluation system should be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work.
- n. 2.946 Standard A1: Defined Role -- The role, responsibilities, performance objectives, and needed qualifications of the evaluatee should be clearly defined, so that the evaluator can determine valid assessment data.
- o. 2.941 Standard U4: Functional Reporting -- Reports should be clear, timely, accurate, and germane, so that they are of practical value to the evaluatee and other appropriate audiences.
- p. 2.931 Standard A8: Monitoring Evaluation Systems -- The personnel evaluation system should be reviewed periodically and systematically, so that appropriate revisions can be made.
- q. 2.914 Standard A2: Work Environment -- The context in which the evaluatee works should be identified, described, and recorded, so that environmental influences and constraints on performance can be considered in the evaluation.
- r. 2.911 Standard A4: Valid Measurement -- The measurement procedures should be chosen or developed and implemented on the basis of the described role and the intended use, so that the inferences concerning the evaluatee are valid and accurate.

- s. 2.844 Standard A7: Bias Control -- The evaluation process should provide safeguards against bias, so that the evaluatee's qualifications or performance are assessed fairly.
- t. 2.741 Standard A5: Reliable Measurement -- Measurement procedures should be chosen or developed to assure reliability, so that the information obtained will provide consistent indications of the performance of the evaluatee.
- u. 2.733 Standard F3: Fiscal Viability -- Adequate time and resources should be provided for personnel evaluation activities, so that evaluation plans can be effectively and efficiently implemented.

44. The Joint Committee (1988) subdivided the standards into four categories that correspond to four basic attributes of sound evaluation: propriety, utility, feasibility, and accuracy. Basically, the standards require that evaluations be proper, useful, feasible, and accurate. The following are the average efficacy ratings within each of the four categories.

- a. 3.308 Propriety Standards
- b. 3.111 Utility Standards
- c. 3.002 Accuracy Standards
- d. 2.925 Feasibility Standards

The tenth research question.

Are differences in the efficacy ratings (the power or capacity to produce the desired effect) on each personnel standard attributable to the model in use, or to the type or size of two-year college? The fourth hypothesis: There is no

significant difference in the efficacy rating for each personnel standard between the evaluation models in use.

45. The clinical supervision model has a significantly higher efficacy rating than the goal setting model at the 0.05 level for the FUNCTIONAL REPORTING (US9) personnel standard.

46. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the SERVICE ORIENTATION (PS1) personnel standard.

47. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the INTERACTIONS WITH EMPLOYEES (PS5) personnel standard.

48. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the CONSTRUCTION ORIENTATION (US6) personnel standard.

49. The clinical supervision model has a significantly higher efficacy rating than either the goal setting model or the common law model at the 0.05 level for the PRACTICAL PROCEDURES (FS11) personnel standard.

50. The clinical supervision model has a significantly higher efficacy rating than either the artistic-naturalistic model or the goal setting model at the 0.05 level for the POLITICAL VIABILITY (FS12) personnel standard.

51. Both the clinical supervision model and the product model have a significantly higher efficacy rating than the artistic-naturalistic model at the 0.05 level for the SYSTEMATIC DATA CONTROL (AS19) personnel standard.

The fifth hypothesis: There is no significant difference in the efficacy rating for each personnel standard between the types of two-year colleges.

52. The fifth null hypothesis is retained, there is no significant difference in the efficacy rating for each personnel standard between the types of two-year colleges.

The sixth hypothesis: There is no significant difference in the efficacy rating for each personnel standard between the sizes of two-year colleges.

53. The sixth null hypothesis is retained, there is no significant difference in the efficacy rating for each personnel standard between the sizes of two-year colleges.

### Conclusions

The following conclusions were drawn from the results of this study.

The first research question:

Who supervises whom and what? What is the span of control? What are each supervisors duties?

1. The appropriate instructor performance evaluator for most two-year colleges is the Dean of Instruction or Academic Affairs. His or her span of control is normally too large. In addition to spending almost nine (9) per cent of their time performing instructor evaluation for an average of over 62 instructors, these instructional leaders have a wide range of additional duties.

The second research question:

What are the models of evaluation being utilized in the two-year college?

2. The clinical supervision model is the most commonly used model of evaluation in two-year colleges.

3. The clinical supervision model is the most prevalent model in all sizes of two-colleges.

4. The clinical supervision model is the model of choice in community colleges, vocational technical colleges, and community technical colleges. The common law model is the most prevalent in junior colleges.

5. Even though a great number of two-year college instructional leaders claim to be using the clinical supervision model, at least a third of those claiming to use the model are not. The clinical supervision model has as elements a pre-observation and a post-observation conference and only 26.6 per cent of the respondents indicated they did both.

6. Most instructor evaluators only conduct one formal evaluation before a summative evaluation report is written.

7. Teacher evaluation is not being augmented by utilizing a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program in most two-year colleges.

8. Smaller two-year colleges and Technical colleges are more likely a 'pay for performance', 'merit pay', 'career ladder', or other type of incentive program is in use.

The third research question:

What are the elements of evaluation being used in instructor performance evaluation?

9. Only a few two-year colleges had instructor performance appraisal processes that appeared to represent a well-developed system in which external and internal environmental demands were balanced to facilitate the attainment of both improvement and accountability.

10. Differences in instructor performance appraisal; the essential elements in use, and current practices (process design) were substantial.

11. The process of instructor performance appraisal is an under conceptualized and underdeveloped activity.

12. Leaders of two-year colleges value instructor performance appraisal because almost all require instructor evaluation even though it is mandated in less than half of the states.

13. A large majority of the two-year colleges use the same evaluation instrument for all instructors.

14. Student feedback is the most frequently used element in two-year college evaluation systems.

15. Peer feedback is an element of two-year college evaluation systems in almost half the institutions. The larger the two-year college the more likely peer feedback is an element of evaluation.

16. Student achievement, academic councils, and outside evaluators are seldom used elements of two-year college evaluation systems.

The fourth research question:

What are the educational and technical backgrounds of those doing the supervision of instructors?

17. A master's degree and previous training in pedagogy is likely to be minimum requirements of instructor evaluators in two-year colleges. However, certification as an instructor evaluator or evaluator training is not a requirement to be an instructor evaluator in two-year colleges.

18. Most instructor evaluators in two-year colleges have completed a teacher preparation program and have at least a master's degree.

19. Vocational Technical colleges place less emphasis on their instructor evaluators having earned degrees beyond the master's degree.

20. Junior colleges and Community colleges place more emphasis on their instructor evaluators having earned degrees beyond the master's degree.

The fifth research question:

What training does the evaluator have in evaluation?

21. Two-year college instructor evaluators are poorly trained in evaluator training. Hence, there is a need for them to obtain additional training in evaluation. Correspondingly, instructor evaluators believe they need more training in evaluation.

22. The Community Technical college instructor evaluators have the most training in evaluation than instructor evaluators while the Community college instructor evaluators have least training.

23. Instructor evaluators from smaller two-year colleges and Community Technical colleges have the highest perceived need for additional training in instructor evaluation.



The sixth research question.

Is the frequency of occurrence of the evaluation models in use dependent on the type of the two-year college? The first hypothesis: The distribution of models in use among the two-year colleges is independent of the five types of two-year colleges.

24. The type of two-year college cannot be used to predict the model of evaluation likely to be in use.

The seventh research question.

Is the frequency of occurrence of the evaluation models in use dependent on the size of the two-year college? The second hypothesis: The distribution of models in use among the two-year colleges is independent of the size of two-year colleges.

25. The use of different models of evaluation cannot be predicted by the size of the two-year college.

The eighth research question.

Is there a relationship between the model in use and the type or size of two-year college? The third hypothesis: There is no relationship between the model in use and the type or size of two-year college.

26. There is little or no relationship between the model in use and the type or size of two-year college.

The ninth research question.

What are the efficacy ratings (the power or capacity to produce the desired effect) for each personnel standard?

27. The Personnel Standards can be used as criteria against which to evaluate alternative models for instructor performance appraisal in two-year colleges.

28. The Personnel Standards can be used in assessing a current instructor performance appraisal system in a two-year college.

29. The Personnel Standards are equally useful for all types of two-year colleges.

30. The Personnel Standards are equally useful for all sizes of two-year colleges.

The tenth research question.

Are differences in the efficacy ratings (the power or capacity to produce the desired effect) on each personnel standard attributable to the model in use, or to the type or size of two-year college? The fourth hypothesis: There is no significant difference in the efficacy rating for each personnel standard between the evaluation models in use.

31. Efficacy ratings (the power or capacity to produce the desired effect) of the Personnel Standards were higher when clinical supervision was identified as the model of instructor performance appraisal.

32. Evaluators felt that they were more successful in creating reports that were clear, timely, accurate, and germane, so that they were of practical value to the evaluatee and other appropriate audiences when using the clinical supervision model.

33. Evaluations of educators were perceived to promote sound education principles, assist in the fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society were met when the clinical supervision model was used instead of the common law or goal setting models.

34. The evaluation appeared more likely to address evaluatees in a professional, considerate, and courteous manner, so that their self-esteem, motivation, professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged when the clinical supervision model is employed.

35. When the clinical supervision model is used evaluations were perceived as being constructive, so that they help institutions to develop human resources and encourage and assist those evaluated to provide excellent service.

36. Personnel evaluation procedures appear to be planned and conducted so that they produce needed information while minimizing disruption and cost when the clinical supervision model is used.

37. The personnel evaluation system can be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work if clinical supervision is the selected model of evaluation.

38. It seems that when either the clinical supervision model or the product model of evaluation is used the information used in the evaluation can be kept secure, and carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.

The fifth hypothesis: There is no significant difference in the efficacy rating for each personnel standard between the types of two-year colleges.

39. Differences in the efficacy ratings (the power or capacity to produce the desired effect) on each personnel standard are not attributable to the type of two-year college.

The sixth hypothesis: There is no significant difference in the efficacy rating for each personnel standard between the sizes of two-year colleges.

40. Differences in the efficacy ratings (the power or capacity to produce the desired effect) on each personnel standard are not attributable to the size of two-year college.

### Limitations

The limitations of this study were:

1. The statistical treatment of the data was limited to answering only the research questions that were formulated a priori.
2. Participation in the study was voluntary, therefore those who chose to respond may have demonstrated a more positive disposition toward

educational research, or toward instructor performance appraisal, than subjects who did not respond.

3. A 40 per cent random sample does not guarantee that different results would not have been obtained if the entire population had been surveyed.

4. Not every state was represented in the data reported.

5. The primary data gathering questionnaire was quite complex and required at least 80 responses.

### Discussion

The results of this study support previous recommendations of related research which strongly suggests the need for improvement in two-year college instructor evaluation. For example, one can view with alarm the fact that most of those evaluating instructor performance are untrained and even though they claim to use a clinical supervision model of evaluation they don't consistently use the basic processes of the model. The extent to which the practices of instructor performance appraisal are being carried out further indicate an alarming need for training.

A number of similarities were discovered by this study and a set of practices common to the majority of two-year colleges could be derived from the data collected. The major approaches being used in instructor evaluation are characterized by a close relationship between the instructor and the supervisor with emphasis on collegial rather than authoritarian orientation. However, evaluation is synonymous with observation and the major emphasis is on summative evaluation; standardized criteria; and comparative judgements. This process usually takes its principle data from the classroom and is designed

to improve the instructor's performance. Therefore, the current practices of instructor performance appraisal in two-year colleges can be categorized as a combination of McGreal's (1983) Common Law Model and Clinical Supervision Model.

The findings of this study can be compared and contrasted to the current research base by considering the conclusions about teacher evaluation published by the Rand Corporation (Wise et al., 1984). Their report stated that the differences in teacher evaluation process design were substantial. This study concludes that differences in the essential elements in use and current practices (process design) of instructor performance appraisal were substantial. Additionally, the results of this study support their conclusions that only a few educational institutions (in this case two-year colleges) had instructor performance appraisal processes that appeared to represent a well-developed system in which external and internal environmental demands were balanced to facilitate the attainment of both improvement and accountability and that the process of instructor performance appraisal is an under conceptualized and underdeveloped activity. However, the Rand Corporation (Wise et al., 1984) reported that there is little consensus about what design process results in successful teacher evaluation while this study concludes that efficacy ratings (the power or capacity to produce the desired effect) of the Personnel Standards were higher when clinical supervision was identified as the model of instructor performance appraisal.

The norms in important procedures are the use of: a single instrument to evaluate most of the instructors; student feedback as a source of evaluation data; and peer feedback for evaluation data in almost half the institutions. On the

other hand; student achievement, academic councils, and outside evaluators are seldom used. Additionally, the practice of observing instructors in the teaching act is a norm, as is the practice of holding post-observation conferences with those instructors. Finally, the current norm is to use a modified clinical supervision approach that does not usually include a pre-observation conference. Figure 46 represents the state of current practice in instructor performance evaluation in two-year colleges.

It appears the span of control, which was in some cases in excess of 100 instructors, is far too large to conduct quality, comprehensive instructor evaluations. Perhaps a checklist or other cursory measure could be accomplished for these large numbers but nothing more. A reasonable span of control for each instructor evaluator is between 10 and 15 people. It is feasible that as many as 20 people could be evaluated properly each year if they were clustered in the same discipline.

The evaluation should include the immediate administrative supervisor and other administrators charged with instructional responsibilities. Two-year colleges need to review carefully the role of a department chair-person who may be voted in by faculty members or are included in the faculty senate or union. In short, department chair-persons need to be clearly identified as administrators if they are to be part of an administrative evaluation team.

Summative evaluation responsibilities should rest in the hands of the professional administrators. Students and faculty are not charged with the firing and/or dismissal responsibilities in two-year colleges. They are not trained or experienced in classroom evaluation techniques and it is unlikely they will establish patterns of defects and deficiencies, establish plans for improvement,

follow-up with face-to-face evaluation, and place faculty on formal notices to remedy their defects and deficiencies.

## The Current Practices of Instructor Evaluation in Two-Year Colleges

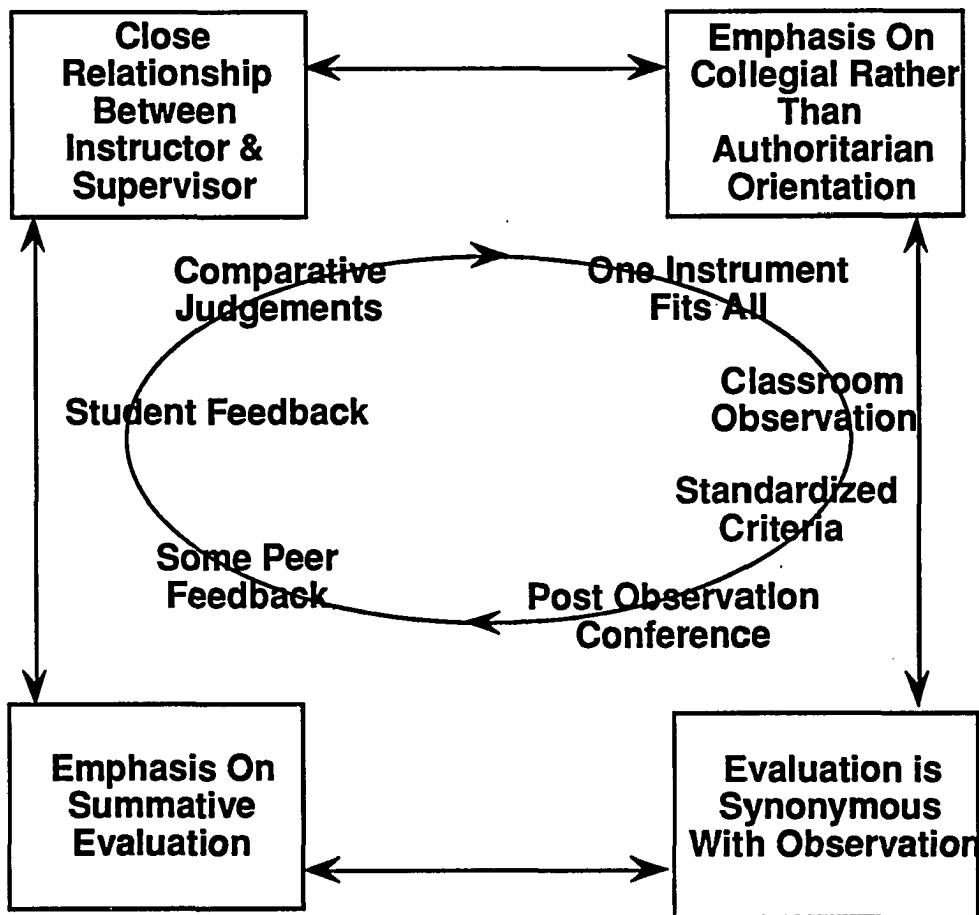


FIGURE 46. The current practices of of instructor performance evaluation in two-year colleges



It will greatly enhance the cause of excellence in education if evaluation policies and procedures are reviewed and strengthened by Governing Boards and top-level administrators. Such a review must include putting a proper weight on the elements as well as carefully selecting the type of elements to be included in evaluation. It should consider the impact that both student and peer evaluation have on the evaluation system. The research on student, self, and peer evaluation show all three systems to be some what weak and may be unreliable as deterrents to poor instruction. However, the students in two-year colleges are older and may be more objective when providing feedback about instructor performance.

#### Recommendations for Practice

The results of this study point to several suggestions for practicing two-year college instructor evaluators.

1. Whenever possible a clinical supervision evaluation model should be used.
2. Both pre-observation and post-observation conferences should be conducted in conjunction with each instructor observation.
3. The use of student feedback as an element of the evaluation system is recommended.
4. The involvement of peers in the evaluation system can help offset personal bias and provide valuable data input.
5. Obtain a sufficient amount of quality training in instructor evaluation to assure self-confidence, inter-rater reliability, and valid reports.

6. Evaluate instructors on research based, critical instructional behaviors.

7. Use the Personnel Standards as criteria against which to evaluate alternative models for instructor evaluation.

8. Use the Personnel Standards in developing an evaluation system or assessing a current system.

### Recommendations for Research

The findings of this study suggest further research. In each suggested study the sample size should be as large as possible.

1. Research should be conducted to identify critical instructional behaviors of two-year college instructors.

2. A research study needs to develop valid and reliable items of evaluation for the behaviors identified as critical instructional behaviors which can discriminate between levels of instructor performance.

3. A research study should be undertaken to examine instrumentation presently being used and to develop prototype instruments to fit the various categories of two-year college instructors.

4. Conduct a study to identify and examine environmental factors influencing instructor performance appraisal. Determine to what extent instructor professional organizations are a key element in developing, implementing and utilizing an evaluation system. Additionally, this study could research the extent that school climate influences the success of the instructor performance appraisal system.

5. Conduct a task analysis of instructor performance appraisal training and develop an ideal training design, materials, and methodology. Determine the amount of contact hours that are needed to most effectively train instructor evaluators.

6. Other research should be conducted which would precisely study the extent that ethics influence instructor performance appraisal. Such a study should result in documents in which the pros and cons of various ethical questions relative to educational evaluation have been examined and expounded upon.

7. Conduct studies to determine if technology, such as the computer, can be used to enhance instructor performance appraisal system utilization.

8. A research study needs to determine what data sources, artifacts, and other items can be used to enhance the evaluation of two-year college instructors. The study should review various examples of evaluation instruments, policies, faculty handbooks, and/or narrative descriptions of the evaluation process and procedures.

9. Other research should be conducted which would precisely study the relationships between evaluation models and instructor performance.

10. As a remedy to the limitations of this study, another study should be conducted using face-to-face interviews with instructor performance evaluators. The sample should be larger than 10 per cent and the written policies, procedures and practices should be critically reviewed.

11. The results of this investigation should be verified. Any replication of this study should add a Likert type response mode to a query about how suc-

cessful the evaluator believes his or her overall instructor performance appraisal system to be.

The means of collecting data should be simplified. A questionnaire which includes fewer items about more specific areas would perhaps provide more concise data. Another researcher might also choose to group the sizes or types of colleges differently for data analysis. Additionally, a sample that is stratified to assure even distribution of size and type of college could enhance the results.

## BIBLIOGRAPHY

- Abrami, P. C. "Dimensions of effective college instruction." Review of Higher Education 8 (1985): 211-228.
- Abrami, P. C., Leventhal, L., and Dickens, W.J. "Multidimensionality of student ratings of instruction." Instructional Evaluation 6 (1) (1981): 12-17.
- Abrami, P. C., Leventhal, L., and Perry, R.P. "Educational seduction." Review of Educational Research 52 (1982): 446-464.
- Acheson, K. A., and Gall, M. D. Techniques in the Clinical Supervision of Teachers. New York: Longman, 1980.
- Aleamoni, L. M. "Student ratings of instruction." In Handbook of teacher evaluation, ed J. Millman, 110-145. Beverly Hills: Sage, 1981.
- Allen, T. F. "Identifying behaviours of the master teacher." Ph.D. diss., Iowa State University, 1985.
- Anderson, J. S. "College Administrators: The Shoemaker's Children." Community and Junior College Journal 54 (6) (1984): 20-21.
- Andrews, H. A. Evaluating for Excellence. Stillwater: New Forums Press, Inc., 1985.
- Andrews, H. A., and Knight, J.H. "Administrative Evaluation of Teachers: Resistance and Rationale." NASSP Bulletin 71 (503) (1987): 1-4.
- Andrews, R. L. The Ecology of School Renewal. Chicago: University of Chicago Press, 1987.
- Arreola, R. A. "Establishing Successful Faculty Evaluation and Development Programs." In Evaluating Faculty and Staff: New Directions for Community Colleges, ed. A. Smith, 83-93. San Francisco: Jossey-Bass, 1983.
- Arreola, R.A. "Evaluating the dimensions of teaching." Instructional Evaluation 8 (2) (1986): 4-12.

- Arreola, R. A. "A Faculty Evaluation Model for Community and Junior Colleges." In Techniques for Evaluating and Improving Instruction: New Directions for Teaching and Learning, ed. L. M. Aleamoni, 65-74. San Francisco: Jossey-Bass, 1987.
- Ashbaugh, C.R. and Kasten, K.L. "Should teachers be involved in teacher appraisal?" NASSP Bulletin 71 (503) (1987): 50-53.
- Aubrecht, J.D. "Reliability, validity, and generalizability of student ratings of instruction." (IDEA Paper No. 6). Kansas State University: Center for Faculty Evaluation and Development, 1981. (ERIC Document Reproduction Service No. ED 213 296).
- Beckham, J.C. Legal Aspects of Teacher Evaluation. Topeka: National Organization on Legal Problems of Education, 1981.
- Berk, R.A. "The construction of rating instruments for faculty evaluation." Journal of Higher Education 50 (1979): 650-669.
- Bevan, J. M. "Faculty Evaluation and Institutional Rewards." AAHE Bulletin 33 (1980): 1-8.
- Blackburn, R.T., and Clark, M.J. "An Assessment of faculty performance: Some correlates between administrators, colleagues, students and self-ratings." Sociology of Education 48 (1975): 242-256.
- Blome, A.C. "Developing a teachers evaluation process." NASSP Bulletin 71 (503) (1987): 36-43.
- Bolton, D. L. Selection and Evaluation of Teachers. Berkeley: McCutchan Publishing, 1973.
- Boyer, E. L. "Principals - The Key People in Strengthening, Improving the Teaching Profession." NASSP Bulletin 70 (490) (1986): 26-32.
- Brandt, R. "On teacher evaluation: A conversation with Tom McGreal." Educational Leadership 44 (7) (1987): 20-24.
- Braskamp, L.A., Brandenburg, D.C., and Ory, J.C. Guidebook for evaluating teaching. Beverly Hills, CA: Sage, 1984.

- Braunstien, D., and Bentson, G. "Student and department chairman views of the performance of university professors." Journal of Applied Psychology, 58 (2) (1973): 244-249.
- Brophy, J.E. "Stability of teacher effectiveness." American Educational Research Journal 10 (1973): 245-252.
- Burke, P., and Lind, K.G. "Performance Assessment Techniques for teacher career ladder advancement." NASSP Bulletin 71 (503) (1987): 27-35.
- Buttram, J.L., and Wilson, B.L. "Promising trends in teacher evaluation." Educational Leadership 44 (7) (1987): 4-6.
- Cameron, W. J. "Perceptions of principals, teachers, and school board members to be considered in a teacher evaluation formula based on student growth as for student characteristics and available facilitation and materials." Ph.D. diss., Iowa State University, 1973.
- Capie, W., Anderson, S. J., Johnson, C. E., and Ellett, C. C. Teacher performance assessment instruments: A handbook for interpretation. Atlanta: Georgia State Department of Education, 1979.
- Centra, J. A. "Colleagues as raters of classroom instruction." Journal of Higher Education 46 (1975): 327-337.
- Centra, J.A. How universities evaluate performance: A survey of department heads. GREB No. 75-5 br. Princeton, NJ: Educational Testing Service, 1977.
- Centra, J. A. Determining Faculty Effectiveness. San Francisco: Jossey-Bass, 1979.
- Centra, J. A. and Potter, D. A. "School and teacher effects: An international model." Review of Educational Research 50 (2) (1980): 273-291.
- Centra, J. A. "Formative and Summative Evaluation: Parody or Paradox?" In Techniques for Evaluating and Improving Instruction: New Directions for Teaching and Learning, ed. L. M. Aleamoni, 47-56. San Francisco: Jossey-Bass, 1987.
- Christen, W. L., and Murphy, T. J. "Inservice Training and Peer Evaluation: An Integrated Program for Faculty Development." NASSP Bulletin 71 (503) (1987): 10-10.

- Chronicle Two-Year College Databook: For 1988-89 School Year. Moravia: CHRONICLE GUIDANCE Publications, Inc., 1988.
- Cogen, M. L. Clinical Supervision. Boston: Houghton Mifflin, 1973.
- Cohen, A. M. "Evaluation of Faculty." Community College Review 2(4) (1974): 12-21.
- Cohen, A. M., and Brawer, Florence B. The American Community College. San Francisco: Jossey-Bass Publishers, 1982.
- Cohen, A. M.; Palmer, James C.; and Zwemer, K. Diane. Key Resources on Community Colleges. San Francisco: Jossey-Bass Publishers, 1986.
- Cohen, P.A. "Effectiveness of student-rating feedback for improving college instruction: A meta-analysis." Research in Higher Education 13 (1980): 321-341.
- Cohen, P.A. "Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies." Review of Educational Research 51 (1981): 281-309.
- Cohen, P.A. "An updated and expanded meta-analysis of multisection student rating validity studies." In Student ratings of instruction: A new look at multisection validity studies. R.P. Perry (chair), Symposium conducted at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1986.
- Cohen, P.A., and Mays, E.R. "Using student rating feedback to improve college teaching." California Journal of Teacher Education 8 (4) (1981): 27-36.
- Cohen, P.A., and McKeachie, W.J. "The role of colleagues in the evaluation of college teaching." Improving College and University Teaching 28 (1980): 147-154.
- Conley, D. T. "Critical Attributes of Effective Evaluation Systems." Educational Leadership 44 (7) (1987): 60-64.
- Costa, A. L., Garmston, R. J., and Lambert, L. "Evaluation of Teaching: The Cognitive Development View" In Teacher Evaluation: Six Prescriptions for Success, eds. Sarah J. Stanley and W. James Popham. Alexandria: Association for Supervision and Curriculum Development, 1988.



- Darling-Hammond, L. "Teacher Evaluation: A Proposal." Elementary School Journal 86 (1986): 531-551.
- Darling-Hammond, L. "The Futures of Teaching." Educational Leadership 46 (3) (1988): 4-10.
- Darling-Hammond, L., Wise, A., and Pease, S. "Teacher Evaluation in the Organizational Context: A Review of the Literature." Review of Educational Research 53 (1983): 285-328.
- Darnell, D. F. "An analysis of the costs of administration in teacher evaluation." Ph.D. diss., Iowa State University, 1984.
- Deegan, W. L.; Tillery, Dale; and Associates. Renewing the American Community College. San Francisco: Jossey-Bass Publishers, 1985.
- Demmon-Berger, D. Effective teaching: Observations from Research. Arlington: American Association of School Administrators, 1986.
- Denham, S. A. "Emotional and Prosocial Correlates of Teachers' Ratings of Preschool Social Competence and Behavior Problems." Paper presented at the annual meeting of the Eastern Psychological Association, Rosslyn, VA, April 1987.
- Doyle, K.O., Jr. Evaluating Teaching. Lexington, MA: D.C. Heath and Company, 1983.
- Duke, D. L. and Stiggins, R. J. Teacher Evaluation: Five Keys to Growth Washington, D. C.: NEA Publications, 1986.
- Dunnette, M.D., and Borman, W.C. "Personnel selection and classification systems." Annual Review of Psychology, 30 (1979): 477-525.
- Dzyacky, K. M. "The relationships of teacher perceptions and administrator time on instructional leadership with school learning climate." Ph.D. diss., Iowa State University, 1988.
- Edmonds, R. "Effective Schools for the Urban Poor." Educational Leadership 37 (6) (1979): 5-13.
- Edmonds, R. "Programs of School Improvement: An Overview." Educational Leadership 40 (10) (1982): 4-11.

- Educational Research Service. Evaluating Teacher Performance. Arlington: Educational Research Service, 1979.
- Edwards, S. L. "Development and analysis of a Structured, Data Capturing Technique (SDCT) for classroom observation." Ph.D. diss., Iowa State University, 1985.
- Eisner, E. "An Artistic Approach to Supervision." In Supervision of Teaching: 1982 ASCD Yearbook, ed. Thomas J. Sergiovanni. Alexandria: Association for Supervision and Curriculum Development, 1982.
- Ellett, C. D., Capie, W. and Johnson, C. E. "Assessing teacher performance." Educational Leadership 38 (3) (1980): 219-220.
- Evertson, C. M. and Holley, P.M. "Classroom observation." In Handbook of teacher evaluation, ed. J. Millman, 73-89. Beverly Hills: Sage, 1981.
- Faast, D. A. "An analysis of the effectiveness of training teacher evaluators in specific steps in the process of clinical supervision and teacher performance evaluation." Ph.D. diss., Iowa State University, 1982.
- Feldman, K.A. "Grades and college students' evaluations of their courses and teachers." Research in Higher Education 4 (1976a): 69-111.
- Feldman, K.A. "The superior college teacher from the students' view." Research in Higher Education 5 (1976b): 243-288.
- Feldman, K.A. "Consistency and variability among college students in rating their teachers and courses." Research in Higher Education 6 (1977): 223-274.
- Feldman, K.A. "Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't." Research in Higher Education 9 (1978): 199-242.
- Feldman, K.A. "The significance of circumstances for college students' ratings of their teachers and courses." Research in Higher Education 10 (1979): 149-172.
- Feldman, K.A. "The seniority and instructional experience of college teachers as related to the evaluations they receive from students." Research in Higher Education 18 (1980): 3-124.
- Feldman, K.A. "Class size and college students' evaluations of teachers and courses: A closer look." Research in Higher Education 21 (1984): 45-116.

- Feldman, K.A. "Seniority and Experience of College Teachers as Related to Evaluations They Receive from Students." Research in Higher Education. 18 (1) (1983): 3-124.
- Feldvebel, A. M. "Teacher evaluation: Ingredients of a credible model." Clearing House 53 (9) (1980): 415-420.
- Fleming, R. W. (Chair). To Strengthen Quality in Higher Education. Summary Recommendations of the National Commission on Higher Education Issues. Washington, DC: American Council on Education, 1982.
- Floden, R. E. and Feiman, S. A consumer's guide to teacher development. East Lansing Institute for Research on Teaching, Michigan State University, 1981.
- Floden, P. M. "Development and assessment of a structured data-recording (SDR) technique for classroom observation." Ph.D. diss., Iowa State University, 1987.
- Fontana, L., and Ochoa, A. "What Are the Issues for Teacher Training." NASSP Bulletin 69 (480) (1985): 15-20.
- Freer, M. L. "Clinical Supervision: Training That Works." NASSP Bulletin 71 (503) (1987): 12-17.
- French, R. L., Malo, G. E., and Rakow, E. A. "What We Have Learned from Tennessee's Career Ladder Experience." Educational Leadership 46 (3) (1988): 70-73.
- French-Lazovik, G. "Peer review: Documentary evidence in the evaluation of teaching." In Handbook of teacher evaluation, ed. J. Millman, 73-89. Beverly Hills: Sage, 1981.
- Frudden, S. H. "A comparison of school administrators' ratings of teaching performance utilizing varied instructional materials in the assessment process." Ph.D. diss., Iowa State University, 1980.
- Furtwengler, C. "Lessons from Tennessee's Career Ladder Program." Educational Leadership 44 (7) (1987): 66-69.
- Gage, N.L. "Paradigms for research on teaching," In Handbook for Research on Teaching, ed. N.L. Gage Chicago: Rand McNally, 1963.

- Gallup, G. H. "The eleventh annual Gallup poll of the public's attitudes toward the public schools." Phi Delta Kappan. 60 (1979): 33-45.
- Garawski, R. A. "Successful teacher evaluation is not a myth." NASSP Bulletin 64 (1980): 1-7.
- Garcia, M. W. "The relationship between teachers' perception of the organizational climate and their perceptions of the effectiveness of the evaluation process." Ph.D. diss., Iowa State University, 1980.
- George, P. S. "Performance Management in Education." Educational Leadership 44 (7) (1987): 32-39.
- Gilmore, G.M. "Student ratings as a factor in faculty employment decisions and periodic review." Journal of College and University Law 10 (4) (1984): 557-576.
- Gilmore, G.M., Kane, M.T., and Naccarato, R.W. "The generalizability of student ratings of instruction: Estimates of teacher and course components." Journal of Educational Measurement 15 (1978): 1-73.
- Glass, G. V. "Teacher Effectiveness." In Evaluating Educational Performance, ed. H. Walberg Berkely: .McCutchan, 1974.
- Glatthorn, A., and Holler, R. "Differentiated Teacher Evaluation." Educational Leadership 44 (7) (1987): 56-58.
- Glueck, W. F. Personnel: A Diagnostic Approach. Plano: Business Publications, 1982.
- Goldhammer, R. Clinical Supervision. New York: Holt, Rinehart, and Winston, 1969.
- Goodlad, J. A place called school: Prospects for the future. New York: McGraw-Hill, 1984.
- Griffen, G. A. and Barnes, Susan "Using Research Findings to Change School and Classroom Practices: Results of an Experimental Study." American Educational Research Journal 23 (4) (1986): 572-586.
- Gudridge, B. M. Teacher competency: Problems and solutions. Arlington: American Association of School Administrators, 1980.

- Haefele, D. L. "How to evaluate thee, teacher -- let me count the ways. Phi Delta Kappan 61 (5) (1980): 349-352.
- Harrington, L. D. "The association of administrators learning styles, background differences, and teacher appraisal judgments." Ph.D. diss., Iowa State University, 1984.
- Harris, B. M. "Resolving Old Dilemmas in Diagnostic Evaluation." Educational Leadership 44 (7) (1987): 46-49.
- Harris, B. M. Developmental Teacher Evaluation. Newton, MA: Allyn and Bacon, Inc., 1986.
- Harris, B. M.; McIntyre, K. E.; Littleton, Jr., V. C.; and Long, D. F. Personnel Administration in Education: Leadership for Instructional Improvement. Boston: Allyn and Bacon, Inc., 1979.
- Hawley, R. C. Assessing Teacher Performance. Amherst, MA: Education Research Associates, 1982.
- Heneman, H. G., III. "Comparison of self- and superior ratings of managerial performance." Journal of Applied Psychology 59 (12) (1974): 638-642.
- Hermann, B. A. "Effective Teacher Evaluation: A Quantitative and Qualitative Process." NASSP Bulletin 71 (503) (1987): 23-30.
- Hidlebaugh, E. "A model for developing a teacher performance evaluation system." Ph.D. diss., Iowa State University, 1973.
- Hinkle, D. E., Wiersma, W., and Jurs, S. G. Applied Statistics for the Behavioral Sciences. Boston: Houghton Mifflin Co., 1979.
- Holdzhom, D. "Appraising Teacher Performance in North Carolina." Educational Leadership 44 (7) (1987): 40-44.
- Huddle, G. "Teacher Evaluation - How Important for Effective Schools? Eight Messages from Research." NASSP Bulletin 69 (479) (1985): 58-63.
- Hughes, C., Lease, S., Kiser, C., and Butchee, J. "Relating Teacher Evaluation to School Effectiveness Research." ASCD Annual Conference, San Antonio, 1987.

- Hunter, M. "Teacher Competency: Problem, Theory, and Practice." Journal of the College of Education 15 (2) (1976): 1-15.
- Hunter, M. Mastery Teaching. El Segundo: TIP Publications, 1982.
- Hunter, M. "Effecting a reconciliation between supervision and evaluation - A reply to Popham." Journal for Evaluation of Educational Personnel. Kalamazoo: College of Education, The Evaluation Center, 1988.
- Joint Committee on Standards for Educational Evaluation. Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill, 1981.
- Joint Committee on Standards for Educational Evaluation. The Personnel Evaluation Standards. Newbury Park: SAGE Publications, Inc., 1988.
- Judkins, M. "Identifying discriminating items for the student evaluation of teachers." Ph.D. diss., Iowa State University, 1987.
- Keefe, J. W. and Jenkins, J. M. Instructional Leadership Handbook. Reston: NASSP, 1984.
- King, J.A. "Beyond classroom walls: Indirect measures of teacher competence" In Handbook of Teacher Evaluation, ed. J. Millman. Beverly Hills: Sage, 1981.
- Kirchner, W. K. "Relationships between supervisory and subordinate ratings for technical personnel." Journal of Industrial Psychology, 3 (3) (1965): 57-60.
- Knapp, M. S. Toward the study of teacher evaluation as an organizational process: A review of current research and practice. Menlo Park: Educational and Human Services Research Center, SRI International, 1982.
- Landy, F.J. and Farr, J. L. "Performance Rating." Psychological Bulletin 87 (1) (1980): 72-107.
- Lee, B. A. "Federal court involvement in academic personnel decisions: Impact on peer review." Journal of Higher Education 56 (1985): 38-54.
- Lewis, A. Evaluating educational personnel. Arlington: American Association of School Administrators, 1982.

- Lezotte, L. W. "School Improvement Based on Effective School Research: A Promising Approach for Economically Disadvantaged and Minority Students." Journal for Negro Education 54, 3 (1985): 17-28.
- Lezotte, L. W. and Bancroft, B. A. "Growing Use of the Effective Schools Model for School Improvement." Educational Leadership 43 (3) (1985): 65-69.
- Licklider, B. L. D.. "Examination of a site-based, peer coaching inservice model." Ph.D. diss., Iowa State University, 1986.
- Look, E. E. "Selecting criteria for evaluation of school principals based upon item discrimination power and job situation." Ph.D. diss., Iowa State University, 1983.
- Lovain, T. B. "Grounds for Dismissing Tenured Postsecondary Faculty for Cause." The Journal of College and University Law 10 (3) (1984): 419-433.
- Lueders, W. J. "Selecting performance criteria for the evaluation of public school superintendents based on item discrimination power." Ph.D. diss., Iowa State University, 1987.
- MacSampson©. Version 1.1. Stanford, CA, Stanford University, 1985.
- Manatt, R. The School Improvement Model: Faculty development through evaluation. Ames: Iowa State University, 1987a.
- Manatt, R. "Lessons From a Comprehensive Performance Appraisal Project." Educational Leadership 44 (7) (1987b): 8-14.
- Manatt, R. P. "Teacher Performance Evaluation: A Total Systems Approach" In Teacher Evaluation: Six Prescriptions for Success, eds. Sarah J. Stanley and W. James Popham. Alexandria: Association for Supervision and Curriculum Development, 1988.
- Manatt, R. P., and Stow, S.B. Developing and testing a model for measuring and improving educational outcomes of K-12 schools: Technical report. Ames, IA: School Improvement Model (SIM), College of Education, Iowa State University, 1986.
- Manatt, R. P., Palmer, K. L., and Hidlebaugh, E. "Evaluating Teacher Performance with Improved Rating Scales." NASSP Bulletin 60 (1976): 21-23.

- Mark, S. F. "Faculty Evaluation in Community Colleges." Community/Junior College Quarterly of Research and Practice 6(2) (1982): 167-178.
- Marsh, H. W. "Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility." Journal of Educational Psychology, 76 (1984): 707-754.
- McCall, R. B. Fundamental Statistics for Behavioral Sciences. San Diego: Harcourt Brace Jovanovich, 1986.
- McGreal, T. L. "Evaluation for Enhancing Instruction: Linking Teacher Evaluation and Staff Development" In Teacher Evaluation: Six Prescriptions for Success, eds. Sarah J. Stanley and W. James Popham. Alexandria: Association for Supervision and Curriculum Development, 1988.
- McGreal, T. L. Successful Teacher Evaluation. Alexandria, VA: Association for Supervision and Curriculum Development, 1983.
- McGreal, T. L. "Effective Teacher Evaluation Systems." Educational Leadership 39 (January 1982): 303-305.
- McGreal, T. L. "Helping Teachers Set Goals." Educational Leadership 37 (February 1980): 414-419.
- McKeachie, W. J. "Student ratings of faculty: A reprise." Academe (1979): 384-397.
- McKeachie, W. J. "Can Evaluating Instruction Improve Teaching?" In Techniques for Evaluating and Improving Instruction: New Directions for Teaching and Learning, ed. L. M. Aleamoni, 3-8. San Francisco: Jossey-Bass, 1987.
- McLaughlin, M. W., and Pfeifer, R. S. "Teacher Evaluation: Learning for Improvement and Accountability and Case Studies." Stanford: Stanford Education policy Institute, January 1986.
- McNeil, J., and Popham, W. "The assessment of teacher competence." In Second Handbook of Research on Teaching, ed. R. Travers, 218-244. Chicago: Rand McNally, 1973.
- Medley, D. M. "The effectiveness of teachers." In Research on Teaching eds. P. L. Peterson and H. J. Walberg, 13-18. Berkley: McCutchen, 1979.



- Medley, D. Teacher Competence and Teacher Effectiveness: A Review of Process-Product Research. Washington, D.C. : American Association of Colleges for Teacher Education, 1982.
- Menges, R.J. and Brinko, K.T. Effects of student evaluation feedback: A meta-analysis of higher education. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1986.
- Miller, R., and Miller, K. "Clinical Supervision: History, Practice, Perspective." NASSP Bulletin 71 (503) (1987): 18-21.
- Millman, J., ed. Handbook of Teacher Evaluation. Beverly Hills: SAGE Publications, Inc., 1981.
- Mitchell, J. K. "Improvement-directed analysis of marginal teachers." Ph.D. diss., Iowa State University, 1985.
- Mohr, M. M. "The Annoyance of a Good Example." Educational Leadership 44 (7) (1987): 74-75.
- Mueller, D. W. "Development of performance improvement commitments for secondary school principals." Ph.D. diss., Iowa State University, 1987.
- Murnane, R., and Cohen, D. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." Harvard Education Review 56 (1986): 1-17.
- Murray, H.G. Evaluating university teaching: A review of research. Toronto, Canada: Ontario Confederation of University Faculty Associations, 1980.
- Nance, S. R. "Training school administrators in the writing of quality professional improvement commitments." Ph.D. diss., Iowa State University, 1986.
- Noriega, F. G. "The high gain teacher." Ph.D. diss., Iowa State University, 1987.
- Norusis, M. J. SPSSx© Introductory Statistics Guide. Chicago: SPSS Inc., 1983.
- Parker, J. W., Taylor, E K., Barrett, R. S., and Martens, L. "Rating scale content: Relationship between supervisory- and self-ratings." Personnel psychology 12 (1959): 49-63.

- Peterson, David. "Rating teacher performance to determine career ladder advancement: An analysis of bias and reliability." Ph.D. diss., Iowa State University, 1988.
- Peterson, Donovan. "Legal and Ethical Issues of Teacher Evaluation: A Research Based Approach." Educational Research Quarterly 7 (4) (Winter, 1983): 6-16.
- Peterson, K. and Kauchak, D. Teacher evaluation: Perspectives, practices and promises. Salt Lake City: Center for Educational Practice, University of Utah, 1982.
- Piele, P. K. The Yearbook of School Law, 1983. Topeka, KS: National Organization on Legal Problems of Education, 1983.
- Piele, P. K. The Yearbook of School Law, 1981. Topeka, KS: National Organization on Legal Problems of Education, 1981.
- Piele, P. K. The Yearbook of School Law, 1979. Topeka, KS: National Organization on Legal Problems of Education, 1979.
- Pinckney, R. D. "An analysis of school building administrator functions, building administrator effectiveness, and measures of school effectiveness." Ph.D. diss., Iowa State University, 1982.
- Poole, L.H., and Dellow, D.A. "Evaluation of Full-time Faculty." In Evaluating Faculty and Staff: New Directions for Community Colleges, ed. A. Smith, 19-31. San Francisco: Jossey-Bass, 1983.
- Popham, W. J. "The dysfunctional marriage of formative and summative teacher evaluation." Journal of Personnel Evaluation in Education 1 (1988): 269-273.
- Popham, W. J. Educational Evaluation. Englewood Cliffs: Prentice-Hall, Inc., 1975.
- Ratcliff, J. L. "Faculty Evaluation as a Measure of Organizational Productivity." Southern Association of Community and Junior Colleges. Occasional Paper Reprint, Vol. 2. No. 1, Jan 1984. Dr. Ratcliff is now the Director of the Center for the Study of Higher Education, Penn State University, State College, PA.
- Rauhauser, W. D. "An analysis of the quality and effectiveness of teachers' job improvement targets." Ph.D. diss., Iowa State University, 1983.

- Redfern, G. B. Evaluating teachers and administrators: A performance objectives approach. Boulder: Westview Press, 1980.
- Rice, R. M. "Effectiveness of a systematic media facilitated approach to training school administrators to conquer postobservation conferences." Ph.D. diss., Iowa State University, 1986.
- Rosenholtz, S. J. "Effective Schools: Interpreting the Evidence." American Journal of Education 93 (3) (1985): 352-388.
- Rosenshine, B. "The stability of teacher effects upon student achievement." Review of Educational Research, 40 (1970): 647-662.
- Rosenshine, B. "Teaching Functions in Instructional Programs." Elementary School Journal 83 (4) (1983): 29-45.
- Roueche, J. E. "Excellence for Students." In Celebrating Teaching Excellence: National Conference on Teaching Excellence and Conference of Presidents, Proceedings, pp. 29-34. Austin, TX: The University of Texas at Austin, 1983.
- Rucker, D. "A study of principals' teaching style preference and teachers' teaching style as a source of bias in teacher evaluation." Ph.D. diss., Iowa State University, 1981.
- Schalock, D. "Research on teacher selection" In Review of Research in Education, Vol. 7, ed. D.C. Berliner, Washington, DC: American Educational Research Association, 1979.
- Schurr, G. M. "Toward a Code of Ethics for Academics." Journal of Higher Education 53 (3) (May-June, 1982): 318-34.
- Schycker, N. K. "Creation and testing of a prescriptive-diagnostic staff development instrument for K-12 teachers." Ph.D. diss., Iowa State University, 1983.
- Scriven, M. "Summative Teacher Evaluation." In Handbook of Teacher Evaluation. National Council on Measurement in Education, ed. J. Millman, 244-271. Beverly Hills: SAGE Publications, 1981.
- Sechrest, L., and Hoffman, P.E. "The philosophical underpinnings of peer review." Professional Psychology 13 (1982): 14-18.

- Seldin, P. Successful Faculty Evaluation Programs. Crugers, NY: Coventry Press, 1980.
- Seldin, P. Changing practices in faculty evaluation. San Francisco: Jossey-Bass, 1984.
- Semones, M. M. "The development and feasibility of a data gathering instrument." Ph.D. diss., Iowa State University, 1987.
- Sergiovanni, T. J. Handbook for Effective Department Leadership: Concepts and Practices in Today's Secondary Schools. Boston: Allyn and Bacon, 1977.
- Sergiovanni, T. J. "Leadership and Excellence in Schooling." Educational Leadership 41 (2) (1984): 28-36.
- Seyfarth, J., and Nowinski, E. M. "Administrator Feedback Can Improve Classroom Instruction." NASSP Bulletin 71 (503) (1987): 47-50.
- Shanker, A. "Separating Wheat from Chaff." Phi Delta Kappan 67 (2) (1985): 108-109.
- Shanker, A. "Teachers Must Take Charge." Educational Leadership 44 (1) (1986): 12-14.
- Shavelson, R., and Russo, N. A. "Generalizability of measures of teacher effectiveness." Educational Research 19 (3) (1977): 171-183.
- Shulman, L. S. "A Union of Insufficiencies: Strategies for Teacher Assessment in a Period of Reform." Educational Leadership 46 (3) (1988): 36-41.
- Sizer, T. Horace's compromise: The dilemma of the American high school. Boston: Houghton-Mifflin, 1984.
- Smith, A. (Ed.). Evaluating Faculty and Staff: New Directions for Community Colleges. San Francisco: Jossey-Bass, 1983.
- Southern Regional Education Board. Faculty Evaluation for Improved Learning. Atlanta: Southern Regional Education Board, 1985.
- Spencer, K. M. "Post observation conferences: Factors related to success." Ph.D. diss., Iowa State University, 1985.

- StatView 512+™. Version 1.0. Brain Power, Inc. Calabases: Abacus Concepts, Inc., 1986.
- StatWorks™. Version 1.2. Philadelphia: Cricket Software, Inc., 1985.
- Stevenson, L. "Teacher performance evaluation as structured within a knowledge-based system: A question of feasibility." Ph.D. diss., Iowa State University, 1987.
- Stiggins, R.J. and Bridgeford, N.J. Performance Assessment for Teacher Development. ERIC Document Reproduction, ED 242 717, 1985.
- Stow, S., and Sweeney, J. "Developing a Teacher Performance Evaluation System." Educational Leadership 38 (1981): 538-541.
- Strike, K. and Bull, B. "Fairness and the Legal Context of Teacher Evaluation." In Handbook of Teacher Evaluation. National Council on Measurement in Education, ed. J. Millman, 303-343. Beverly Hills: Sage Publications, 1981.
- Stufflebeam, D.L., and Brethower, D.M. "Improving personnel evaluations through professional standards. Journal of Personnel Evaluation in Education, 1 (1987): 125-155.
- Sullivan, C. G. Clinical Supervision: A State of the Art Review. Alexandria: Association for Supervision and Curriculum Development, 1980.
- Sweeney, J. "Improving the Post-Observation Conference - Planning Makes a Difference." NASSP Bulletin 67 (463) (1983a): 33-36.
- Sweeney, J. "The Post Observation Conference: Key to Teacher Improvement." The High School Journal 66 (2) (1983b): 61-66.
- Sweeney, J. "Developing A Strong Climate: The Key to School Improvement." National Forum of Educational Administration and Supervision 3 (1986): 3.
- Sweeney, J., and Stow, S. "Performance Improvement: A People Program." Education 101 (3) (1981): 267-269.
- Tesch, S., Nyland, L., and Kernutt, D. "Teacher Evaluation - Shared Power Working." Educational Leadership 44 (7) (1987): 26-30.
- Thomas, M. D. Evaluation of educational personnel. Bloomington: Phi Delta Kappa Educational Foundation, 1979.

- Thornton, G. C. "The relationship between supervisor- and self appraisals of executive performance. Personnel Psychology, 21 (1968): 441-455.
- Thorson, J., Miller, R., and Bellon, J. "Instructional Improvement Through Personnel Evaluation." Educational Leadership 44 (7) (1987): 52-54.
- Uhl, Perry L. "Selecting criteria for the evaluation of school counselors based on item discrimination power." Ph.D. diss., Iowa State University, 1988.
- Veldman, D. J. and Brophy, J. E. "Measuring teacher effects on pupil achievement. Journal of Educational Psychology 66 (1974): 319-324.
- Vlaanderen, R. Trends in competency-based teacher certification. Denver: Education Commission of the States, 1980.
- Weber, J. R. "Teacher Evaluation as a Strategy for Improving Instruction. Synthesis of Literature." Elmhurst: North Central Regional Educational Laboratory, 1987.
- Weinbach, W., and Randolph, L. "Peer review for tenure and promotion in professional schools." Improving College and University Teaching 32 (1984): 81-86.
- Westerberg, Timothy R. "The effects of classroom observation by principals on conference climate and classroom performance of teachers." Ph.D. diss., Iowa State University, 1983.
- Wicks, Thomas G. "The development and testing of an optical scanner-based system of classroom data analysis." Ph.D. diss., Iowa State University, 1988.
- Wilson, R.C., Dienst, E.R., and Watson, N.L. "Characteristics of effective college teachers as perceived by their colleagues." Journal of Educational Measurement 10 (1973): 31-37.
- Wise, A., & Darling-Hammond, L. "Teacher evaluation and teacher professionalism." Educational Leadership January (1985): 28-30.
- Wise, A. E.; Darling-Hammond, L.; McLaughlin, M.W.; and Bernstein, H.T. Teacher Evaluation: A Study of Effective Practices. Santa Monica: Rand Publications, 1984.

Wuhs, Susan K. and Manatt, Richard P. "The Pace of Mandated Teacher Evaluation Picks Up." The American School Board Journal 170 (May, 1983): 28.

## ACKNOWLEDGMENT

The writer wishes to express his sincere appreciation to dissertation and program of studies committee co-chairpersons, Dr. Richard P. Manatt, Professor of Professional Studies in Education and Director of the School Improvement Model Project and Dr. Larry Ebbers, Professor of Professional Studies in Education and Chair of the Department; for their guidance and help.

Also, I offer sincere thanks for scholarly assistance, personal interest, and friendship during this study to the remaining members of my committee; Dr. Jim Sweeney, Professor of Professional Studies in Education and Educational Administration Section Leader; Dr. Harold S. "Sandy" McNabb, Jr., Professor of Plant Pathology and Professor of Forestry; Dr. Shirley Stow, Associate Professor of Educational Administration and co-director of the School Improvement Model Project; and Dr. Donald J. McKay, Associate Professor of Industrial Education and Technology.

Finally, I dedicate this study to my wife, Elaine, and my children, Kristine and James. Words cannot express what they mean to me. All have made sacrifices so that I could work toward the degree. To Elaine, much is owed: she was always there when needed -- to type, to encourage, to help, to support, to listen, and to prod. This study would never have been finished without her.



APPENDIX A.  
SUMMARY OF THE STANDARDS

Part one of the book The Personnel Evaluation Standards (Joint Committee, 1988) summarizes the standards very concisely. As a result, that summary has been reproduced in abbreviated form in the following pages.

The Propriety Standards require that evaluations be conducted legally, ethically, and with due regard for the welfare of evaluatees and clients of the evaluations.

Standard P1: Service Orientation

Evaluations of educators should promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society are met.

Explanation: The primary purpose of education personnel evaluation is to guide educators to deliver quality educational services.

Rationale: Educational systems exist to meet the needs of students and the community; so all elements of those systems, including personnel evaluation, should be directed toward achieving that purpose.

**Standard P2: Formal Evaluation Guidelines**

Guidelines for personnel evaluations should be recorded in statements of policy, negotiated agreements, and/or personnel evaluation manuals, so that evaluations are consistent, equitable, and in accordance with pertinent laws and ethical codes.

**Evaluation:** Formal guidelines are the written statements that define the purpose, procedures, and substance of the evaluation.

**Rationale:** Evaluation must be carried out in a consistent, equitable, and legal manner, regardless of who is evaluating and who is being evaluated.

**Standard P3: Conflict of Interest**

Conflicts of interest should be identified and dealt with openly and honestly, so that they do not compromise the evaluation process and results.

**Evaluation:** Conflict of interest arises when an evaluator's own goals and biases exert, or might exert, inappropriate influence on a judgment or decision.

**Rationale:** Conflict of interest can undermine the entire personnel evaluation system.

**Standard P4: Access to Personnel Evaluation Reports**

Access to reports of personnel evaluation should be limited to those individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured.

**Evaluation:** Much of the information in personnel evaluation reports is of such a nature that access to it must be restricted to those persons with a professional need to see and use the reports.

**Rationale:** Access to personnel evaluation is a highly sensitive issue.

**Standard P5: Interactions with Evaluatees**

The evaluation should address evaluatees in a professional, considerate, and courteous manner, so that their self-esteem, motivation, professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged.

**Explanation:** The evaluator must be cognizant of and responsive to the evaluatee's personal and professional needs.

**Rationale:** When the evaluator and evaluatee share a sense of professionalism and basic human dignity, they are less likely to be anxious and feel negative toward the evaluation.

The Utility Standards are intended to guide evaluations so that they will be informative, timely, and influential.

**Standard U1: Constructive Orientation**

Evaluations should be constructive, so that they help institutions to develop human resources and encourage and assist those evaluated to provide excellent service.

**Evaluation:** Evaluations are constructive when they promote the success of students, educators, and organizations.

**Rationale:** When personnel evaluation is constructive, it encourages and supports educators and their organizations to fulfill their goals and responsibilities.

**Standard U2: Defined Uses**

The users and the intended uses of a personnel evaluation should be identified, so that the evaluation can address appropriate questions.

**Evaluation:** Evaluation planners should identify and consult with each user group to clarify the purposes of the evaluation.

**Rationale:** Personnel evaluations should be guided by their intended use.

**Standard U3: Evaluator Credibility**

The evaluation system should be managed and executed by persons with the necessary qualifications, skills, and authority, and evaluators should conduct themselves professionally, so that evaluation reports are respected and used.

**Evaluation:** Institutions should take great care in appointing, training, supporting, and monitoring the persons who manage and implement personnel evaluation systems.

**Rationale:** The acceptance of an evaluation depends heavily upon the evaluatee's perceptions of the evaluator's authority, expertise, professionalism, sensitivity, trustworthiness, and efficient and effective performance.

#### Standard U4: Functional Reporting

Reports should be clear, timely, accurate, and germane, so that they are of practical value to the evaluatee and other appropriate audiences.

**Explanation:** To be useful, the information should be reported when the user needs it and when the evaluatee can best apply it to improve performance.

**Rationale:** Since evaluation reports become the basis for important judgments about the practice, status, and development needs of professionals, they should provide clear, useful, and relevant information; and they should be provided at times and in ways that facilitate decision making and action.

#### Standard U5: Follow-Up and Impact

Evaluations should be followed up, so that users and evaluatees are aided to understand the results and take appropriate actions.

**Explanation:** Those in charge of coordinating or conducting evaluations should not only reach sound assessments but should recognize that purpose and value lie in application of the findings.

**Rationale:** Follow-up activities should be a natural outgrowth of the evaluation process.

The Feasibility Standards call for evaluation systems that are as easy to implement as possible, efficient in their use of time and resources, adequately funded, and viable from a number of other standpoints.

**Standard F1: Practical Procedures**

Personnel evaluation procedures should be planned and conducted so that they produce needed information while minimizing disruption and cost.

**Explanation:** Personnel evaluation procedures are the series of actions that give a plan its practical effect.

**Rationale:** Impractical procedures can be inefficient and needlessly disruptive, detracting from individual performance and effectiveness and organizational efficiency.

**Standard F2: Political Viability**

The personnel evaluation system should be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work.

**Explanation:** Personnel evaluation policies and procedures should provide the educator, the evaluator, and other interested parties with a common focus and set of directives concerning the goals and methods of personnel evaluation.

**Rationale:** If personnel evaluation policies and procedures are understandable, cooperatively developed, acceptable to all interested parties, and officially adopted, they are likely to assure continued cooperation within the personnel evaluation program.

**Standard F3: Fiscal Viability**

Adequate time and resources should be provided for personnel evaluation activities, so that evaluation plans can be effectively and efficiently implemented.

**Explanation:** Personnel evaluation systems require substantial resources to function effectively.

**Rationale:** The justification for expenditures for personnel evaluation may be any or all of the following: better selection of new employees, improved personnel performance, improved services to students, and improved operation and general welfare of the organization.

The accuracy standards require that the obtained information be technically accurate and that conclusions be linked logically to the data.

**Standard A1: Defined Role**

The role, responsibilities, performance objectives, and needed qualifications of the evaluatee should be clearly defined, so that the evaluator can determine valid assessment data.

**Explanation:** All parties to the evaluation process should have the same understanding of the position requirements before the evaluation process is designed.

**Rationale:** A carefully developed and sufficiently detailed and delineated description of the role, responsibilities, performance objectives, and qualifications is prerequisite to specifying relevant assessment criteria.

**Standard A2: Work Environment**

The context in which the evaluatee works should be identified, described, and recorded, so that environmental influences and constraints on performance can be considered in the evaluation.

**Explanation:** The context in which a person works can affect performance. Among the many contextual factors that can influence or constrain performance are those associated with organizational structure and process, such as educational goals and objectives, curriculum mandates, leadership and supervisory practices, financial resources, and decision-making policies.



**Rationale:** Holding educators accountable for the effects of variables they cannot control or influence is likely to lead to resentment and low morale.

**Standard A3: Documentation of Procedures**

The evaluations procedures actually followed should be documented, so that the evaluatees and other users can assess the actual, in relation to intended, procedures.

**Explanation:** The documentation should focus on the steps, forms, appeal procedures, reporting and recording schedule, follow-up, and due process procedures.

**Rationale:** The effectiveness of an evaluation is linked to how well the evaluator, evaluatee, and other appropriate persons understand and accept the evaluation procedures.

**Standard A4: Valid Measurement**

The measurement procedures should be chosen or developed and implemented on the basis of the described role and the intended use, so that the inferences concerning the evaluatee are valid and accurate.

**Explanation:** Evidence for the validity of an evaluation procedure must be assembled and available.

**Rationale:** Validity is the single most important issue in the assessment of any evaluation process.

**Standard A5: Reliable Measurement**

Measurement procedures should be chosen or developed to assure reliability, so that the information obtained will provide consistent indications of the performance of the evaluatee.

**Explanation:** A reliable measure is one that provides consistent information about the performance being assessed.

**Rationale:** Individual personnel evaluation instruments and the procedures used should have levels of reliability that are acceptable for intended uses.

**Standard A6: Systematic Data Control**

The information used in the evaluation should be kept secure, and carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.

**Explanation:** Information should be carefully processed and checked for accuracy at every stage of collection, storage, and retrieval.

**Rationale:** Personnel decisions resulting from evaluations can be no better than the information upon which they are based.

**Standard A7: Bias Control**

The evaluation process should provide safeguards against bias, so that the evaluatee's qualifications or performance are assessed fairly.

**Explanation:** There is constant potential for the intrusion of bias in personnel evaluations.

**Rationale:** The presence of bias can entirely undermine an evaluation system.

**Standard A8: Monitoring Evaluation Systems**

The personnel evaluation system should be reviewed periodically and systematically, so that appropriate revisions can be made.

**Explanation:** Personnel evaluation involves purposes, procedures, instruments, reports, and uses of findings.

**Rationale:** Personnel evaluation is difficult to do well and is subject to mistakes and complaints.

APPENDIX B.

THE INITIAL QUESTIONNAIRE

Lloyd O. Roettger  
 January 25, 1989

Dear Colleague,

I'm a Ph.D. candidate in Educational Administration and Higher Education. It is my intent to conduct a national study about two-year college instructor performance evaluation as my dissertation. You are being asked to identify those primarily responsible for instructor evaluation in your institution. I will be contacting those you name with a data gathering questionnaire. Examples:

- Norma Jones, Dean of Instruction
- John Brown, Business Division Chair
- Larry Green, Electronics Department Head
- LouAnn Lovely, Practical Nursing Program Supervisor
- Garry Goodwrench, Auto Mechanics Program Leader

Please complete the bottom portion of this page and return it to me as soon as possible. Thank you very much for your assistance.

Name of Institution: _____	<u>Type of Institution</u>
	<input type="checkbox"/> Junior College
Address: _____	<input type="checkbox"/> Community College
	<input type="checkbox"/> Technical College
City, State, Zip: _____	<input type="checkbox"/> Vocational-Technical Coll.
	<input type="checkbox"/> Vocational/Trade College
Approximate FTE _____	<input type="checkbox"/> Other _____

	<u>Name</u>	<u>Title</u>
Evaluator #1:	_____	_____
Evaluator #2:	_____	_____
Evaluator #3:	_____	_____
Evaluator #4:	_____	_____
Evaluator #5:	_____	_____
Evaluator #6:	_____	_____

Please return to:  
 Lloyd O. Roettger  
 112 N. 18th St.  
 Marshalltown, Iowa 50158

APPENDIX C.

THE PRIMARY DATA GATHERING QUESTIONNAIRE

## Two - Year College

<b>SUPERVISION / EVALUATION of INSTRUCTION QUESTIONNAIRE</b>
--

**Attention:** You have been selected as a participant in a nationwide research project about professional instructor performance evaluation in the two year college. Your responses are valued and will help in developing the most complete set of data available about the "state-of-the-art" of personnel evaluation of two-year college instructors. (If you could benefit from an abstract of the complete report please indicate your name and mailing address at the end of this questionnaire.)

Position TITLE of Person Completing this Questionnaire:

\_\_\_\_\_

Please check any duties that you perform as an instructional leader for your institution:

Assist students	_____	Supervise Instructors	_____
Supervise Other Administrators	_____	Supervise Support Staff	_____
Formally Evaluate Instructors	_____	Formally Evaluate Other Administrators	_____
Formally Evaluate Support Staff	_____	Support Instructors	_____
Supervise the curriculum	_____	Monitor student progress	_____
Support improvement of instruction	_____	Provide orderly environment	_____
Maintain physical facilities	_____	Maintain community relations (PR)	_____
Promote professional growth	_____	Emphasize student achievement	_____
Assist in economic development	_____	Serve the community	_____

Please indicate the number of persons you supervise/evaluate by their titles:

Other Non-Teaching Administrators	_____	Classroom Instructors	_____
Other Teaching Administrators	_____	Clinical Instructors	_____
Technicians	_____	Laboratory Instructors	_____
Support Staff	_____	Shop Instructors	_____
Others	_____	i.e.	_____

Please indicate the approximate percentage of time you spend annually doing formal evaluations of instructors: \_\_\_\_\_

Do you have an institution or division-wide evaluation model? Yes \_\_\_ No \_\_\_

If "Yes", please send an example of the process or model.

In his book Successful Teacher Evaluation, Thomas L. McGreal provides a taxonomy for the classification of evaluation models and procedures. The five statements that follow are representative of the models identified. Select the statement that best describes your evaluation model / process:

\_\_\_\_\_ A. The model is characterized by high supervisor - low teacher involvement; evaluation is synonymous with observation; major emphasis is on summative evaluation; standardized criteria; and comparative judgements. This process usually relies on definitions, procedures, and processes that are traditional.

\_\_\_\_\_ B. The model is characterized by an emphasis on an individualized approach to evaluation. Instructors and evaluators meet and confer to set and monitor goals. Generally, no checklist of criteria is used. Self-evaluation may be a component of this model.

\_\_\_\_\_ C. The model is characterized by evaluation that is based upon the results or outcomes of student achievement tests or on competency-based evaluations, but not on methods, styles, or processes. Generally, the instruments for assessing student growth are norm-referenced tests and criterion-referenced tests.

\_\_\_\_\_ D. The model is characterized by a close relationship between the instructor and the supervisor with emphasis on collegial rather than authoritarian orientation. It takes its principal data from the classroom and is designed to improve the instructor's performance.

\_\_\_\_\_ E. The model is characterized by a belief that teaching is an art, that the quality of the performance the instructor exhibits is likened to an aesthetic experience. The evaluation is more subjective and, perhaps, less precise.

How many formal observations are made before a summative evaluation report is written? \_\_\_\_\_

Do you conduct a pre-observation conference before each formal evaluation? Yes \_\_\_\_\_ No \_\_\_\_\_

Do you conduct a post-observation conference after each formal evaluation? Yes \_\_\_\_\_ No \_\_\_\_\_

The Joint Committee on Standards for Educational Evaluation, chaired by Daniel L. Stufflebeam, recently published a book entitled The Personnel Evaluation Standards. The book enumerated twenty-one standards in four basic areas. Please use the following scale to rate the quality of the model or process your institution uses for instructor evaluation on each of the standards. SA = Strongly Agree, A = Agree, N = Not Applicable, D = Disagree, and SD = Strongly Disagree

**The Propriety Standards:** The Propriety Standards require evaluations are conducted legally, ethically, and with due regard for the welfare of evaluatees and clients of the evaluations.

Circle Your Response

- |    |   |   |   |    |  |
|----|---|---|---|----|--|
| SA | A | N | D | SD | 1. Evaluations of educators promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society are met.   |
| SA | A | N | D | SD | 2. Guidelines for personnel evaluations are recorded in statements of policy, negotiated agreements, and/or personnel evaluation manuals, so that evaluations are consistent, equitable, and in accordance with pertinent laws and ethical codes.                              |
| SA | A | N | D | SD | 3. Conflicts of interest are identified and dealt with openly and honestly, so that they do not compromise the evaluation process and results.   |
| SA | A | N | D | SD | 4. Access to reports of personnel evaluation is limited to those individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured.  |
| SA | A | N | D | SD | 5. The evaluation process addresses evaluatees in a professional, considerate, and courteous manner, so that their self-esteem, motivation, professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged. |



**The Utility Standards:** The Utility Standards are intended to guide evaluations so that they will be informative, timely, and influential.

- SA A N D SD 6. Evaluations are constructive, so that they encourage and assist those evaluated to provide excellent service.
- SA A N D SD 7. The users and the intended uses of personnel evaluation are identified, so that the evaluation can be guided by its intended use and so that everyone is clear as to its purpose.
- SA A N D SD 8. The evaluation system is managed and conducted by persons with the necessary qualifications, skills, and authority; and evaluators conduct themselves professionally, so that evaluation reports are respected and used.
- SA A N D SD 9. Reports are clear, timely, accurate, and germane, so that they are of practical value to the evaluatee and other appropriate audiences such as merit pay or career ladder committees, tenure and promotion committees, and performance review boards..
- SA A N D SD 10. Evaluations are followed up, so that evaluatees are aided to understand the results and take appropriate actions.

**The Feasibility Standards:** Feasibility Standards call for evaluation systems that are as easy to implement as possible, efficient in its use of time and resources, adequately funded, and viable from a number of other standpoints.

- SA A N D SD 11. Personnel evaluation procedures are planned and conducted so that they produce needed information while minimizing disruption and cost.
- SA A N D SD 12. The personnel evaluation system is developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work.
- SA A N D SD 13. Adequate time and resources are provided for personnel evaluation activities, so that evaluation plans can be effectively and efficiently implemented.

**The Accuracy Standards:** The Accuracy Standards require that the obtained information be technically accurate and that conclusions be linked logically to the data.

- SA A N D SD 14. The role, responsibilities, performance objectives, and needed qualifications of the evaluatee are clearly defined, so that all parties have the same understanding of the position requirements before the evaluation process is designed.
- SA A N D SD 15. The context in which the evaluatee works is identified, described, and recorded, so that environmental influences and constraints on performance can be considered in the evaluation.
- SA A N D SD 16. The evaluation procedures actually followed are documented, so that the evaluatees and other users can assess the actual, in relation to intended, procedures.
- SA A N D SD 17. The measurement procedures are chosen or developed and implemented on the basis of the described role and the intended use, so that the inferences concerning the evaluatee are valid and accurate.
- SA A N D SD 18. Measurement procedures are chosen or developed to assure reliability, so that the information obtained provides consistent indications of the performance of the evaluatee.
- SA A N D SD 19. The information used in the evaluation is kept secure, and carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.
- SA A N D SD 20. The evaluation process provides safeguards against bias, so that the evaluatee's qualifications or performance are assessed fairly.

SA A N D SD 21. The personnel evaluation system is reviewed periodically and systematically, so that appropriate revisions can be made.

Do you use the same evaluation instrument for all instructors? Yes \_\_\_ No \_\_\_

If "No" why not? \_\_\_\_\_

Please attach a copy of each instructor evaluation instrument used.

Does your state require that instructors be evaluated? Yes \_\_\_ No \_\_\_

Does your college require that instructors be evaluated? Yes \_\_\_ No \_\_\_

Is peer feedback a part of your instructor evaluation process? Yes \_\_\_ No \_\_\_

Is an academic council used to evaluate performance? Yes \_\_\_ No \_\_\_

Is student feedback a part of your instructor evaluation process? Yes \_\_\_ No \_\_\_

Is student achievement a part of your instructor evaluation process? Yes \_\_\_ No \_\_\_

Are outside evaluators a part of your instructor evaluation process? Yes \_\_\_ No \_\_\_

Does your college utilize a "pay for performance", "merit pay", "career Ladder", or other type of incentive program? Yes \_\_\_ No \_\_\_

**Demographics about you:**

Did you complete a teacher preparation program? Yes \_\_\_ No \_\_\_

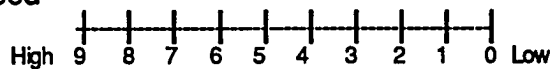
Highest earned degree and area? \_\_\_\_\_

How many hours of evaluator training have you had? \_\_\_\_\_

What certification in evaluation do you hold? \_\_\_\_\_

Is certification required? Yes \_\_\_ No \_\_\_

Circle the level at which you perceive the need for additional training?



**Type of institution**

\_\_\_\_\_ Junior College

\_\_\_\_\_ Community Technical College

\_\_\_\_\_ Community College

\_\_\_\_\_ Vocational Technical College

\_\_\_\_\_ Technical College

Other \_\_\_\_\_

Size of Institution \_\_\_\_\_ FTE (approximate for Fall 1988)

**Please Return To:**

Lloyd O. Roettger  
Iowa State University, College of Education  
School Improvement Model Project  
E005 Lagomarcino Hall  
Ames, Iowa 50011  
(or use the enclosed label)

**If you'd like a copy of the results:**

Name \_\_\_\_\_  
Address \_\_\_\_\_  
City \_\_\_\_\_  
State \_\_\_\_\_ Zip \_\_\_\_\_

APPENDIX D.  
THE LETTER THAT ACCOMPANIED THE PRIMARY DATA GATHERING  
QUESTIONNAIRE

Lloyd O. Roettger  
March 10, 1989

Dear Colleague,

I'm a Ph.D. candidate in Educational Administration and Higher Education. It is my intent to conduct a nationwide study about two year college instructor performance evaluation as my dissertation. You have been identified as an instructional leader in your college by a member of your college's administration. Therefore, I am asking you to take a few moments of your time to participate in this effort to create the most complete set of data available about the "state-of-the-art" in personnel evaluation of two-year college instructors.

Dr. Larry Ebbers, Professor and Department Executive Officer for Professional Studies in Education at Iowa State University and Dr. Richard Manatt, Professor and Director of the School Improvement Model Project at Iowa State University, are my co-chairs and major professors. The dissertation project is sampling two-year colleges in every state. The sample is stratified by size and type of institution; therefore, it is critical that every questionnaire be accounted for, so please respond if at all possible.

I thank you very much for your assistance and for your time. Without the help of good folks like you, this research could never succeed.

Appreciatively I Remain,

Lloyd O. Roettger  
(515-752-4249)

