

## INFORMATION TO USERS

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

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

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.
2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University  
Microfilms  
International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106  
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

7921269

**SUWANASUK, PRAPAPUN**  
**AN EVALUATION MODEL FOR TEACHER PREPARATION**  
**PROGRAMS IN HIGHER EDUCATION IN THE**  
**SOUTHWEST.**

**THE UNIVERSITY OF OKLAHOMA, PH.D., 1979**

University  
Microfilms  
International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106

AN EVALUATION MODEL FOR TEACHER PREPARATION PROGRAMS  
IN HIGHER EDUCATION IN THE SOUTHWEST

APPROVED BY

Michael Langensack  
W. Higgins  
Edward J. Johnson  
Gene Ringleton

DISSERTATION COMMITTEE

## DEDICATION

To my parents, Prasert Suwanasuk and Sudjai Suwanasuk.

## ACKNOWLEDGEMENTS

Gratitude is sincerely expressed to Dr. Michael Langenbach who served as chairman of the doctoral committee and director of this study. His sincere interest and valuable assistance throughout the doctoral program are acknowledged with appreciation. Deepest gratitude is also expressed to the other members of the doctoral committee, Dr. Thomas W. Wiggins, Dr. G. Gene Pingleton, and Dr. Lloyd J. Korhonen for their interest and assistance in the writing to this study.

The author also wishes to express her appreciation to Mr. Sompoch Iamsupasit, her parents, brothers, and sisters, Prasert, Sudjai, Praprit, Prapussorn, Supreeya, and Nijja for their encouragement, support, patience, and understanding throughout all phases of her graduate study.

## TABLE OF CONTENTS

	Page
LIST OF FIGURES.....	vi
Chapter	
I. INTRODUCTION.....	1
II. HISTORY OF EVALUATION.....	6
III. THEORETICAL FRAMEWORK.....	12
IV. RELATED LITERATURE.....	20
V. DESIGN OF THE STUDY.....	31
VI. GENERAL EVALUATION MODEL.....	36
VII. COLLEGE OF EDUCATION MODEL.....	46
VIII. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	102
BIBLIOGRAPHY.....	108

## LIST OF FIGURES

	Page
1. A layout of statements and data to be collected by the evaluator of an education program.....	15
2. General Evaluation Model.....	39
3. College of Education Model.....	47
4. Needs assessment processes.....	50
5. Format of the questionnaire.....	52
6. Presenting data.....	55
7. Setting and ranking goals processes.....	58
8. The format of rating scale.....	65
9. A means-ends perspective of planning.....	68
10. Program planning processes.....	70
11. Program improvement processes.....	80
12. Summative evaluation processes.....	86
13. A representation of the process of judging the merit of an educational program.....	88
14. Some possible outcome patterns from the introduction of an experimental variable at point X into a Time Series of Measurements, $O_1 - O_8$ .....	94

AN EVALUATION MODEL FOR TEACHER PREPARATION PROGRAMS  
IN HIGHER EDUCATION IN THE SOUTHWEST

CHAPTER I

INTRODUCTION

One problem that most of the teacher colleges in the Southwest area face today is how the effectiveness of the teacher preparation programs should be assessed. In order to assess the teacher preparation programs, an evaluation plan needs to be developed. Also an evaluation model should be selected or developed to be used in the evaluation plan. Unfortunately, most of the evaluation models that are available today are ineffective for use in teacher preparation programs. Thus, an evaluation model needs to be developed. Before it can be developed, the nature of the teacher preparation program has to be studied.

There are many colleges and universities in the Southwest area which offer teacher preparation programs for state residents and non-residents. Basically, most of the teacher preparation programs are established to supply the state with teachers. Most of the teacher preparation programs in these colleges and universities in the Southwest area have common characteristics in their requirements (based on the College Bulletins from the University of Arkansas, Louisiana State University, Northwestern State University, University of New Mexico, New



Mexico State University, University of Oklahoma, Oklahoma State University, University of Texas, and Texas Tech University). All of them offer doctoral programs in their Colleges of Education in addition to Bachelors and Masters' programs.

In each university, the following generalizations apply:

1. Each student has to take between 48-57 semester hours in Academic Foundations, such as English, Mathematics, Science, History, Physical Education, and Political Science.

2. Each student has to take between 24-70 semester hours in Specialized Areas which can be Mathematics, Science, Music, History, Reading, and Language Arts.

3. Each student has to take between 18-44 semester hours in Professional Education which is composed of:

- a) Basic Psychology, which includes Educational Psychology, Child Psychology, and Adolescent Psychology.

- b) Foundations of Education, which includes History and Philosophy of Education, School in American Culture, and Social Foundations.

- c) Curriculum and Instruction, which includes Materials and Methods in Teaching (specific subject areas such as Mathematics, Science, etc.) and Methods of Teaching.

- d) Student Teaching.

4. Elective course can be taken in order to complete a certain amount of the requirements of semester hours for the degree. Total requirements for Bachelor Degrees vary from 126 to 136 semester hours.

Most of the courses in Professional Education are offered within the College of Education except that one college (LSU) offered Basic Psy-

chology in the Psychology Department. Figures show that only 14-34 percent of the credits toward the B.Ed. are earned within the College of Education. This show that programs in the College of Education are highly interdisciplinary when compared with other colleges such as Engineering or Business.

The number of semester hours required vary from college to college and from major area to major area. The University of Oklahoma requires students to take 52 semester hours in Academic Foundation, 26 semester hours in Professional Education, and the rest in Specialized Areas and Electives. These hours are required in order to complete a Bachelors Degree of Education.

#### Problem Statement

How can the worth and effectiveness of the courses that have been offered in a College of Education be judged? How can it be known that the programs are worth-while and serve the needs of the society? How can it be known to what extent the programs achieve their goals? If they do not achieve their goals, how can they be improved? These questions lead to:

Can an evaluation model be devised to be used with teacher preparation programs?

Most of the Colleges of Education in the Southwest States have similar functions and characteristics. The primary function is to provide degree and teacher certification programs for both undergraduate and graduate students who plan a career in teaching in elementary and secondary schools. It is not a practical idea to uncritically use one of the many evaluation models which have been developed in other institutions in the evaluation of a college's program. The best way to obtain program

evaluation in a college is to develop or to modify a model based on the functions and needs of the college; a model that is general and practical enough to evaluate all types of courses and programs that have been offered in the Colleges of Education.

Because there are so many colleges and universities in the Southwest area, it is difficult to develop any general evaluation model that can be applied to each of them. To solve this problem, a general evaluation model will be developed along with a specific model.

There are many evaluation models that have been developed in recent years. Most of them grew out of the needs institutions and programs had, and continue to have, regarding examination and analysis of their methods and products.

Steele (1977) classified these models into six groups.

Group 1: Evaluation as Input into Decision Making

- a. CIPP Model (Stufflebeam, 1971)
- b. CSE Model (Alkin, 1969b)
- c. Discrepancy Evaluation (Provus, 1971)
- d. IPI Formative Evaluation (Lindvall & Cox, 1970)

Group 2: Evaluation of Program Parts

- a. Program Evaluation and Review Technique (PERT)
- b. Social System Model (Loomis, 1960)

Group 3: Evaluation--Kind of Data; Types of Activities

- a. Countenance Model (Stake, 1967)
- b. System Role Model (Knox, 1969)

Group 4: Evaluation Process

- a. Appraisal Model (Harris, 1947)

- b. Data Management (Stufflebeam et al., 1971)
- c. Transactional Evaluation (Rippey, 1972)

Group 5: Results--Attainment of Objectives

- a. Tyler Model (Tyler, 1950)
- b. Program Planning Budgeting System (PPBS)
- c. NEA Model (Taba & Sawin, 1962)

Group 6: Evaluation of Outcomes and Effects

- a. Goal-Free Evaluation (GFE) (Scriven, 1972b)
- b. Accountability Model (Browder, 1971)

There are other evaluation models that have not been classified to date.

Purpose of the Study

The purpose of this study is to develop an evaluation model that can be implemented into the teacher preparation programs in the colleges and universities in the Southwest area. This model will be used as a guideline in providing information on the effectiveness of the programs. It, also, will provide the basic ideas on how and/or whether the programs should be improved or maintained or whether they should be terminated.

## CHAPTER II

### HISTORY OF EVALUATION

J. M. Rice, who has been called the "pioneer and path-maker among American Scientific students of education" (Ayres, 1918), is a pioneer of educational evaluation in so much as he carried out one of the first large scale assessments of an educational practice and product. In 1892 Rice administered spelling tests to thousands of pupils in thirty-six cities across the nation and talked with 1,200 teachers. He used his finding--that there was no relation between time spent on spelling drill and ability to spell--to argue for a reduction in time spent on teaching spelling. He carried out what was probably the first serious study that took careful account of educational outcomes. Thus, the first mass objective achievement test was, in fact, used for curriculum evaluation. But the idea caught on slowly.

During World War I, the test and measurement movement became popular. The United State Army wanted to use tests to classify personnel. To serve this need, Army Alpha was developed. Army Alpha is a group intelligence test based on the work done by Arthur Otis (the first individual intelligence test had been developed by Binet, called the Binet-Simon Scale in 1905). In 1923 the first standardized achievement battery was published; the Stanford Achievement Test for use in the elementary grades. Before 1930, several kinds of standardized tests were developed and had been admi-

nistered widely (Wrightstone, Justman & Robbins, 1956). The purpose was to see the differences among individuals.

During 1930's, tests were given almost exclusively for the purpose of making judgments about individuals. The concept of using tests to help to evaluate the school program and curriculum had not yet been developed. In the period of 1920's and 1930's, administrators and accrediting agencies still used descriptive features of the school program in judging adequacy. Instead of collecting direct evidence of educational impact, they judged schools in terms of size of budget, students/staff ratio, square feet of laboratory space, and the number of advanced credits accumulated by the teacher in order to make inferences about program quality.

Another movement in evaluation during 1930's was led by Ralph W. Tyler. His Eight-Year Study (Smith & Tyler, 1942) has become a landmark in evaluation study. The principles and procedures advocated by Tyler have served as the basis for most of the major efforts in evaluation since that time. The Eight-Year Study was financed by the Progressive Education Association. The purpose of the study was to determine how students of progressive secondary schools compared with students of traditional schools. Smith and Tyler, (1942) concentrated on the objectives of educational programs. They defined the objectives of programs in behavioral terms and made them the basis of instrument development and evaluation.

In the late 1930's, the word "evaluation" came into popular usage in education. Evaluation expresses a broader concept than measurement. Monroe (1945) has distinguished between measurement and evaluation by indicating that in measurement the emphasis is upon single aspects of subject-matter achievement or specific skills and abilities, whereas in evaluation

the emphasis is upon broad personality changes and major objectives of an educational program. Taba (1962) made it more clearcut when she stated that measurement is only one part of evaluation. Educational measurement usually tends to concentrate on narrow, specific, and well-defined characteristics. While evaluation depends on measurement, its concern is with a broader profile of characteristics and attainments.

During World War II, evaluation became the main function in selecting and training personnel. Matching the right person to the right job was a major objective.

The most significant improvement in evaluation occurred after Russia launched Sputnik into space in 1957. The curriculums of many schools were changed and reformed shortly thereafter. Millions of dollars were spent revising content and methods of courses. Congress and federal agencies began to demand evidence that their investments were producing desired results.

In response to the demand, Cronbach (1963) provided a new dimension to the context of evaluation. Like Tyler, Cronbach believed that educational goals and objectives have to be set, but his framework for the evaluation process was different. He evaluated not only the product, but the process also. Cronbach said that evaluation should be used to understand how the course produces its effects and what parameters influence its effectiveness (p. 675).

There are three types of decisions that evaluation should serve: course improvement, decisions about individuals, and administrative regulation. Cronbach did not consider it worthwhile to evaluate the product (end of the course) when evaluation is used to improve the course. He

stated that evaluation used to improve the course while it is still fluid contributes more to improvement of education than evaluation used to appraise a product already placed on the market (p. 675). He also did not see the importance of comparing one course with another. He warned evaluators to be aware of the process of evaluation because he believed that evaluation was a diversified activity and that no one set of principles would suffice for all situations.

After the enactment of Titles I and III of the Elementary and Secondary Education Act (ESEA) in 1966, most of the educational programs were required to be accountable. Because of accountability requirements, a new era of evaluation began. New theories have been developed. The techniques of evaluation have become more advanced and accurate.

The most important concepts regarding evaluation introduced by Scriven (1967) were formative and summative evaluation. Formative evaluation is concerned with the evaluation context, formulating goals and objectives. Summative evaluation is an evaluation of a completed product, aimed at the potential consumer (see Sanders & Cunningham, 1973, 1974). Scriven conceived evaluation as a process of judging the worth of something (program, for example).

A Countenance model was developed by Stake in 1967. Stake (1967) stated that description and judgment were two basic acts of evaluation. He presented his evaluation model in the form of a matrix system. His model is composed of four columns and three rows. The first two columns are called descriptive, and are composed of intents and observations. The last two columns are called judgmental and are composed of standards and judgments. Each column has three rows, called antecedents, transactions,



and outcomes. Stake also suggested that there are two principal ways of processing descriptive evaluation data; finding the contingencies among antecedents, transactions, and outcomes and finding the congruence between intents and observations. The Countenance model can be used in the classroom as well as in national programs (see Borich, 1971; Draayer, 1974).

One of the most widely used evaluation model is the CIPP (Context, Input, Process, and Product Evaluation) model (Stufflebeam, 1968, 1971; Stufflebeam et al., 1971). This model was developed as a means to systematically provide timely evaluation information for use in decision making.

Another decision management model is the Center for the Study of Evaluation (CSE) model (Alkin, 1969). The CSE model is similar to the CIPP model. Alkin (1969) stated that the first two and the last need areas in the CSE model are somewhat similar, respectively, to "context," and "product," as presented by Stufflebeam (1968). What he referred to as "process" we have chosen to think of as two separate stages, program implementation and program improvement (p. 3). He also pointed out the differences are the emphasis on the descriptions of need areas of the CSE model and the stages of the CIPP model.

The most controversial concept in evaluation is Goal-Free Evaluation (Scriven, 1972b). Scriven (1972b) stated that evaluation of stated goals is unnecessary (p. 2). He recognized goals when he indicated goals are only a subset of anticipated effects; that they are the ones of special importance, or the one distinctive of the project. He was concerned that an evaluator would not look at other anticipated and unanticipated effects except the stated goals, while sometimes other anticipated effects were

more important than the stated goals. He also pointed out that when an evaluator knows the goals, he could develop tunnel vision upon seeing a set of goal statements and use only the stated goals as clues for identifying outcomes variables. Scriven (1972b) recommended the use of Goal-Free Evaluation in evaluation.

## CHAPTER III

### THEORETICAL FRAMEWORK

Smith and Tyler (1942) stated that evaluation should be conducted because it can:

1. provide a periodic check which gives direction to the continued improvement of the program of the school.
2. help to validate some of the important hypotheses upon which the program operates.
3. furnish data about individual students essential to wise guidance.
4. give a more satisfactory foundation for the psychological security of the staff, of parents, and of students.
5. supply a sound basis for public relations (pp. 5-11).

The main purpose of evaluation in education is to improve curriculum (Scriven, 1967), but evaluation also can contribute to the construction of a curriculum, the prediction of academic success, or the improvement of an existing course. Cronbach (1963) stated that evaluation is a fundamental part of curriculum development, not an appendage. Its job is to collect facts that a course developer can and will use to do a better job, and facts from which a deeper understanding of the educational process will emerge (p. 683). Weiss (1972b) looked at evaluation as a tool to provide information and

resolve problems related to curriculum programs, instructional strategies, innovative structures, and, on a broader scale, the total school enterprise.

There are five approaches to evaluation that will be mentioned in this study. They are the CIPP model (Stufflebeam, 1968), CSE model (Alkin, 1969), Countenance model (Stake, 1967), Tyler model (Tyler, 1942), and Goal-Free Evaluation (Scriven, 1972b).

### CIPP Model

The CIPP (Context, Input, Process, and Product) evaluation model was developed by Stufflebeam in 1968. This model is generally known as the decision management model (Worthen & Sanders, 1973; Popham, 1975; and Steele, 1977). Stufflebeam (1971) defined evaluation as "the process of delineating, obtaining, and providing useful information for judging decision alternatives" (p. 2). According to the CIPP Model there are four kinds of decisions: planning, structuring, implementing, and recycling, which respectively are served by context, input, process and product evaluation.

Context Evaluation. The purpose of context evaluation is to systematically provide information that can be used by decision-makers to make planning decisions regarding the establishment of new objectives, modification of existing objectives, or confirmation of present objectives.

Input Evaluation. The purpose of input evaluation is to identify and assess alternative program strategies for achieving given objectives and to provide information to assist in detailing particular strategies.

Process Evaluation. The purpose of process evaluation is to provide information during the implementation stages of a project or program which can assist program managers to operate the program according to its

design, improve the program design as effects are indicated under operating conditions, and to make structuring decisions which could not be made during the preparation of the program design.

Product Evaluation. The purposes of product evaluation are to relate outcomes to objectives and to assess the overall worth of a procedure in terms of its effects.

### CSE Model

The CSE model has been developed at the Center for the Study of Evaluation, UCLA by Alkin in 1969. This model also has been known as decision management model. Alkin defined evaluation as the process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives.

CSE model has five areas of evaluation that may be identified.

Needs Assessment. It involves stating potential educational goals or objectives, deciding which of these are of highest priority, and determining how well the existing educational program is meeting these objectives.

Program Planning. It involves making decisions about the kinds of programs or combinations of programs (or program components) that should be adopted to meet the problems identified in the needs assessment.

Implementation Evaluation. It focuses on whether the procedures specified in the program plan are actually carried out in the intended manner.

Progress Evaluation. It is aimed at determining the extent to which the program is actually making gains towards achieving its objectives.

Outcome Evaluation. They lead to final judgments regarding the

general worth of a total program.

### Countenance Model

Stake developed the Countenance Model in 1967. This model has been known as a judgmental strategy model. Stake (1967) emphasized description and judgment. He said both of them are essential in evaluation. He arranged his model as a matrix which can be seen in the Figure 1.

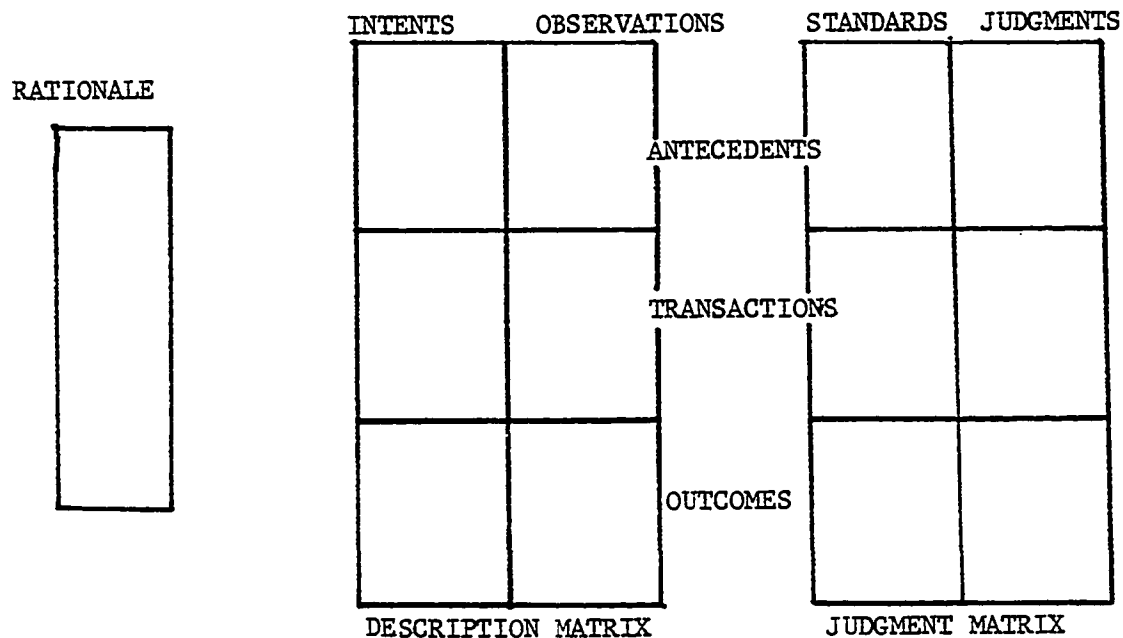


Figure 1: A layout of statements and data to be collected by the evaluator of an education program. (From "The Countenance of Educational Evaluation" by R. Stake, Teacher College Record, 1967, 68, 523-540.) (p. 529).

The horizontal side of the matrix includes the three major categories of data to be examined.

1. Antecedents are conditions existing before the teaching that may affect the outcome.
2. Transactions are events that occur while a program is in progress.
3. Outcomes are the consequences, or the product, of the program.

To improve the program, the evaluator is concerned not only with the information recorded under the three major gestalts, but also with the relationship or contingency that may or may not exist among the variables under study.

The vertical side of the matrix describes the two kinds of descriptive data (intentions and observations) and the two kinds of judgmental data (standards and judgments).

Intentions are descriptions of the planned-for environmental conditions, the planned-for teaching methods and content coverage, and intended student behavior. Emphasis is given to what teaching as well as what learning is intended.

Observations are descriptions of what actually did occur related to the antecedents, transactions, and outcomes. This category of data can include direct observations. It also includes the realm of data about transactions and outcomes secured through the use of tests, check lists, and other instruments.

Stake suggested two principal ways of processing descriptive evaluation data: finding the contingencies among antecedents, transactions, and outcomes and finding the congruence between intentions and observations.

Standards are necessary to establish for making judgments about program worth. Standards may be absolute--a reflection of personal judgment--or relative, representing a comparison with the descriptive data of alternate programs.

Judgments can be done by using absolute standards and relative standards.

### Tyler Model

The Tyler model is a goal-attainment model. Tyler (1942) stated that evaluation provides a means for the continued improvement of education and for an ever deepening understanding of students with a consequent increase in the effectiveness of our educational institutions (p. 501). He believed that educational objectives are important in the education process which should effect changes in the behavior of the students. He also suggested six steps in performing an evaluation.

1. It is necessary for the school to formulate a statement of its educational objectives, then these statements of objectives are classified into major types.
2. Define each of these types of objectives in terms of behavior.
3. Identify situations in which students can be expected to display these types of behavior so that we may know where to go to obtain evidence regarding this objective.
4. Selection and trial of promising methods for obtaining evidence regarding each type of objective.
5. Select on the basis of this preliminary trial the more promising appraisal methods for further development and improvement.



6. Devise means for interpreting and using the results of the various instruments of evaluation (pp. 498-500).

In collecting data, Tyler (1942) suggested using a variety of techniques such as observational records, anecdotal records, questionnaires, interviews, check lists, records of activities, products made and the like. In the process of interpreting the data, Tyler suggested each school and college should develop methods for interpretation and use the results of appraisal to improve the educational program and guide individual students more wisely.

#### Goal-Free Evaluation

Goal-Free Evaluation (GFE) was proposed by Scriven in 1972. GFE is one of the most controversial concepts in evaluation because Scriven (1972b) said that consideration and evaluation of stated goals was unnecessary (p. 1). This does not mean that he wants to omit goals. He suggested an evaluator should be aware of all the factors that affect the outcome of the program. He called all of these effects anticipated effects (p. 2). Goals are only a subset of the anticipated effects. He also said that goals are often stated so vaguely that they include not only desirable activities but also activities considered undesirable by almost anyones standards (p. 2). Almost all projects either fall short of their goals or over-achieve them. So, he claims, it is a waste of time to rate the goals.

Scriven (1972b) suggested that GFE should be used in summative evaluation. He believed that the less the external evaluator hears about the goals of the program, the less he will develop tunnel vision, and the

more attention he will direct to looking for actual effects. It can be said that GFE will encourage the evaluator to be attentive to a wider range of program outcomes. Scriven is not recommending GFE as a replacement for Goal-Base Evaluation (GBE), but as a supplement to more goal-oriented frameworks.

## CHAPTER IV

### RELATED LITERATURE

The evaluation process which emphasized the behavioral objectives proposed by Tyler has been criticized recently (Guba, 1969; Airasian et al., 1972). Guba (1969) stated that evaluation no longer focused solely on the student, but could provide insights about curriculum and other educational procedures as well. Evaluation came to have utility not only for judging a product (student achievement, for example), but also a process (the means of instruction, for example). Airasian et al. (1972) presented another point of view. They claimed the Tyler rational for curriculum development and evaluation can lead to premature closure, with respect to both the selection of course objectives and the evaluation of the impact of the curriculum. The process of evaluation should not only emphasize the intended outcomes but also the unintended outcomes, or side-effects (Airasian et al., 1972; Scriven, 1972b). They also pointed out that one of the principal difficulties with Tyler's model is that it was primarily teacher-centered.

Finn (1972) classified the evaluation procedure proposed by Tyler (1942) as a measurement. He said evaluation practices in Tyler's model are actually employed by teachers in the classroom setting (p. 96). They frequently focus upon the assignment of grades or scores to students. Finn suggested that educational evaluation should be broader than collecting information

from the students, but all types of information concerned with educational enterprise should be considered also.

The formative evaluation that was proposed by Scriven (1967) is playing an important role in improving courses and educational programs (Cronbach, 1963; Sorenson, 1971). Sorenson (1971) stated eight principles in formative evaluation which are helpful to evaluators. Sanders and Cunningham (1973, 1974) provided some ideas in the framework of formative evaluation. They suggested four types of formative evaluation activities:

1. Predevelopmental Activities -formative evaluation work that occurs before formal product development has started. Formative evaluation tasks related to the evaluation of needs, tasks, or other planning activities fall into this category. The techniques that can be used in these activities are sampling techniques, Q-sort, and task analysis.

2. Evaluation of Objectives Activities -formative evaluation work directed to judging objectives in product development. The emphasis of work is on the provision of reliable information about the worth of goal statements produced by the product developer. The techniques that can be used in these activities are questionnaires, delphi technique, and content analysis.

3. Formative Interim Evaluation Activities -formative work dealing with the appraisal of early product development efforts. Formative evaluation activities in this category are interim payoff evaluation work, interim intrinsic evaluation work, and the evaluation of program or project operations. The techniques and procedures that can be used in these activities are descriptive information, content analysis, critical appraisal, and student tryout.

4. Formative Product Evaluation Activities -formative evaluation work that focuses on the appraisal of a finished draft of the proposed product. Strategies such as validation studies, cost analysis, descriptive analysis, and goal-free evaluation can be used in these activities.

Sanders and Cunningham (1973, 1974) also suggested three sources of information that the evaluator should consider. Those sources of information are:

1. Internal Information -information that could be generated by inspecting the product itself.
2. External Information -information concerning the effects of the product or its components on the behaviors of students, teachers, parents, and other relevant groups directly involved in the use of the materials.
3. Contextual Information -information concerning the conditions under which the materials are expected to function.

Sorenson (1971) pointed out some advantages in doing formative evaluation. He said:

The advantages to be accrued from formative evaluation seem clear. Rather than waiting to find out at the end of a program whether it has been successful by running an outcome evaluation, it is more useful to direct that effort toward improving the program by testing and refining it while it is still under development. In addition to the greater economy of effort and time offered by this approach is the increased quality and effectiveness of the instructional program. (p. 17)

Scriven (1967) proposed the idea of summative evaluation which is mostly used in accountability programs, counseling programs, educational programs, social programs, and many more. The idea of summative evaluation is to assess the overall effectiveness of a program. The design of evaluation studies in the summative evaluation is similar to the design of research studies. Then the concept of evaluation and research should be distinguished. Sommer (1977) distinguished between evaluation and research according to the following:

1. Although research and evaluation complement one another, it is important to realize that the generation of new knowledge (research) does not require judgmental statements and that evaluation does not require the generation of new knowledge so much as it does a quantification and costing of factors already identified as important and relevant.
2. The problems in research tend to be self-generated by the researcher and there is a premium (sometimes a requirement) for originality. Problems in evaluation come from the outside and there is no encouragement for originality.
3. Researchers tend to begin their studies with hypotheses, evaluation avoid hypotheses as a matter of policy. Instead, the evaluation team uses questions stated as matters of degree.
4. Researchers tend to work in their own space; evaluation teams operate in the field mostly in other people's areas.
5. Deadlines in evaluation are natural, critical, and are frequently contractual. Evaluation is intended to guide policy and if a report comes in late, the study will be meaningless. Research deadlines are external and arbitrary.

Many evaluators such as Airasian (1974), Cook, Cook, and Mark (1977), Riecken and Boruch (1974), and Weiss (1972b) suggested that quasi-experimental designs (Campbell & Stanley, 1963) are the best designs to be used in program evaluation. Quasi-experimental designs control for both internal and external validity. Internal validity is the basic minimum without which any experiment is uninterpretable. External validity is concerned with generalizability. Internal validity is emphasized in evaluation.

Collecting data and using statistics to analyze the data are important tasks in summative evaluation. Randomization, matching group, analysis of variance, t-test, and multivariate analysis are popular techniques that have been used in evaluation. The summative evaluator should consider the techniques of presenting the data. Cost-benefit analysis also should be conducted in summative evaluation (Weiss, 1972, pp. 60-91).

Draayer (1974) reported the use of the Countenance model to evaluate the Experience Teacher Fellowship Program in Economic Education, 1969-1970. He claimed that program evaluation based on the Countenance model contributes to accountability and credibility by providing the insights and rationale for making modifications during the operation of the program and planning future programs.

The Educational Products Information Exchange (EPIE) expanded the Countenance model to increase information yield about new educational products (Borich, 1971). EPIE expanded the Countenance model to deal with internal interaction of product components by cross-analyzing data in three matrices: intents, observations and judgment of the producer, analyst and

user across various conditions of product use. Simple extension of the model will provide a worthwhile increase in the amount of information EPIE can offer users.

In the context of curriculum evaluation, the Countenance model was criticized by Westbury (Westbury, 1970). He said the suggestions from the Countenance model are not close enough to curricular phenomena to be immediately helpful; they do not direct an evaluator precisely enough to the phenomena he is supposed to look at (p. 251). From this point of view, another system approach to curriculum evaluation has been developed by combining the concepts and techniques from the CSE model and the Countenance model by Lutterodt in 1975 (Lutterodt, 1975).

The CIPP model was revised by the Phi Delta Kappa Study Committee on Evaluation in 1971 (Stufflebeam et al., 1971). The purpose of revising the CIPP model was to help educational practitioners improve their evaluation efforts. The Study Committee claimed that the revised model is an extension of the original CIPP model. Gephart (1972) pointed out that one of the major points in the extension was the recognition that all four types of evaluation (context, input, process, and product) do not go on all the time in all types of educational programs. The type of evaluation depends on the stage of planning decision. If a planning decision is made that calls for a change then input, process, and product evaluation are needed. Otherwise these are not needed at all.

The PDK report (CIPP model) was heavily criticized by Scriven (1972a). He said:

...to make a much tougher and more pedagogical claim about CIPP analysis, it seems to be about the most



complicated and confusing way of analyzing the practical procedures of evaluation that I can imagine, and it is certainly the most complicated one that I have ever seen. (p. 136)

Scriven claimed there were many mistakes in the PDK report (Scriven, 1972a, p. 133). The PDK report stated that the purpose of evaluation was not to prove but to improve. Scriven (1972a) said that this idea was mistaken (p. 133). The evaluation process should not stick to the concept of formative evaluation. He said that evaluation suffered for a long time from being regarded as simply summative, but we did not have to swing so far over as to say that it is never summative (p. 133). Another mistake was in the concept of product evaluation which turned out to be both summative and formative evaluation. Scriven pointed out that the actual process of evaluation--the nature of evaluation, in one sense --is usually the same in both cases. But the role evaluation plays and, in a sense, the kind of entity evaluated is different. The feedback from formative evaluation is to the project director or whoever is in charge of the project. In summative evaluation the feedback is to a consumer (p. 135).

Randall (1969) tried to simplify the CIPP model and made it simple enough to be applied to any program evaluation. He suggested some tasks that should be done in each type of decision (planning, structuring, implementing, and recycling), the kinds of information that should be derived from each type of evaluation (context, design, process, and product), and also the source of information. He emphasized timing and relevant information. He said the best information is of utterly no use if it does not arrive in time to base a decision on it.

Findlay (1971) applied the CIPP model to the Center for Vocational and Technical Education at Ohio State University. He showed that the CIPP model could be applied with a high degree of relevance to the Center's complex program and organization (p. 43). In setting a center-wide evaluation system, Findlay said there were four major requirements that have to be met by the system and the CIPP model can help to satisfy those requirements. First, the system must make possible on a continuing basis the identification of high priority problems in vocational and technical education. Second, it should permit selection of the most effective combination, within the Center's available resources, of activities for coping with high priority problems. Third, it should enable the Center to detect and revise or eliminate ineffective activities. Fourth, it should include procedures for assessing the actual impact of the Center's program on target needs in vocational and technical education (p. 43). Findlay (1971) also said the general effect of applying the CIPP model approach is not to change the Center's program structure but to strengthen it by making evaluation a routine, explicit process with high visibility (p. 45).

Merriman (1970) divided the process of evaluation into two sections. Context evaluation is the first section, which is called a system approach. Input, Process, and Product evaluation is the second section, called a task-problem centered approach. He pointed out that the context evaluation which serves planning decisions is concerned with the entire system. The planned solution strategy has to be selected first, then input, process, and product evaluation can be processed (p. 54). He also suggested that an evaluation department should be composed of four teams.

Context Evaluation Team. This team should include personnel who have the training to work with the variables which will be included in the system, as well as a system analyst and data processor.

Input, Process, and Product Evaluation Team. This team should include personnel with training in evaluation and research as well as an understanding of a school system.

The Direct Service Team. This team may be formed in the event of the need for extensive, continuous needs for input, process, and product information relevant to a particular problem area.

Supportive Services Team. This team will provide the service in clerical, typing, and editing activities (pp. 55-57).

The CSE model had been revised by Klein et al. in 1971, but the basic concept remained the same (Klein et al., 1971). There are some changes in the model, such as the first step of doing evaluation. The first step was system assessment, but it has been changed to needs assessment. More flow charts were added containing the tasks that should be done in each step of the evaluation. Those changes are helpful to the evaluator in setting program evaluation.

Goal-Free Evaluation (GFE) (Scriven, 1972b) became one of the most controversial techniques in the field of evaluation. Stufflebeam (1972) seemed to agree with the general idea of GFE, but he was not so sure that GFE could be done. He said it is questionable that GFE can or should be goal-free. The strategy is potentially useful, but far from operational and replicable (p. 5). Alkin (1972) did not agree with the concept of GFE. He stated that if goals are not clear why not change the procedure of determining the goal (p. 6). Condemning the Goal-Base Evaluation (GBE) procedure

because of inadequacies in its execution does not solve the problem. Performing a better job of GBE does offer some hope (Alkin, 1972, p. 6). Popham (1972) stated that the possibility of keeping GFE completely uncontaminated by goal preferences is unrealistic (p. 7). Kneller (1972) suggested two alternative directions in which the side-effects of the project can be observed without using GFE. He said first, evaluators should be trained to observe both goals (and outcomes) and side-effects; second, the researchers should specify as many likely side-effects as possible within the original research design (p. 7).

#### Summary

The concept of evaluation was changed from product oriented to product and process oriented. The product oriented is called summative evaluation and the process oriented is called formative evaluation. The purpose of summative evaluation is to assess the overall program to see whether or not the program has been successful. The information taken from the summative evaluation is used to decide whether the program should be improved, modified, or terminated. The formative evaluation is used to provide information for the on-going program which helps the program planner to improve the program while it is in progress. In order to be sure that the results from both the summative and the formative evaluations are valid and reliable, the strategies from research should be applied to evaluation studies. Internal and external validity should be well controlled. Quasi-experimental designs should be used in the evaluation study. Evaluation and research use the same strategies. Because of this the concepts of evaluation and research should be distinguished. Evaluation should not emphasize only the intended outcomes but side-effects of the

program also. The scope of the evaluation study should include a cost-benefit analysis.

There are many contemporary evaluation models available today. The CIPP model, CSE model, Countenance model, Tyler model, and Goal-Free Evaluation are the most prominent evaluation models that have been applied in a variety of programs throughout the country.

## CHAPTER V

### DESIGN OF THE STUDY

Before an evaluation model is designed, the purpose of evaluation should be stated. The purpose of evaluation in this study is to provide relevant information to decision-makers so they can determine how and/or whether to improve, maintain, or terminate a course or program.

In the process of developing an evaluation model, the conceptual frameworks of Stufflebeam (1968), Alkin (1969), Stake (1967), Scriven (1967), and Tyler (1942) were used as guidelines.

There were three criteria used to develop the model:

1. It is precise, which means that elaborate forms of measurement are usually devised to describe the phenomena of interest.

2. It is specific, which means that the model deals with only a select number of phenomena, purposefully avoiding complex summary or aggregate characteristics of an event.

3. It is verifiable, which means that hypotheses are posed to check the precision of the model, i.e., empirical evidence is accumulated that eventually determines the model's accuracy and usefulness.

### General Evaluation Model (GEM)

There are four stages in the basic general evaluation models that are proposed in this study. All of the four stages are basic requirements in most of the evaluation models today. These four stages are:

1. Setting Goals and Objectives. Goals and objectives must be set before any program can be set. Without goals and objectives, the effectiveness of an educational program is hard to define.

2. Planning the Program. The program planners have to set the program according to the goals and objectives to make sure that those goals and objectives will be reached.

3. Construction of Measuring Instrument. After the program has been set, selecting and developing the measurement instrument will be done. The validity and reliability of measurements must be determined.

4. Collecting and Analyzing of Data. After the data have been collected and analyzed, the outcome should be judged against the goals and objectives of the program. The outcome will provide information regarding the effectiveness of the program.

All of the four stages above are expanded and implemented into the College of Education Model.

### College of Education Model (CEM)

The model is composed of seven stages. Each stage has certain requirements that have to be met. The seven stages are:

1. Needs Assessment

- a. Definition of terms.

- b. What types of information should be collected?

- c. Who should be the respondents?
- d. Who should conduct needs assessment?
- e. How needs assessment could be conducted?
- f. What should be done with information once it has been collected?

2. Setting Goals and Objectives

- a. Definition of terms.
- b. Who should make decision on goals and objectives?
- c. How goals and objectives should be set?
- d. What budget, personnel, and facilities factors should be considered in setting goals and objectives?

3. Program Planning

- a. Definition of terms.
- b. Who should plan the program?
- c. How should the program be planned?
- d. How can the program planners be certain that the program will reach the goals?
- e. What are the roles of the evaluators and the administrators in program planning?

4. Program Implementation

- a. Definition of terms.
- b. How should the program be implemented in the real setting?
- c. What types of information should be collected in this stages?
- d. How should it be collected?
- e. What are the roles of the evaluators and the administrators



in program implementation?

5. Program Improvement

- a. Definition of terms.
- b. What types of information should be collected in this stage?
- c. How should it be collected?
- d. Who should collect this information?
- e. What should be done with the information once it has been collected?
- f. What are the roles of the evaluators and the administrators in program improvement?

6. Summative Evaluation

- a. Definition of terms.
- b. What types of information should be collected in this stage?
- c. How should it be collected?
- d. Who should collect this information?
- e. What types of standards should be set in order to make judgments?
- f. What are the roles of the evaluators and the administrators in summative evaluation?

7. Follow-Up Study

- a. Definition of terms.
- b. What types of information should be collected?
- c. How should it be collected?
- d. Who should be the respondents?

- e. Who should conduct follow-up study?
- f. What should be done with information once it has been collected?

In order to develop the proposed model, an extensive review of the literatures dealing with program evaluations and evaluation models, especially as related to teacher preparation programs was made. The methodologies used to assess results of the various programs were identified and adjusted to the College of Education Model. All the literature was searched by using ERIC search, books in print, and Government documentation concerned with program evaluations.

## CHAPTER VI

### GENERAL EVALUATION MODEL

This chapter and the following chapter will provide the necessary details to answer the question which was stated in Chapter I:

Can an evaluation model be devised to be used with a teacher preparation program?

To answer the question in the affirmative, two evaluation models were developed. The framework from which these two models was drawn and synthesized consists of the previous studies on the evaluation models, evaluation studies, selecting and developing measuring instruments, research designs, teacher preparation programs, and related fields. The evaluation models were based on three criteria. They have to be precise, specific, and verifiable.

An extensive search for information had been done through the GIPSY (General Information Processing System) which provided the documents printed in both CIJE (Current Index of Journal of Education) and ERIC (Education Resources Information Center): Research in Education, Books in Print and Government reports also had been used.

In order to be sure that the models were valid, the concepts, techniques, strategies, and measuring instruments presented in the models were selected from reliable research works and outstanding articles in the area of evaluation which were written by prominent evaluators such as Scriven,

Stake, Stufflebeam, Tyler, Airasian, Alkin, and others. The research works were reliable if they were well determined in design, sampling, and data collection (Wechsler et al., 1976). The design should control for internal and external validity. Sampling should be done by systematic randomization or at least by matching. The data should be collected by valid and reliable measuring instruments.

#### Program Planning Committee (PPC)

Before any evaluation plan is implemented into the teacher preparation programs, a Program Planning Committee (PPC) would be necessary. The PPC will be responsible for reporting the effectiveness of the program, developing, defining and/or reviewing goal statements, planning or modifying the program, and evaluating the program. The PPC is also responsible for providing information to the staff if there are any changes in the program. If seminars or workshops are necessary in implementing the new program, the PPC will be responsible for that too. The PPC should work directly with the dean of the college, which will help to eliminate any delay in the planning and evaluating processes.

King et al. (1974) suggested that the PPC should be relatively small, twelve to fifteen members, including representatives from administration, faculty, students, alumni, in-service teachers, the superintendent or principal, and an in-house evaluator (an in-house evaluator is the person who is on the staff or faculty in the college). These groups should be represented because they are the groups of people who have been affected or will be affected directly or indirectly by the programs provided in the college. The in-house evaluator is necessary for the PPC because he has responsibility for assessing and providing information for planning and

modifying the program. The in-house evaluator should come from the faculty or the staff of the college because he will be supportive, nonthreatening, dedicated to producing a success, and tolerant of ambiguity of objectives and curriculum development procedures (Scriven, 1967). Scriven (1967) also pointed out that the out-side evaluator (out-side evaluator is the professional evaluator who is hired to conduct summative program evaluation in the college) should not be part of the committee because he may "dampen the creative fires of a productive group," slow down the development process by urging that objectives be clarified, or lose his independence. At least one of the faculty who joined the committee also should be a curriculum development expert. The chairperson of the committee should represent the college's administration.

#### General Evaluation Model (GEM)

The General Evaluation Model (GEM) is composed of four stages, (see Figure 2). These four stages are the basic fundamentals in any evaluation processes. Most of the evaluation models available today also have these four stages as their foundation in developing their models.

Setting Goals and Objectives. Goals and objectives are required in most of the planning and evaluating processes. Goals are the desired outcomes and are usually stated in broad terms. Objectives are the description of the performance that the learners should be able to exhibit before they are considered competent (Mager, 1975). Objectives should be stated in behavioral terms because they can be observed and measured (Mager, 1975; Tyler, 1942; Popham, 1973; Gagne', 1965). Objectives can be classified as subsets of goals.

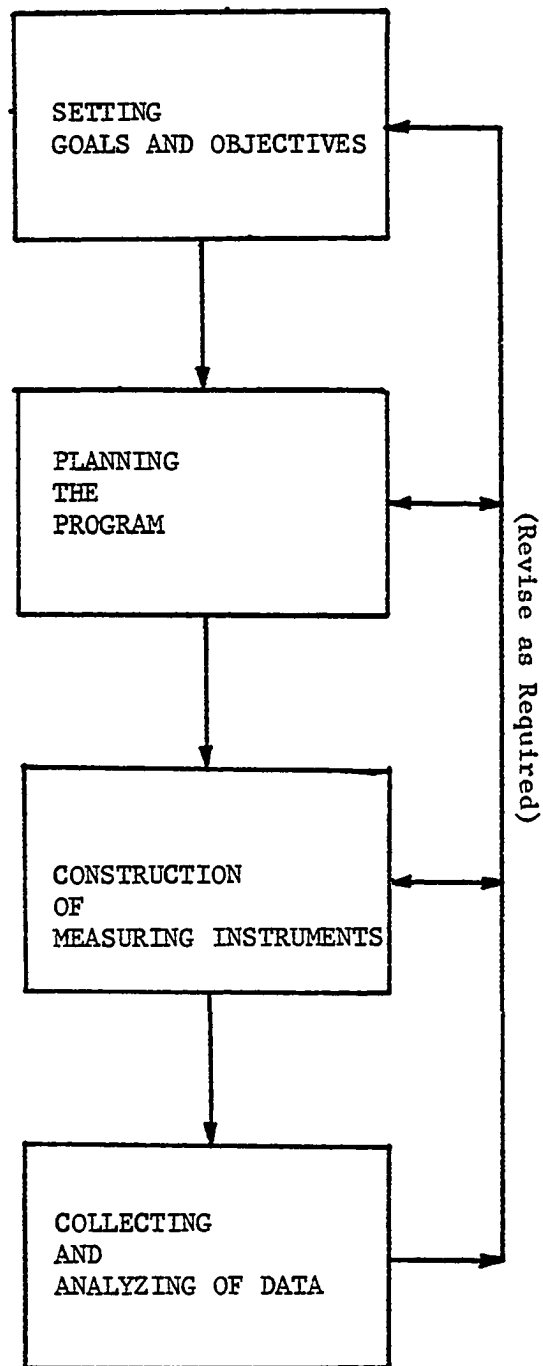


Figure 2: General Evaluation Model.

Goals should be set according to the needs of the groups of people (the community members, the faculty, and the students). The needs of the groups of people can be assessed by using needs assessment. Needs assessment is the process of defining the discrepancy between "what is" and "what should be" (Kaufman, 1972; Berrie, 1976; Popham, 1972a). The results from needs assessment will show the need areas perceived by the community members, the faculty, and the students. The varied sources make it difficult to rank these need areas according to their importance. In order to eliminate the ranking problem, the Delphi Technique (Uhl, 1971; Cyphert & Gant, 1970), Q-sort (Stephenson, 1953; Whiting, 1959; Downey, 1960), rating, and money game (Mullen, 1974) can be used. The Delphi Technique itself also can be used to collect information on what the groups of people want the program to be.

The evaluator will provide all the information on the need areas perceived by the groups of people and also should provide the idea of how the statements of goal should be set. Goals Analysis, written by Mager (1972), is a useful instrument in setting and analyzing goals. For writing objectives, the concept from Bloom's Taxonomy (Bloom, 1956; Krathwohl et al., 1964) and the learning structure analysis of Gagne' (1965) are helpful.

Planning the Program. After the program goals have been set, the program can be planned. The problem in the planning stage is concerned with an uncertain future which makes the planning process difficult. What the PPC can do is to forecast or predict the likelihood of the outcomes that should occur when certain instructional methods or techniques are implemented into the program. The evaluator will play the role of a

detective in which he has to find out whether or not the instructional methods or techniques that were proposed in the program planning have been studied or used before. He also has to find out if there are any programs that are similar to the program planned which have been implemented in other institutions before. As the detective, he has to know the sources from which he can collect information. Today much information is stored in computers. GIPSY is one of the most efficient and economical systems that provide information from CLJE and ERIC. Other sources that the evaluator can search for information are government agencies (such as Health, Education, and Welfare, and State Departments of Education); books in print, and persons and agencies that collect and process descriptive data.

In planning the strategies for implementing and developing the program, PERT (Program Evaluation and Review Technique) (Miller, 1963), PPBS (Planning Programming-Budgeting System) (McCullough, 1966; Hovey, 1970) and system analysis (Cleland & King, 1968; Kaufman, 1972) are helpful instruments.

In the process of program planning, the financial situation has to be considered. The evaluator should compare the cost of operating the instructional methods or techniques with the effect of the instructional methods or techniques. The instructional methods or techniques that yield the higher effect with the lower cost should be considered for use in the program. At the same time, the evaluator should assess the resources that are available in both the college and the community as well.

Before the evaluator recommends the instructional methods or techniques that should be used in the program, he has to assess the reliability of the instructional methods or techniques. The estimation of the relia-



bility of the instructional methods or techniques can be done by letting the experts in the area predict the possibilities of each instructional method or technique. There are two methods that can be used to obtain the prediction from the experts. First, using a confrontation technique by inviting all of them together and letting them discuss the possibility of each of the instructional methods or techniques. Second, using the Delphi Technique, the evaluator will construct the questionnaires on the instructional methods or techniques and send them to all the experts and let them predict the possibilities of each instructional method or technique. By using the Delphi Technique, the evaluator eliminates the problem of domination by a single person often found in the confrontation method.

Construction of Measuring Instruments. Before the program will be implemented into the real setting, the evaluator has to find out what measuring instruments should be used. Interviews, questionnaires, observations, psychometric tests (attitudes, values, personality, preferences), teacher made tests, standardized achievement tests, etc., can be used to assess the effectiveness of the program. Measuring instruments should be selected according to these three criteria: reliability, validity, and economy. Reliability is concerned with the consistency of the test. The test will be reliable if it is administered to two or more comparable groups of students and yields similar results. Validity will answer the question of do the tests measure what we want them to measure. If the tests do not measure what we want them to measure, the results can be misleading. After the evaluator is satisfied with the reliability and the validity of the tests, he has to consider the cost of each of the tests before he can select any of them. Sometimes the cost of the tests is the priority criterion.

Collecting and Analyzing Data. After the program has been implemented, the data have to be collected. There are two purposes in collecting the data. First of all, the data are collected while the program is in progress for the purpose of improving the program, which is called formative evaluation. Another purpose is to find out the effectiveness of the overall program, which is called summative evaluation (Scriven, 1967; Bloom et al., 1971; Stufflebeam, 1971). Most of the evaluators suggest that true experimental designs and quasi-experimental designs should be used in the summative evaluation because both of them control well for internal and external validity. In the evaluation process, internal validity is most important because it is concerned with the interpretability of the data. The statistical techniques will be selected according to the type of designs and the data. T-test, analysis of variance, analysis of covariance, factor analysis, and regression analysis are among the most popular statistical techniques that have been used in evaluation studies recently.

Scriven (1972b) suggested that Goal-Free Evaluation (GFE) should be used in the summative evaluation. He thinks that the evaluator should know less about the program goals. Scriven is concerned that the more the evaluator knows about the program goals, it is likely that he will develop tunnel-vision upon those goals and will forget about the side-effects of the program. These side-effects are sometimes more important than the program goals themselves (p. 2). Alkin (1972), Popham (1972b), and Kneller (1972) do not think that GFE can be or should be used in the summative evaluation, but agree with Scriven that side-effects of the program should get attention also.

Because the results from summative evaluation have an effect on the decision of whether the program should be improved, modified, or terminated, the report on the summative evaluation should be as free from bias as possible. In order to eliminate the bias in the report, most of the evaluators suggest that summative evaluation should be conducted by professional evaluators from outside the institution.

There are some controversies in the process of formative evaluation. Scriven (1967) suggests that formative evaluation should be conducted as the same technique as summative evaluation. Cronbach (1963) and Stake (1967) do not believe that comparison between the experimental group and the control group or comparable groups should be done in the formative evaluation. They point out that the question of the formative evaluation is whether the program has an effect on the students performance not whether or not the program has been successful. They suggest that the outcomes of the program should be compared with the absolute standard which is set by the PPC.

Formative evaluation will be conducted while the program is in progress. If the outcomes do not meet the standard, the program should be modified or changed. The decision of changing or modifying the program should be done immediately after finding out that the program does not accomplish the standard. In both summative and formative evaluation, the evaluator should collect as much information as he can. This information should deal with both the goals and the side-effects of the program. This information will provide an indication of the side-effects of the program as well as the success of the program.

After the data have been collected and analyzed, the report should be written. The report should be written in simple form and presented in both the area of the programs success and the area of the programs failure. The judgement of the outcomes and the recommendation should be included in the report. The report should be presented on time.

## CHAPTER VII

### COLLEGE OF EDUCATION MODEL

The College of Education Model (CEM) was developed by expanding the concepts from GEM so it can be applied to the teacher preparation programs. CEM is composed of seven stages which are as follows: needs assessment, setting goals and objectives, program planning, program implementation, program improvement, summative evaluation, and follow-up study. (see Figure 3)

#### Stage 1: Needs Assessment.

Most evaluators define educational need as a measure of the discrepancy between current outcomes and desired outcomes (Kaufman, 1972; Grotelueschen & Gooler, 1972; Berrie, 1976; Popham, 1972a; Yuskiewicz, 1975). Therefore, needs assessment in education is the process of assessing or determining the extent of the discrepancies that exist in educational operation as related to student outcomes.

Center for Community Needs Assessment (1973a) stated that:

Needs assessment can be a tool to help education be more responsive to the needs of the citizenry-- showing up discrepancies based on facts rather than guesswork. It is a tool so much better than

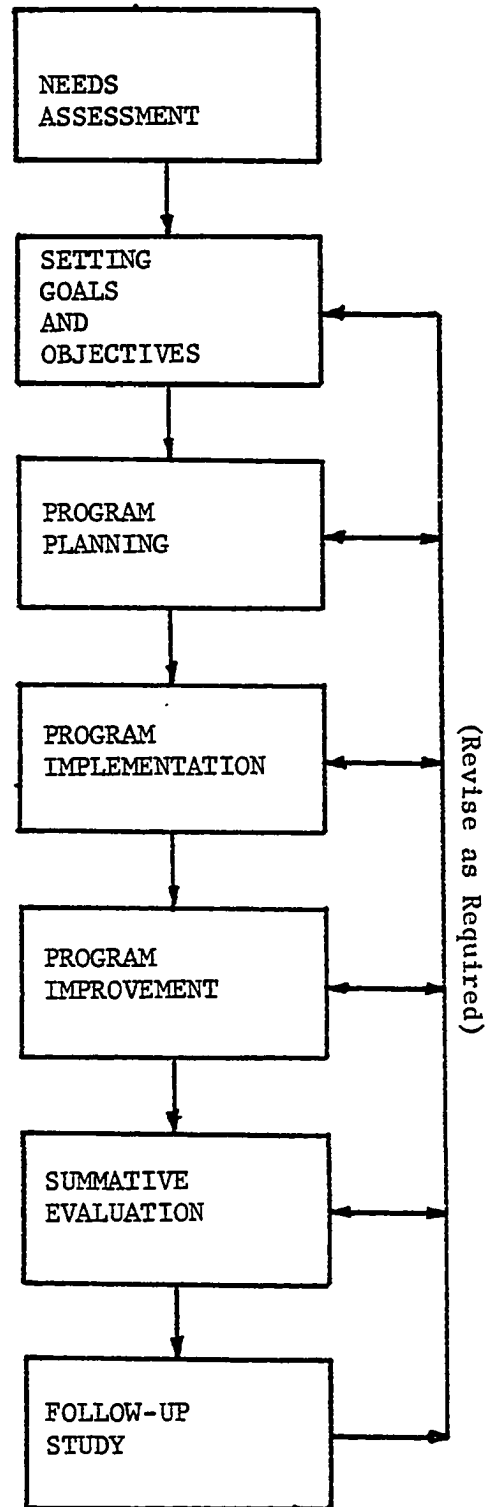


Figure 3: College of Education Model.

anything we have now and can be revolutionary in bringing about positive change for the most important human process of all--the education of man. It can be the vehicle by which the formal system moves from its existing state to what it should be. (p. 24)

Needs assessment can be used as a tool in determining program goals and objectives. Baker (1972) defines it as the appraisal of the operation of a system to determine what program goals should be. Needs assessment is also used to decide which of the program goals and objectives are of highest priority and need to be accomplished in given time (Klein et al., 1971; Popham, 1972a).

Needs assessment also can be used as a tool in planning and developing educational programs (Rookey, 1975; Berrie, 1976; Coffing & Hutchinson, 1974). Sweigert (1968) perceives needs assessment as a tool to provide base-line data for use in measuring the progress made in improving the efficiency and the effectiveness of the educational enterprise. Fitzgerald (1972) and Fine (1969) also perceive it as a tool in evaluating the effects of the program on the learner.

Because of the accountability movement and the awareness of the college toward the needs of the community, needs assessment has become an essential part in planning an educational program. There are many colleges and universities that have collected information from needs assessment and use it as a data base for planning and improving educational programs. Governor State University conducted an extensive needs assessment by using more than 1,000 respondents from government,

education, business, industry, and the arts. These respondents were chosen from the local and national levels. The purpose of the study was to collect information for setting educational objectives and developing the educational programs in the university (Norton, 1970). The English Department of the University of California, Los Angeles, conducted a needs assessment to develop a new freshmen composition program (Spooner-Smith, 1976). In order to evaluate the teacher education program in the college, Lehman College decided to conduct a needs assessment (Frankel, 1974). The Center for Community Needs Assessment (1973b) surveyed the needs of the community for seven community colleges in Florida. The information collected was given to the college administrators to use for planning an effective curriculum. Bakersfield College (California) also surveyed the needs of the community within a three mile radius of the college and found out that vocational education should be the priority goal of the college (Larsen, 1973). For more needs assessment programs see Galvin et al. (1975), Skelhorne (1975), Moore (1975), McClurg (1975), and Niemi (1976).

The needs assessment must have at least three characteristics (Kaufman, 1972).

1. The data must represent the actual world of learners and related people, both as it exists now and as it will, could, and should exist in the future.

2. No needs determination is final and complete; we must realize that any statement of needs is in fact tentative, and we should constantly question the validity of our needs statements.

3. The discrepancies should be identified in terms of products



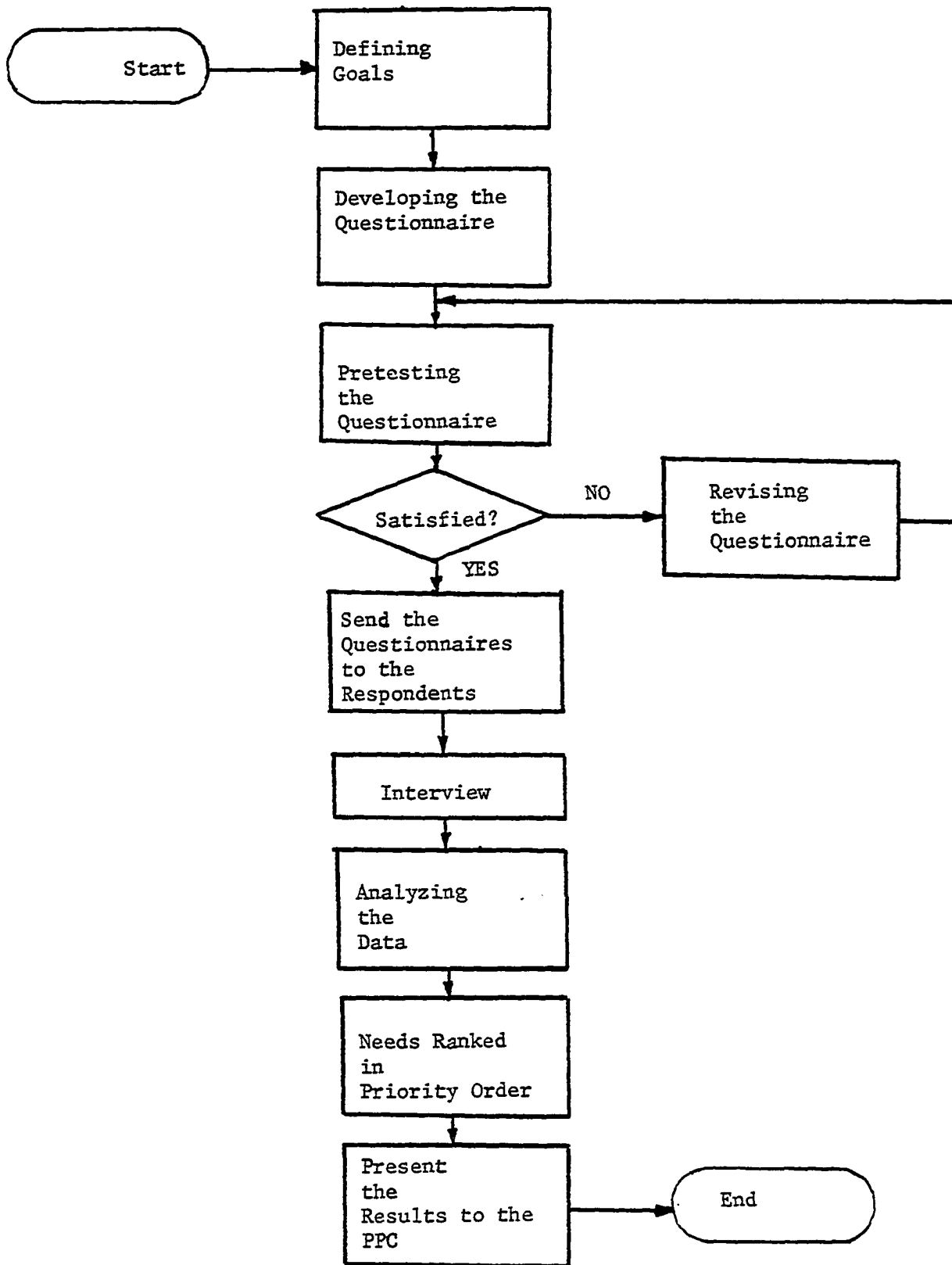


Figure 4: Needs assessment processes.

or actual behaviors (ends), not in terms of process (or means) (p. 29).

How needs assessment should be conducted is shown in the flow-chart in Figure 4.

Defining Goals. This is the first step in doing needs assessment. It seems to be the most difficult step in assessing the needs. Actually, most of the Colleges of Education in the Southwest area already established their goals, but their goals are often broad, hazy, ambiguous, and hard to specify. Then, the goals should be defined and narrowed down into the scopes and areas that the PPC are interesting in. In the process of defining goals, the concept that the PPC should be aware of is that goal statements have to be clear, specific, and measurable (Weiss, 1972b). The process of defining goals should be under the supervision of an in-house evaluator.

Developing the Questionnaires. When the PPC approves the selected goal areas, the questionnaires will be developed. The questionnaire is the most appropriate instrument, because it is relatively inexpensive, easily administered, and can be given to a large number of respondents simultaneously. In the process of construction of the questionnaires, the evaluator has to decide whether he is going to use closed or open-ended questions. Closed questions limit the respondents to answer the specific issues, while open-ended questions give more freedom to the respondents in answering the questions. Open-ended questions are very useful when the relevant dimensions are not known, or when the interest of the PPC lie in the exploration of a process or of the individual's formulation of an issue. Most of the questionnaires developed recently have both types of questions.

## Needs Assessment Quesstionnaire

- 5 - Always
- 4 - Often
- 3 - Sometimes
- 2 - Rarely
- 1 - Never
- 0 - Do not Know

.....

Should Exist	To what extent do you think the College of Education, OU should provide each student to:	Actually Exists
_____ A.	construct a lesson plan.	_____
_____ B.	prepare directions for a substitute teacher.	_____
_____ C.	teach a lesson effectively.	_____
_____ D.	establish criteria for evaluation of lessons, units or courses.	_____
_____ E.	assist students to develop study habits.	_____
etc.	etc.	etc.

---

Figure 5: Format of the questionnaire.

Figure 5 shows the format of the questionnaire that is widely used in conducting needs assessment.

From the format of the questionnaire in Figure 5, the respondents have to answer both "should exist" and "actually exist." The information from the "actually exist" side will help the PPC know the perception of the respondents about the existing program in the college. This information can be used as a criterion reference in program evaluation. The information from the "should exist" side will help the PPC know what the respondents would like to see the college provide in the educational program. The discrepancies between the score of "should exist" and the score of "actually exist" is the "index of need." This "index of need" will show which goals should be set, improved, or terminated.

Another instrument that can be used in conducting needs assessment is the Delphi Technique. Governor State University used this technique to obtain the data on what goals toward which the university should operate in the future (Norton, 1970). The disadvantage of using the Delphi Technique is the PPC will obtain only the data that the school should provide in the future. This technique lacks the data on how the respondents actually perceive the educational program as it exists.

Pretesting the Questionnaires. Pretesting the questionnaires should be conducted after the tentative questionnaires have been completed. The pretest will provide a means of catching and solving unforeseen problems in the administration of the questionnaires, such as the phrasing and sequence of questions, or its length. It may also indicate the need for additional questions or the elimination of others (Sellitz et al., 1976). In conducting pretest, Parten (1950) suggests that using only 20 to 30

respondents should be enough. If the results are satisfactory, the questionnaires will be ready to administer in the real setting. If the results are not satisfactory, the questionnaires should be revised again and again until they are satisfactory.

Collecting the Data. There are two steps that need to be taken in the process of collecting the data. First of all, the questionnaires have to be sent out to all the respondents and after the questionnaires have been returned, the process of an interview should be followed. The interview technique is used to enhance full and accurate data collection.

The problem that the evaluator has to face in the first step is concerned with defining the population and sample. In most needs assessment studies, three segments of the population have always been used. There are students, staff and faculty, and community members (in-service teachers, principals or superintendents, members of local boards of education, parents, and alumni) (Kaufman, 1972; Fine, 1969; Yuskiewicz, 1975; Center for Community Needs Assessment, 1973b). The evaluator has to decide how large a sample he wants to use in the process of data collection. Defining the size of sample is a crucial point, because it is concerned with internal validity. Usually the size of the sample is decided upon the area that the college serves most. In the areas that have more than 500 teachers, an adequate random sampling size of 20% is recommended (Yuskiewicz, 1975).

Once the questionnaires have been returned, the evaluator has to find out what areas from which he wants to gain more information. Then the interview questions will be developed according to those areas. When the process of interviewing has been completed, analyzing the data is the next step.

---

	Should Exist	- Actually Exist	= Index of Need
Goal B	4.9	1.2	3.7
Goal A	4.5	2.1	2.4
Goal C	3.6	2.3	1.3
Goal D	2.5	2.5	0.0
Goal E	2.1	3.5	- 1.4
etc.	etc.	etc.	etc.

---

Figure 6: Presenting data.

Analyzing the Data. The average score of "should exist" and "actually exist" of each goal should be calculated. The average score of "actually exist" minus the average score of "should exist" will show a discrepancy. The discrepancy will show the "index of need." Figure 6 shows how the data should be presented.

If the "index of need" equals zero, the goal is being achieved satisfactorily. If the "index of need" is negative, above standard performance on the goal is being made. A positive discrepancy would indicate that the goal is not being met and is, therefore, an educational need.

Needs Ranked in Priority Order. Needs will be ranked in priority order according to the degree of the "index of need." The need area that has higher score on the "index of need" will be ranked higher than the need area that has ranked a lower score on the "index of need." After the need areas have been ranked, the results will be presented to the PPC.

## Stage 2: Setting Goals and Objectives.

Goals are general statements of desired outcomes, which have long-range implications and which may involve complex human behaviors and aspirations (Hayman & Napier, 1975). Objectives on the other hand, are narrower, specific, and usually short-range. Objectives should be stated in behavioral terms which the student may perform in order to demonstrate achievement. They are thought to contribute to the envisioned final goals. The distinction between goals and objectives is a matter of convention, and the division point of these concepts is indistinct, because the difference between them is in degree rather than in kind (Kaufman, 1972; Hayman & Napier, 1975).

Goals lie at the heart of the educational planning process, whether one is planning a curriculum or a single classroom lesson (Davis, 1976; Dressel, 1976). Bloom et al. (1971) suggests that goals should be a starting point for developing the instruction and evaluation process. Goals are not only used for educational planning and evaluation process but for the basis of communication, understanding, and support from the public also. Peterson (1970) stated:

It seems essential in these time that colleges articulate their goals: to give direction to present and future work, to provide an ideology that can nurture internal cooperation, communication, and trust, to enable appraisal of the institution as a means end system, to afford a basis for public understanding and support. Indeed, the college without the inclination or will to define itself, to chart a course for itself, can look forward either to no future--to a kind of half-life of constantly responding to shifting pressures--or to a future laid down by some external authority. (p. 11)

Weiss (1972b) suggests that the statement of the goals should be: clear so that the evaluator knows what to look for; specific and able to be translated into operational terms; and made visible and measurable.

At this stage, the PPC and evaluator have to work together closely on setting and refining the program goals. The evaluator will take responsibility for clarifying the statement of program goals. Minimizing all the fuzziness of statements and modifying all the statements to the degree that their meaning is understood, or in other words to the degree that you



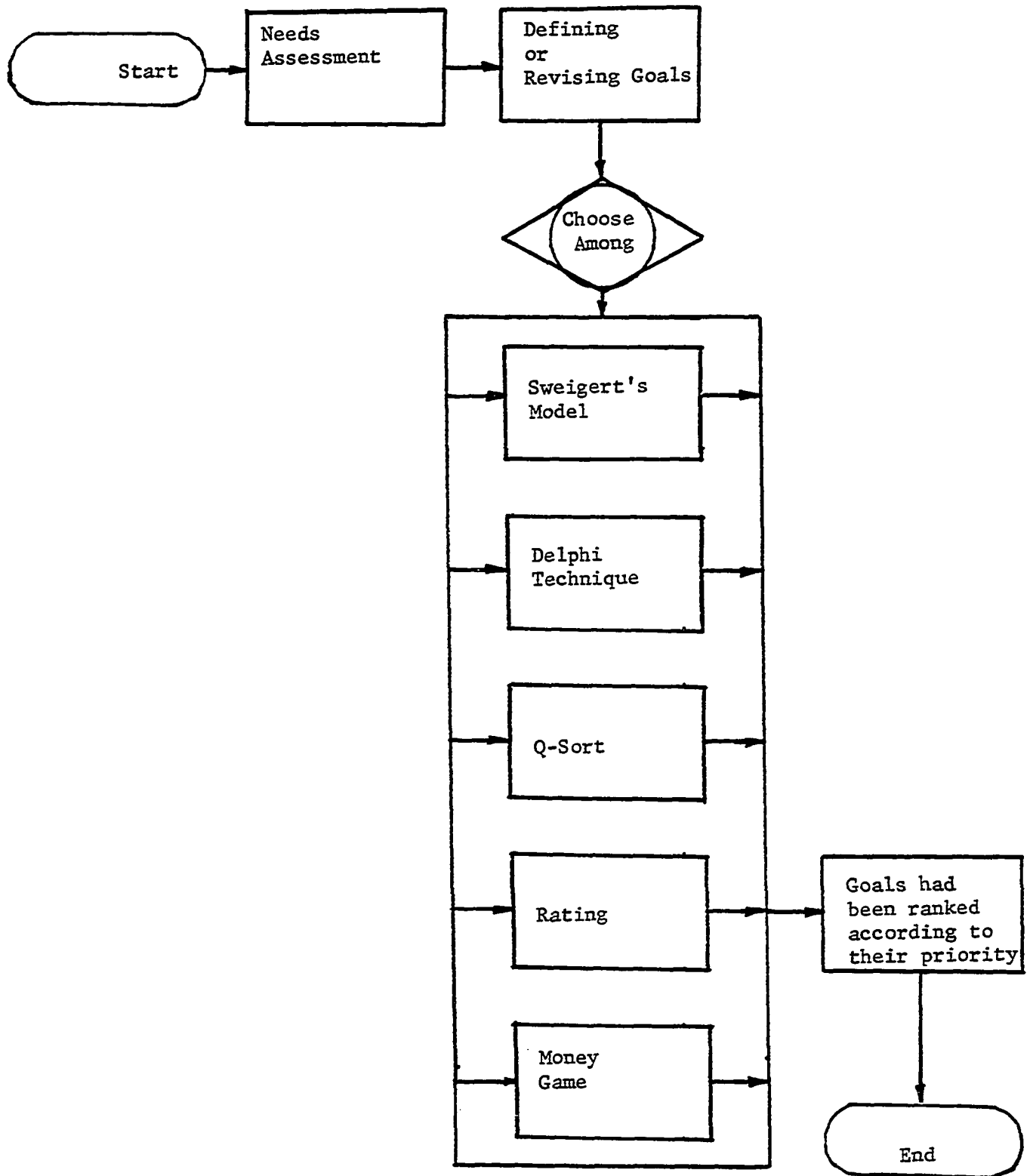


Figure 7: Setting and ranking goals process.

will know one when you see one (Mager, 1972). Mager wrote the book Goal Analysis which is helpful in clarifying program goals. He suggests five steps:

- Step 1: Write down the goal.
- Step 2: Jot down, in words and phrases, the performance that, if achieved, would cause you to agree that goal is achieved.
- Step 3: Sort out the jottings. Delete duplications and unwanted items. Repeat Step 1 and 2 for any remaining abstractions (fuzzies) considered important.
- Step 4: Write a complete statement for each performance, describing the nature, quality, or amount you will consider acceptable.
- Step 5: Test the statements with the question, if someone achieved or demonstrated each of these performances, would I be willing to say he has achieved the goals? When you can answer yes, the analysis is finished.

As Stake (1970) states a goal statement is a value judgement, and the process of setting goals by the administrators and the faculty within the college will represent the values of the college. They will not represent the values of the community. If the concept that the college is part of the community and is being established to serve the community still stands, then the values of the community should be a part of educational goals. Then, the input from the community should be sought. Figure 7 showed how the educational goals should be set.

There are at least 5 techniques that can be used to let the community members, the students, and the faculty participate in the process

of setting educational goals.

1. Needs assessment.
2. Delphi technique.
3. Q-Sort.
4. Rating.
5. Money game.

Needs Assessment. This technique has been presented in full detail in the first stage of the CEM. The information from needs assessment will show the degree of discrepancies between "what is" and "what should be" of the need areas. The need area that has the highest discrepancy will be considered as a priority goal for improvement. There is a problem in selecting the priority goal areas within the process of needs assessment because there are three groups of people participating in this process. If there are discrepancies in perception in the priority goal areas among the groups of people who participate in this process, then the question of which priority goal areas should be selected for the program planning is raised.

There are many techniques that can be used to select the priority goal areas from the results of needs assessment, such as Delphi Technique, Q-sort, rating, and money game. In the case that the PPC does not want the evaluator to do any further analysis, then Sweigert's Model (Sweigert, 1969) would be an appropriate alternative for selecting the priority goals areas. Sweigert suggests that:

1. If any goal areas are perceived as highest need by all three groups of the participants, then those goal areas should be considered as priority goals.

2. If any goal areas are perceived as highest need by the community members and by either the faculty or the students, then those goal areas should be considered as priority goals also.

3. If any goal areas are perceived as highest need by only the community members, then those goal areas should be kept for consideration in the future.

4. If any goal areas are not perceived as high need by the community members, then those goal areas should be dropped out from any consideration.

Delphi Technique. Delphi Technique was developed by RAND Cooperation. The objective of the Delphi Technique is to obtain a consensus among experts without bringing them together in a face-to-face meeting. This is achieved by having them complete a series of questionnaires in depth, interspersed with controlled opinion feedback.

Originally, the Delphi Technique was developed as a tool for forecasting the events of the future. The respondents were asked to focus on what is likely to happen. For example, during the fall of 1965, the University of California, Los Angeles used the Delphi Technique with twenty graduate business students to forecast the gross national product, defense expenditures, and fourteen other business indexes (Pfeiffer, 1968). Gordon and Helmer (1966) also used the Delphi technique to forecast the events related to scientific, population, and war that would occur from twenty to fifty years from now.

When the Delphi Technique is applied to the field of education, the basic concept is changed. The respondents are asked to focus on what they would like to see happen rather than what is likely to happen.

Cyphert and Gant (1970) applied this concept when they used the Delphi Technique to seek information on what should be prime targets for the School of Education, University of Virginia, to concentrate its energies and resources on in the next decade. Norton (1970) used this technique to find out the goals toward which the Governor State University should work.

An Institutional Goals Inventory (IGI) has been developed by Educational Testing Service, which combined the concepts of needs assessment and Delphi Technique together to make the results more reliable (Uhl, 1971).

The Delphi Technique can be applied to the CEM in the following manner.

1. List all the goal areas that are perceived as high need by the participants from the results of needs assessment.
2. The participants are then asked to evaluate the total list of the goal areas and list 3 or 5 goal areas according to their importance.
3. Each participant receives the list and a summary of responses to the goal areas, and is asked if he wants to revise his opinion.
4. Each participant again receives the list, an updated summary, and is asked for a final chance to revise his opinion.

By using the Delphi technique, the discrepancies in perception of the priority goal areas among the groups of the participants will be eliminated. The results from the Delphi technique will show a ranked list of the important goal areas that all the participants agree on.

Q-Sort. Q-sort is another techniques that can be used to reach an agreement on the discrepancies in perception. Q-sort was developed by

Stephenson (1953) and used for the study of the traits or characteristics of individual cases. Since then, Q-sort has been used in a variety of studies such as Revie (1956), Whiting (1959), and Downey (1960). Revie used Q-sort to describe both the teacher's and the school psychologist's concept of pupils. Whiting had nurses, patients, and physicians sort statements concerning the importance of various aspects of the nurse's work. Downey also used Q-sort to determine how sixteen educational goals were rated by a variety of people.

Q-sort can be applied to the CEM in the following manner:

1. Place the high need goal areas that are perceived by the participants from the results of needs assessment on the cards, one to a card.
2. Randomly order the cards and give them to the participant.
3. Tell the participant to sort the cards into predetermined distribution, for example, Downey (1960) asked the participant to sort the 16 cards into 7 piles with the number in each pile set as follow.

Most Important							Least Important
1	2	3	4	3	2	1	

4. Collect the cards and calculate appropriate statistics (such as analysis of variance and correlation) on resultant data.

In using Q-sort, the evaluator has to bear in mind that there are two basic types of Q-sort: structured and unstructured. Structured Q-sort includes a set of rules whereby a certain number of cards must be placed each in a certain number of piles such as in the case of Downey's study. In using unstructured Q-sort, the cards are placed into numbers of piles

according to the sorter's own perception of where they should be placed. Jones (1956) suggests that unstructured Q-sort should be used.

Rating. Rating is the easiest technique to find the congruence in perception among the groups of the participants. The participants will be asked to rate the various goals by checking such terms as absolute importance, great importance, medium importance, and etc.. Figure 8 shows the typical rating from that can be used to ask the importance of possible goals of the College of Education.

Money Game. A Money Game has been developed by Mullen (1974) at the University of Georgia called BONANZA game. The purpose of the BONANZA game is to help to discover the goal area priorities which exist in any particular school community. In the BONANZA game, nine areas on the 3 R's: social studies; science; work world; the arts; health, P. E., and safety; making choices; relationship with others; and development of self, are provided. A player (participant) is given 20 play monetary bills, each in the denomination of \$100, and spends the money according to the amount of educational time and effort he feels should be devoted to a particular area. These will provide an opportunity for the participants to purchase the kind of education they feel is needed.

The BONANZA game is an effective instrument for selecting priority goal areas in elementary and secondary school levels. The BONANZA game has not been applied to the higher education level yet, but the concept and the technique of the game can be used, and promise the same effectiveness at higher education level.

In finding the congruence in perception among the groups of the participants in the priority goal areas, the evaluator has to decide which

These are the GOALS of College of Education. How important do you think each of these GOALS should be?

AI - Absolute Importance  
 GI - Great Importance  
 MI - Medium Importance  
 LI - Little Importance  
 NI - No Importance  
 DN - Don't Know

GOALS	AI	GI	MI	LI	NI	DN
Development and refinement of instructional skills.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clarifying an image of what teaching is and what it could become.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Improvement in communication skills.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
etc.				etc.		

---

Figure 8: The format of rating scale.



instrument is best for the present situation. The present situation will be determined by the time and budget. If the time and the budget are limited, it may be that only using needs assessment would be enough. But if unlimited time and budget are available, the Delphi Technique is recommended because it yields more reliable results. Whatever instruments are used, the evaluator will come up with the list of the priority goal areas. Those priority goal areas will be ranked according to their importance based on the perception on the participants. The list of the goal areas will be presented to the PPC. The PPC and the evaluator have to discuss and select the goal areas that the College of Education should provide or improve on in its educational programs. Time and resources (budget, personal, instruments, and utility) will be the key factors in selecting the educational goals. If, as is sometimes the case, the first priority goal areas cannot be obtained because of a lack of time and resources, the second or third priority goal areas should be selected.

Once a set of goal areas have been agreed on, the PPC and the evaluator may translate each goal area into instructional objectives. These instructional objectives should provide tangible indications that process in a given goal area is being achieved.

### Stage 3: Program Planning.

When the program goals and objectives have been set, the program can be planned. Planning means laying out a course of action that we can follow that will take us to our desired goals (Churchman, 1968). The problem that is faced in the planning process is concerned with the uncertain future, which is hard to predict. As Quade (1975) says planning is forecasting, the value of which does not necessarily lie in whether or

not it comes true. Also the social environment is in a constant state of flux. This means that planning must be continuous and dynamic; it must anticipate change. Rittel and Webber (1973) point out that there is no general theory of planning. Even worse, planning is faced with problems which have no solutions in the sense of definitive and objective answers.

Why should the planning be done? Grotelueschen and Gooler (1972) said that planning should be done because it leads to a more efficient and effective allocation of resources, as systematic planning minimizes haphazard expenditures of funds and personnel (p. 8). Davies (1976) points out that by planning, it is possible to anticipate and hopefully avoid unforeseen difficulties, as well as to ensure that the potential of the situation is more effectively realized. Planning should be done to avoid future problems by anticipating them, and to obtain a more desirable future by working toward it in the present (Wildavsky, 1973).

Davies (1976) stated that:

Planning is anticipatory decision-making. It involves deciding what to do and how to do it before any concrete action is taken. It involves decision, after due consideration of the nature of relationships between the ends that you wish to accomplish and the means that you have available to use. When both ends and means are known and agreed, no real problem or difficulty exists. (p. 5)

Davies also provides three paths of planning which can be seen in Figure 9.

From Figure 9, Path X represents a systematic approach toward curriculum development. It involves first defining the goals to be achieved

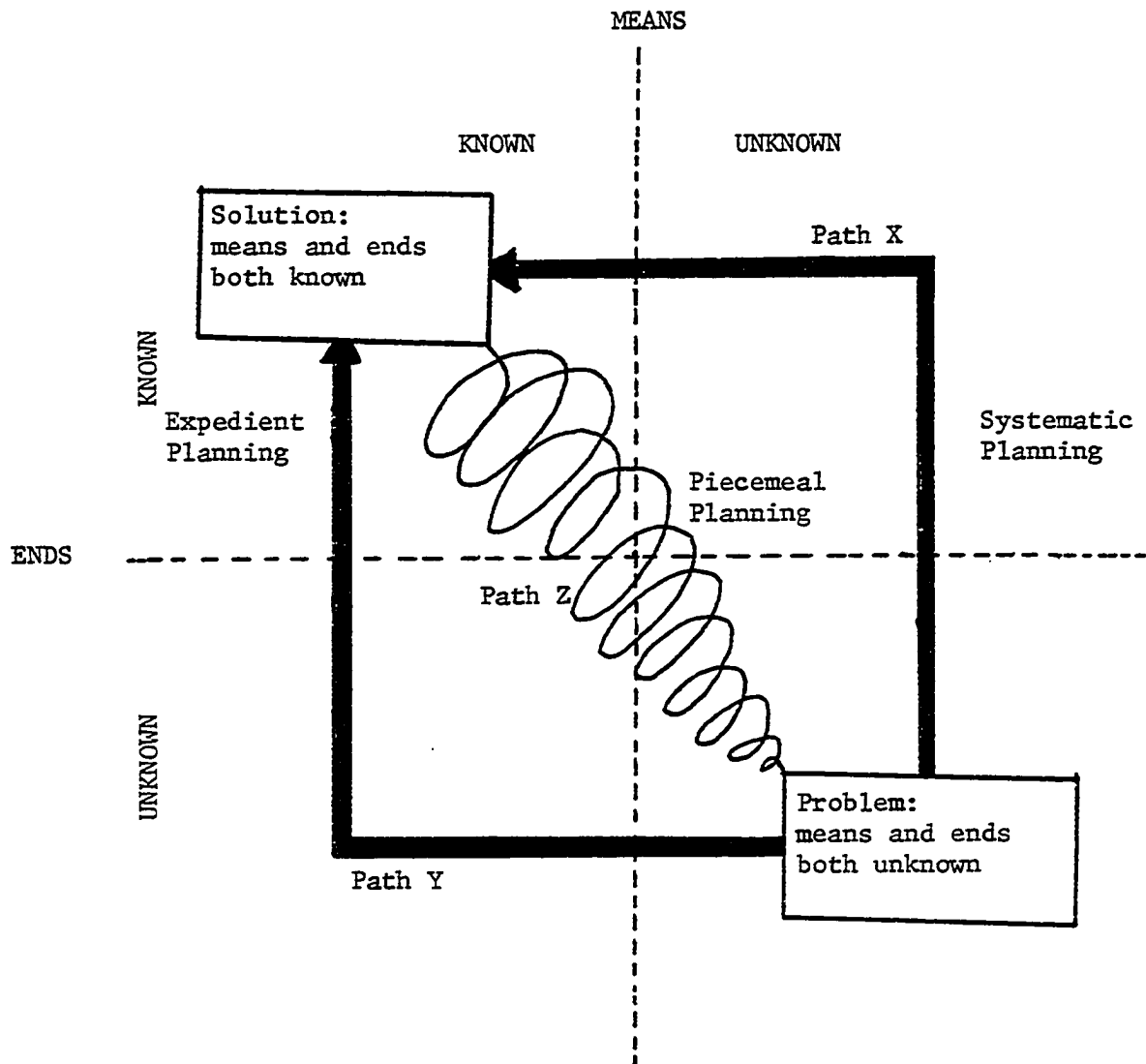


Figure 9: A means-ends perspective of planning. (From I. K. Davies, Objectives in Curriculum Design. New York: McGraw-Hill, 1976.) (p. 5).

and then selecting the means or procedures necessary for realizing them.

Path Y involves first defining the means or procedures to be used, and then determining those objectives that best accommodate the limitations or constraints that you have imposed upon yourself. This path is much closer to the traditional way of operating with the limits set by many educational situations and budgets, as circumscribed by the economic situation.

Path Z involves successive approximations for which no initial blueprint or master plan is necessary. This approach may start with some more important goals, before considering the resources needed; then back to a more precise definition of objectives, before turning once again to the means available, etc.. As a result of this constant turning and re-cycling backwards and forwards, more and more information becomes available, and a more realistic approach to development is sometimes possible.

It is very interesting to point out that most program planning starts with Path X and after the program has been implemented into the real setting, Path Z seems to be developed. It is likely that Path Z is more realistic in dealing with the uncertain situation than the other paths especially in the social environment.

In the program planning, the PPC will develop several alternative instructional methods as possibilities to fulfill the need (goal) areas. The PPC not only develops several alternative instructional methods but also the strategies for proceeding with those alternatives. There are several tools that can be helpful to the PPC in developing the strategies such as PERT (Miller, 1963; Case, 1969) PPBS (McCullough, 1966; Hinrichs & Taylor, 1969; Hovey, 1970; Whritner & Antin, 1972), and system analysis

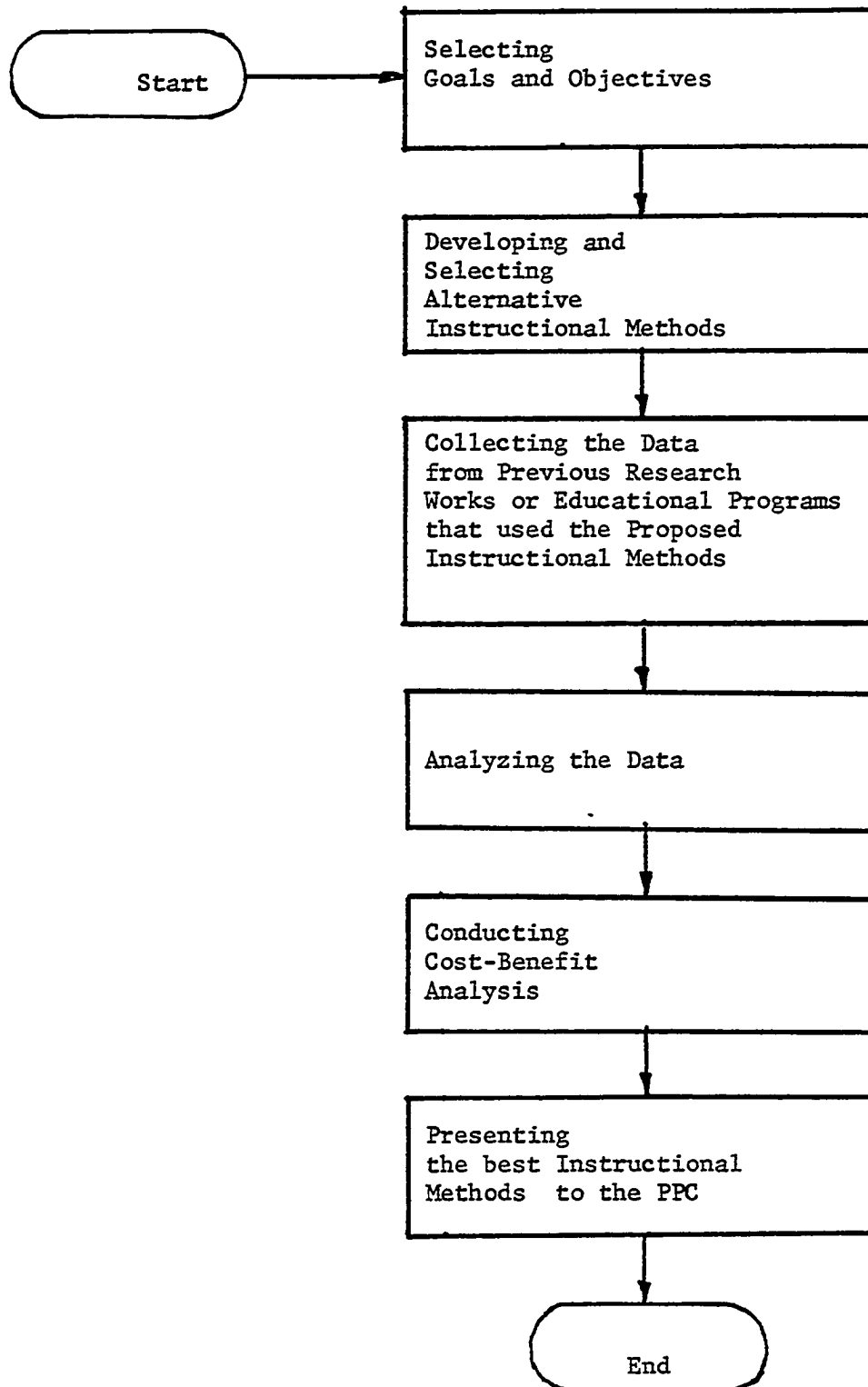


Figure 10: Program planning processes.

(Kaufman, 1968, 1971a, 1971b; Cleland & King, 1968). The learning structure analysis of Gagne' (1965) is useful for instructional planning.

In the process of program planning, the PPC have to bear in mind that good planning must be flexible and ready to be changed or modified. The process of program planning is shown in Figure 10.

After the tentative program has been planned with a variety of alternative instructional methods that can be used to approach the program goals, the PPC will ask the evaluator to provide information on the possible consequences of selecting each alternative. The evaluator also has to provide the information on the cost to obtain each alternative. With this information, the PPC will be able to select the most effective alternative instructional methods at the lowest cost.

Collecting the Data. Data collection in the stage of program planning is different from other stages. The evaluator does not have to develop evaluation designs or select or construct measuring instruments. What he has to do find out whether or not there has already been any research done or whether any evaluation studies performed are similar to the program designed. If previous studies have been done, it must be known how they process their programs. What are the outcomes of the programs? What are the consequences of the programs? These information will increase the likelihood of predicting the success of the designed program.

Actually, collecting the data from the primary resources (field studies) will give a more accurate and clear picture of the effect of the program designed but it is not economical. As Freeman and Sherwood (1970) said:

Field studies are exceedingly costly, and require a

major commitment of personnel. Moreover, there is a long wait for the results. Finally, efforts to speed up the research, or economize, almost invariably result in studies so incompetent that it would have been better not to have made them at all. (p. 42)

Then, it will be better in terms of time and finances for the evaluator to collect the data from the sources of secondary data (the data which have already been collected and processed (Freeman & Sherwood, 1970; Grotelueschen & Gooler, 1972; Alkin, 1969)).

Sometimes the best sources of secondary data are stored in computers. One of the most efficient and economic system is GIPSY (General Information Processing System). GIPSY is a sophisticated information handling tool developed at the University of Oklahoma (Office of Information Systems Programs, 1977). The system is simple and easy to use, even by the non-computer professional. The system is well designed in eliminating the problems of too much information by providing the precise descriptors, plus using the Boolean logic. The system provides four data bases, one for CLJE and three for ERIC. The information provided by the system includes the title, the author(s), publication date, the name of the journal, accession number of the report, and the abstract.

The data also can be collected from government agencies such as HEW (Health, Education, and Welfare) and the State Departments of Education. Another source is books in print which is published every year. There also are many reports which may never be published or may be published later that are available. Access to these reports is not easy. The best access to these reports is to establish a relationship with the persons

and agencies that collect and process descriptive data.

Analyzing the Data. After the evaluator has collected the reports from the previous research works which are similar to the program designed or using the same instructional methods proposed in the program designed, he has to ask himself is this research reliable. What are the researchers trying to tell? Are they biased? What types of research designs do they use? Are those designs controlled for both external and internal validity? Do they use reliable measurements? How do they select their subjects for the study? Do they have concrete evidence to make their conclusions? The evaluator has to answer these questions before he can make any judgments or recommendations on the likelihood of the success of the program designed. The evaluator can also use content analysis (Berelson, 1954; Kerlinger, 1973) to analyze the reports from the previous studies.

There are at least three more factors that the evaluator should assess before he makes any report and recommendation on the program designed to the PPC. First, he has to assess the resources that are available in both the college and the community. He has to find out whether the staff has to be trained in order to operate the new program. Does the college have to hire new personnel to operate the program? Does the college have enough teaching-learning instruments required in the program? Does the college have enough utilities for operating the new program? Does the college have enough budgeted to operate the entire new program? Are there any resources such as personnel, budget, or instruments available within the community?

Second, the evaluator has to figure out the reliability of his prediction on the effectiveness of the program designed. He can do this



by acquiring the opinions of the experts. He can use face to face discussion with the experts or use Delphi Technique. The Delphi Technique is a more preferable technique because it saves time and eliminates the conflict among the experts from the process of face to face discussion. It also eliminates the schedual conflict among the experts. Another technique that the evaluator can use is simulation game. In this technique, the computer will be programed to simulate the future of the program. The evaluator will ask the computer to forecast what would happen if a certain alternative were adopted, and by means of various kinds of judgments and facts stored in the computer, a prediction will come out.

Third, the evaluator has to assess the likely costs and benefits of alternative means proposed in the program designed. In other words, the evaluator has to conduct a cost-benefit analysis.

Cost-Benefit Analysis. Cost-benefit is primarily an economic analysis. The idea of cost-benefit analysis is simple. The evaluator has to identify the costs and benefits of the alternative means into a common unit of measure--dollars. When the costs and benefits of the alternatives are certain, the selection can be done. The alternatives can be selected according to either yields, the largest benefits for given cost, or the alternative which will yield the least cost for a given level of benefits. The idea of the cost-benefit analysis sounds simple as mentioned, but it is difficult to implement in reality. As Tripodi et al. (1971) mentioned, cost-benefit analysis can be valid only if the indicators of benefits can be translated into a monetary unit. Some benefits may not be registerable in monetary equivalents, or just as problematic, the amount of assigned economic value may be purely arbitrary. That will jeopardize

the decision. This problem occurs frequently in the outcomes of education. How can the change in the student's concept, interest, or self concept be valued in the monetary unit? Weiss (1972b) suggests that agreement on the value of the outcomes should be set by using a system of weights. The system of weights should be developed according to the values of gains for specific outcomes.

Beside the problem of identifying the outcomes into the monetary unit, cost-benefit analysis can provide a clear cut process of selecting the alternative means proposed in the program designed. It is ascertained that the alternative means that yield the highest benefits with the lower costs should be selected.

After all the information has been collected, the extensive report should be presented to the PPC for discussion. The recommendation on the program selected will be presented to the dean of the college for final review and decision.

#### Stage 4: Program Implementation.

Program implementation is the crucial stage in determining whether the well planned program will have a chance to succeed in the actual setting. Unfortunately, it seems to have received little conceptual attention from anyone (Guba, 1967). With little attention paid to the implementation of the program, it is possible that some parts of the program will be poorly implemented or not implemented at all. Williams and Elmore (1976) said:

The fundamental implementation question remains whether or not what has been decided actually can be carried out in a manner consonant with that underlying decision.

More and more, we are finding, at least in that case of complex social programs, that the answer is no.-- It is possible that past analysis and research that ignored implementation issues may have asked the wrong questions, thereby producing information of little or no use to policy making. (p. xi)

Implementing the new program into the organization (such as a College of Education) is not a simple issue. The most significant problem in the implementation stage is concerned with the members of the organization. As Hage and Aiken (1970) stated,

Organizations can have operations researchers carefully design a system for handling a new program, but the human element is seldom adequately considered in the implementation of a new product or service. There will be mistakes that will have to be corrected. Alteration of the existing structure will also create conflicts and tensions among the members of the organization. (p. 100)

Naturally, when the organization wants to create a new program, a special group of people will be formed to take on the responsibility of the new program. These people will need power to display their authority in the pursuit of success for the new program. It sometimes causes resistance from members of the organization, which in turn decreases their contribution to the new program. That could make the program fall short of its stated goals.

Another problem dealing with the members of the organization is that when the new program has been designed, they are usually required to utilize new materials and practices, but they will often tend to return to more familiar methods of operation (Skager, 1969). Hage and Aiken (1970) and McLanghlin (1976) suggested that those problems could be avoided by sharing the power of decision making with the members of the organization. A staff meeting should be conducted and the plan of the new program should be presented to them. Staff meetings will give staff a chance to communicate with the program committee, to share ideas, and to provide each other with encouragement and support. But Hage and Aiken add that in letting the staff share in the decision making, the plan may be curtailed by their diffuse and irrelevant suggestions.

Another problem in the implementation stage concerns the program itself. Sometimes the programs are well planned in the paper, but some parts of the program are difficult to implement in the actual setting because of time, budget, personnel, and instruments. As a consequence, the plan has to be altered. Because of the unforeseen events, Williams and Elmore (1976) suggest that the program should be planned with specificity, flexibility, and incentives. The program should leave some options available for the staff to modify to suit their particular situation. Guba (1967) suggests that training should be introduced in case the staff is not familiar with the new instruments or techniques that are being introduced into the new program.

The role of the evaluator in this stage is to determine whether or not the program is being implemented according to the plan. If the program cannot be implemented, or is implemented poorly in the actual

setting, the evaluator has to find out the reasons and suggest alternative solutions; and present his report to the PPC for discussion and decisions. This process has to have complete cooperation between the evaluator and the PPC because timing is the crucial element in deciding the outcomes of the program. To avoid rushing the decision, the evaluator has to be aware of the situation while the programs are being implemented.

The instruments that the evaluator can use to collect the information in the stage of implementation are: systematic observation, questionnaires, interviews, class records, and lesson plans. In systematic observation, checklist should be developed. The checklist will ask whether the on-going program has certain specified elements. The observers should be well trained and know what they should look for. In classroom observation, two observers should be used so the reliability of the observers can be checked. Weikart and Banet (1976) warn that constant observation can make the instructors become quite paranoid. In that case, other instruments can be used such as interviewing the students or the instructors about the events in the classroom, checking classroom records and lesson plans, and questionnaires. These options are sometimes a better idea than direct observation of the class.

By controlling the stage of implementation well, the evaluator can make sure whether or not some events in the implementation stage have an effect on the outcome of the program. Such as, if the actual operations do not meet specifications of the program planned, and the program fails, then the evaluator can say that failing to achieve the specifications of the plan may be the cause of the failure of the program. Also if the program is well implemented and the program still fails, then the evaluator

can at least eliminate the fault of the implementation and concentrate on other causes.

#### Stage 5: Program Improvement.

With most of the program evaluations that have been studied in recent years, the emphasis has been on whether or not the programs have succeeded (summative evaluation). Unfortunately, the results from summative evaluation have little to do with the improvement of the program because the program is finished. It would be a better idea and would benefit the students and the program if the evaluator and the PPC would pay more attention to the effect of the program on the students while the program is operating. As Cronbach (1963) said "evaluation, used to improve the course while it is still fluid, contributes more to improvement of education than evaluation used to appraise a product already placed on the market."

Evaluation in this stage is essential in the process of improving the inovative program. It is often that some of the processes that have been designed in the planning stage do not yield the expected results, then they need to be modified so the program can be improved and successful. The evaluator plays an important role in this stage by collecting as wide a variety of information on the program as possible, making judgments, providing alternative strategies, and presenting them to the PPC immediately. This information will enable the PPC and the evaluator to modify and improve the program. The decision to modify the program has to be made immediately because timing is the crucial part in the success of the program. Because timing plays an important role in the program improvement, Alkin and Fink (1974) and Butman and Fletcher (1974) suggest that the

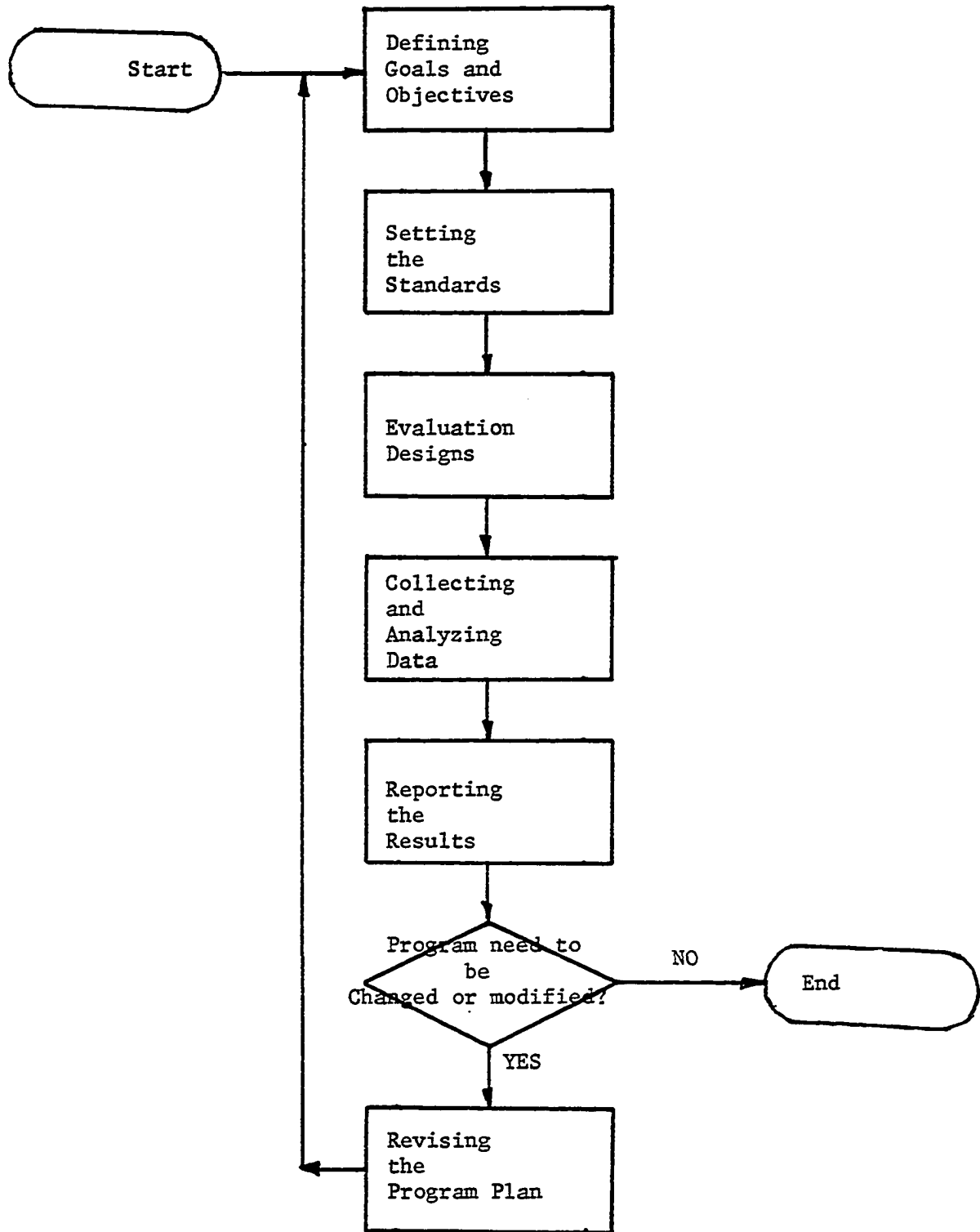


Figure 11: Program improvement processes.

evaluator should be one of the PPC so the information and the decision can be provided quickly. In the program improvement stage, the in-house evaluator is generally favored over an outsider, because he knows more about the program. He also works at the same place with most of the members of the PPC and he can reach them on time if the program needs to be modified or changed (Rutman, 1977; Weiss, 1972b). The processes of program improvement was shown in Figure 11.

Most of the evaluation processes in the stage of program improvement are similar to summative evaluation but with different purposes. Summative evaluation is conducted to see whether the program is effective or ineffective, while the program improvement aims to determine what effects the program has on the student's performance.

Defining Goals and Objectives. Again, the evaluator has to review all the program goals and objectives from the previous stage. The evaluator has to be sure that all the statements of the goals are clear and specific. The set of the statements of the objectives have to be derived from the statements of the goals. Actually, the statements of the objectives should be stated in behavioral terms so they can be observed and measured. The concept from Bloom's Taxonomy (Bloom, 1956; Krathwohl et al., 1964) and the learning structure analysis of Gagne' (1965) are helpful in setting the educational objectives. They are also helpful in constructing the test for the program.

Setting the Standards. As Provus (1969) said the judgment can be made by a comparison between the student's performance and the program standards. Then, the beginning task of the evaluator in this stage is to obtain program standards. There are two types of standards according to



Stake (1967), absolute standards and relative standards. Relative standards are involved in comparing the results from an innovative program with the another comparable programs or previous programs, while absolute standards compare the results of the innovative program with the standards of excellence such as, all the students in the new program should score at or above the 70th percentile, based on national norms, on a standardized mathematics achievement test. Cronbach (1963) and Stake (1967) suggest that absolute standards should be set in the stage of program improvement. Scriven (1967) on the other hand, does not believe that absolute standards should be used at all in the program evaluation. Standards will be discussed more in the summative evaluation stage.

Evaluation Designs. In evaluation designs in the program improvement stage, the evaluator has to decide how he is going to select the respondents to be studied, how the data should be collected, and when it should be collected. In developing evaluation designs, the evaluator can use true experimental designs in which he has to set a control group using randomization in selecting the respondents. If randomization is prohibited, then quasi-experimental designs can be used. Both true experimental designs and quasi-experimental designs will suggest to the evaluator how to select the respondents to be studied and when the data should be collected. These designs control well both internal and external validity. In the program improvement stage, internal validity is heavily emphasized. The true experimental designs and quasi-experimental designs will be fully discussed in the summative evaluation.

There are so many instruments that can be used to collect the data such as observations, interviews, questionnaires, psychometric tests

(attitudes, values, personality, interests), tests of information (skills, knowledge), rating scale, and etc.. The evaluator will select from those measuring instruments according to the program goals and objectives and also the information that he thinks might influence the effectiveness of the program.

There are some controversial beliefs in the process of evaluation designs among leading evaluators. Cronbach (1963) and Stake (1967) do not think that experimental designs should be used in the program improvement. They said that the question of the program improvement is whether or not the program has an effect on students. This type of question does not need any comparison between the outcomes of the new program and the comparable programs at all. It only wants to find out if there are any changes in the student's performance after the program has been implemented. Scriven (1967), Weiss (1972b), and Sanders and Cunningham (1974) on the other hand think that experimental designs should be used in both program improvement and in summative evaluation. In using experimental designs, they can be sure that internal validity will be well controlled, which the suggestion from Cronbach and Stake are weak in this area. By using the concept from Cronbach and Stake, time series design fits their idea and it also controls internal validity.

Collecting and Analyzing Data. There are many measuring instruments that can be used in collecting the data as mentioned in the previous step. The evaluator has to bear in mind that he should not concentrate only on the information about whether or not the program has an effect on the performance of the students according to the program goals but consider also the information on the side-effects of the program (Scriven, 1967;

Cronbach, 1963). Information needs to be collected by the evaluator other than the students' test scores are attitudes (Cronbach, 1963), satisfaction (Baker, 1974), socioeconomic status, faculty experience, teaching style (Sanders and Cunningham, 1974) and etc.. This information should be collected before the new program is implemented so that it can be used as base line data. It also should be collected during the program operation and after the end of the program. This information is valuable to the evaluator and the PPC to make a decision on whether the program should be modified beside the information on students' test scores.

After the evaluator collects all the data, he has to decide what statistical analysis should be used. He has to select the right statistical techniques that will strengthen the results. Statistical techniques will be selected according to the nature of the data that obtained from the program.

Reporting the Results. The evaluator has to report the results immediately so if there are some processes that need to be changed or modified, the PPC will have a chance to make a decision on time. In the report, the evaluator has to judge whether any particular portion of the program has a positive and negative effect on students performance. If any particular portions of the program do not have any effect or have negative effect on the student's performance, the evaluator has to find out why and provide the alternative strategies for changing or modifying the program. Not only the effect of the program on the student's performance should be reported to the PPC but the reaction from other participants such as faculty, teachers, principals, and parents should be reported also. The form of the report need not to be formal like the summative

evaluation report because it will save time, and timing is the crucial part in the process of program improvement.

Stage 6: Summative Evaluation.

After the program has been completed, the PPC has to report the outcomes and the judgment on the effectiveness of the program to the dean of the college so he can decide whether the program should be continued, improved, or terminated. The effectiveness of the program should not be determined only by how well the goals of the program have been met, but also by the side-effects of the program. Scriven (1972b) supports this idea when he points out in his GFE that most of the evaluators tend to over look the side effects of the program which might well be the crucial achievement.

Because the information from summative evaluation has a tremendous effect on the program, the summative evaluator should not only be objective about the program being evaluated but also seem to be objective. As Anderson et al. (1975) said both being objective and seeming to be so are important if the conclusions reach by the summative evaluator are to be valid and reliable. Scriven (1967) suggests that summative evaluation should be conducted by external evaluators because they are professional and well trained. By using external evaluators, it can be assured that there are no "whitewashes" or favorably biased reports being written. External evaluators also have an advantage over in-house evaluators on administrative confidence, objectivity, and autonomy (Weiss, 1972b). External evaluators should be brought in after the program has been planned and before the program will be implemented.

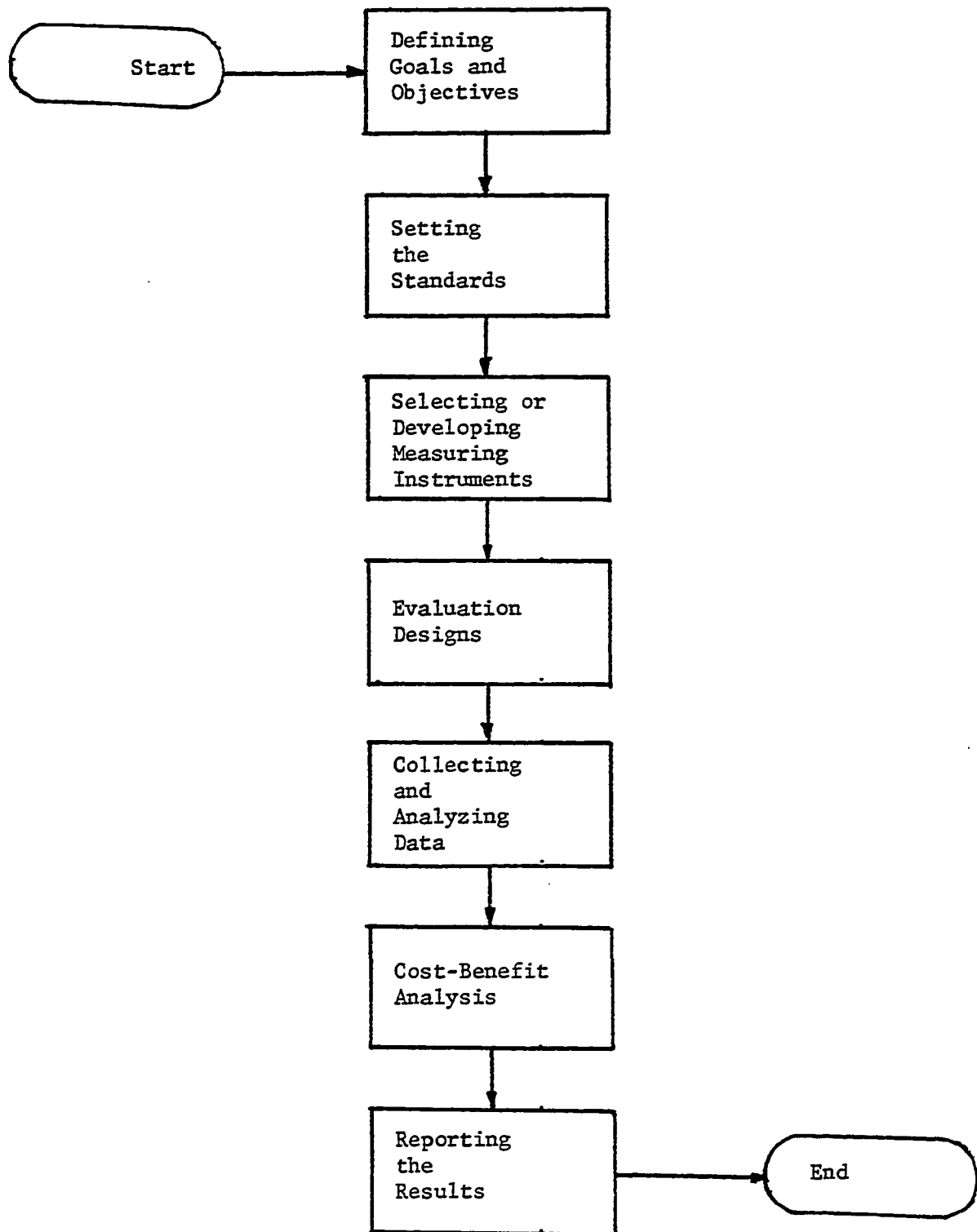


Figure 12: Summative evaluation processes.

There are still some problems with the external evaluators. First of all, they lack an understanding of the program and secondly, they are expensive. In the case that the PPC has to be faced with a tight budget, in-house evaluators should be the best choice. The processes of summative evaluation is shown in Figure 12.

Defining Goals and Objectives. Defining goals and objectives in the summative evaluation stage seems to be a revision of the second stage (Setting Goals and Objectives) of the CEM. The evaluator should look through the previous file of goals and objectives of the program and find out whether are there any fuzzy goals left. If there are some left, the evaluator has to make them clear so they can be observed and measured.

Setting the Standards. As Scriven (1967) said, evaluation is the judgment of merit or worth. Before the evaluation study can be processed, the PPC and the evaluator have to set the standards of the program. There are two basic standards of judgment which Stake (1967) presented in his Countenance Model, one is an absolute standard and the other is a relative standard. (see Figure 13) Absolute standard is the standard that has been set by the PPC and the evaluator. An example of an absolute standard would be a statement such as all the students in the new program should score at or above the 80th percentile, based on national norms, on the GRE (Education section). The relative standard refers to how the students in the new program performed in comparison with the previous programs or other comparable programs.

Cronbach (1963) suggests that relative standard should not be used in evaluation. As he states the aim to compare one course with another should not dominate the plans for evaluation. He also points

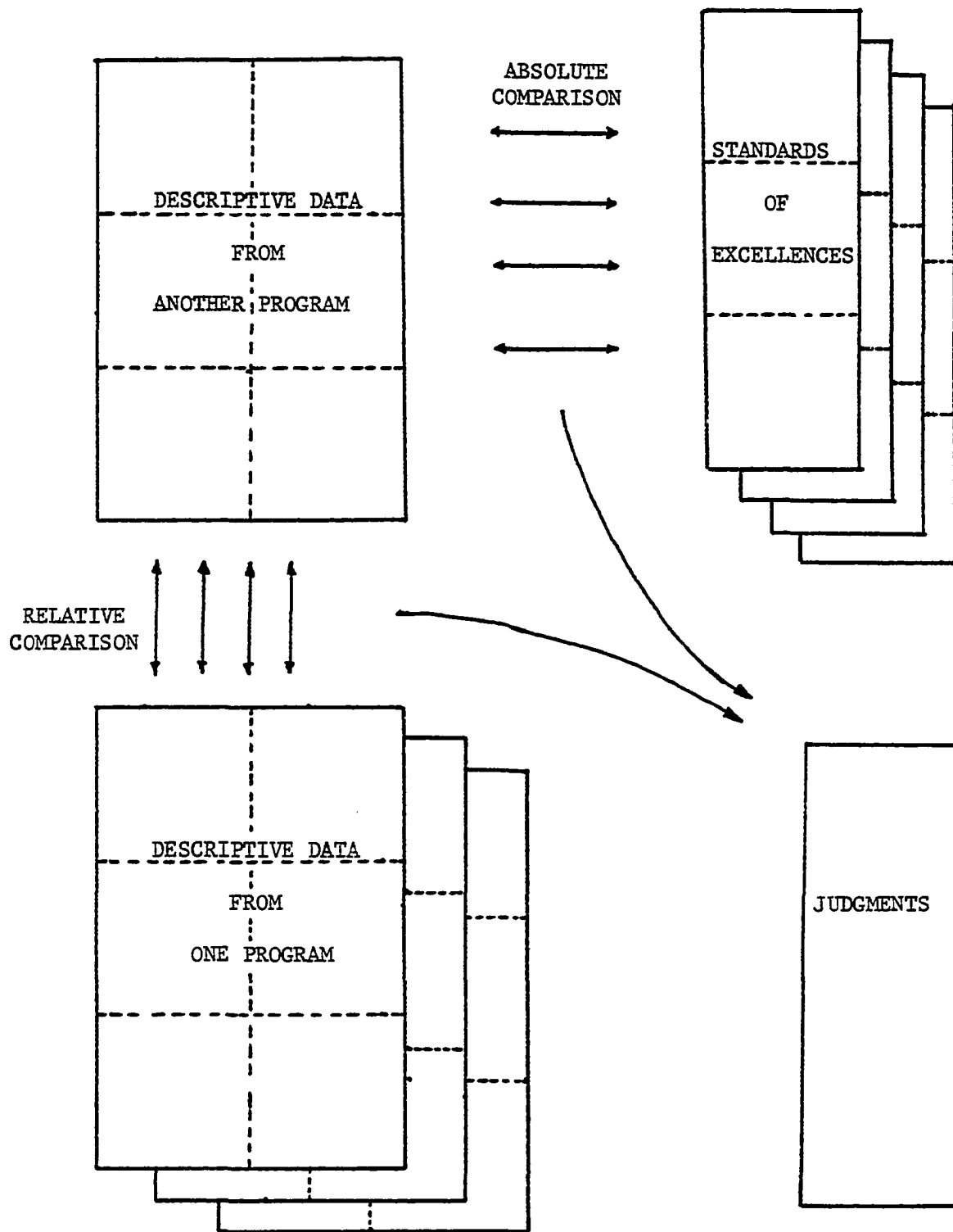


Figure 13: A representation of the process of judging the merit of an educational program. (From "The Countenance of Educational Evaluation" by R. Stake, Teacher College Record, 1967, 68, 523-540.) (p. 537).

out that the experimentation has to be done by comparing one course with another. There are some weak point in the social experimentation such as failure to equate the classes taking the competing courses and Hawthorne effect which could possibly jeopardize the interpretation of an outcomes. Cronbach also suggests that a formal study should be designed primarily to determine the post-course performance of a well described group with respect to many important objectives and side-effects. Scriven (1967) does not agree with Cronbach. He said:

We are interested in curricula because they may prove to be better than what we now have, in some important way.--Comparative evaluations are often very much easier than noncomparative evaluations, because we can often use tests which yield differences instead of having to find an absolute scale and then eventually compare the absolute scores. (p. 64)

Stake (1967) suggests that either the absolute or relative standard should be used according to the question that the PPC want to answer. If the questions are "which is best?" and "which will do the job best?", then the relative standard should be used. The absolute standard should be used when the question is how can we teach it better. The relative standard is always tied with the process of experimental designs which are employed in most of the summative evaluations. If the evaluator decides to use experimental designs in his study, then relative standards should be used.

Selecting or Developing Measuring Instruments. After the goals and objectives have been cleared, the evaluator has to select or develop



instruments to measure those goals and objectives. There are three criteria in selecting or developing measuring instruments.

1. Reliability. Kerlinger (1964) states that to be interpretable, a test must be reliable. Reliability means consistency, dependability, and accuracy. The reliability of measurement is insured when two or more applications of the same measure or a comparable measure under comparable conditions yield the same results.

2. Content Validity. Content validity is the representativeness or sampling adequacy of the content--the substance, the matter, and the topics--of a measuring instrument (Kerlinger, 1964). Content validity is necessary for the measuring instruments because without content validity we can not say that students have mastered or failed the specific instructional objectives.

3. Budget. The third criterion is also important. There are many measuring instruments that are available in a variety of costs. The evaluator has to determine which measuring instruments will give the most effectiveness within a price range that the program can afford.

Evaluation Designs. When an evaluator designs an evaluation study, he has to ask himself if the design controls internal and external validity. Internal validity is the basic minimum without which any experiment is uninterpretable. External validity relates to the problem of the generalizability of the results of the study. There are eight factors that will jeopardize the internal validity if the evaluator does not control them (Cambell & Stanley, 1963). They are:

1. History: the specific event occurring between the first and second measurement in addition to the experimental variable.

2. Maturation: processes within the respondents operating as a function of the passage of time per se (not specific to the particular events).

3. Testing: the effects of taking a test upon the scores of a second testing.

4. Instrumentation: in which changes in the calibration of a measuring instrument.

5. Statistical regression: operating where groups have been selected on the basis of their extreme scores.

6. Selection: biases resulting from differences between types of individuals recruited for comparison groups.

7. Experimental mortality: or differential loss of respondents from the comparison groups.

8. Selection-maturation interaction, etc..

Campbell and Stanley also discussed the four factors that can jeopardize the external validity. They are:

1. The reactive or interaction effect of testing.
2. The interaction effects of selection biases and the experimental variable.
3. Reactive effects of experimental arrangements.
4. Multiple-treatment interference.

External validity does not get much attention in evaluation studies especially at the local level. Most evaluators are concerned only with internal validity, because it has a direct effect on the results of the study. If they do not control internal validity, it is very difficult for them to make a conclusion on the results of the study, because they really

do not know which factors have an effect on the respondents.

There are two techniques that can be used to control extraneous factors and to make research internally valid. They are control or comparison groups and randomization. These two techniques are employed in the true experimental designs. This is the true experimental design that has been used frequently in evaluation studies.

Pretest-Posttest Control Group Design.

R	O <sub>1</sub>	X	O <sub>2</sub>	Experimental
R	O <sub>3</sub>		O <sub>4</sub>	Control

R will represent randomization, O will refer to some process of observation or measurement, and X will represent the exposure of a group to an experimental variable or event, the effects of which are to be measured.

Pretest-Posttest Control Group Design has well controlled internal validity (Campbell & Stanley, 1963). Both experimental and control groups are randomly assigned. This design can be used to compare the new program and the tradition program. If the difference between O<sub>1</sub> and O<sub>2</sub> is greater than the difference between O<sub>3</sub> and O<sub>4</sub>, the program is a success. The t-test and analysis of variance are frequently used to analyze the data from this design.

There are some problems in using true experimental designs in educational settings. Weiss (1972b) said that the controlled experiment is often impossible in an action setting. Using randomized assignment is very difficult to achieve. At times even after it has been achieved, participants may drop out during the course of the program, a factor which

the evaluator has no control over. Guba and Stufflebeam (1968) also criticize using true experimental designs in evaluation. They said true experimental designs do not facilitate the program implementation and the program improvement at all. Another problem of true experimental designs is the Hawthorne effect. Hawthorne effect refers to changes in behavior that occur when the subjects in an experiment are aware of their special status (Anderson et al., 1975).

Another type of design that controls internal and external validity is quasi-experimental designs. Quasi-experimental designs have been developed by Campbell and Stanley (1963). One of the prominent distinctions between quasi-experimental designs and true experimental designs is the absence of a randomized assignment of subjects to treatments. Quasi-experimentations also have the advantage of being practical when the conditions prevent true experimentation (Weiss, 1972b).

These are the most used quasi-experimental designs in the evaluation studies.

Time-Series Design. It involves a series of measurements at periodic intervals before the program begins and continuing measurements after the program ends. It can be diagrammed thus:

O<sub>1</sub>    O<sub>2</sub>    O<sub>3</sub>    O<sub>4</sub>    X    O<sub>5</sub>    O<sub>6</sub>    O<sub>7</sub>    O<sub>8</sub>

This design is helpful when experimental control is impossible. This design controls most of the factors that will jeopardize internal validity except history. It is obvious that the students learn various things beyond those under the experimenter's control in the classroom. Figure 14 shows some possible outcome patterns for time series.

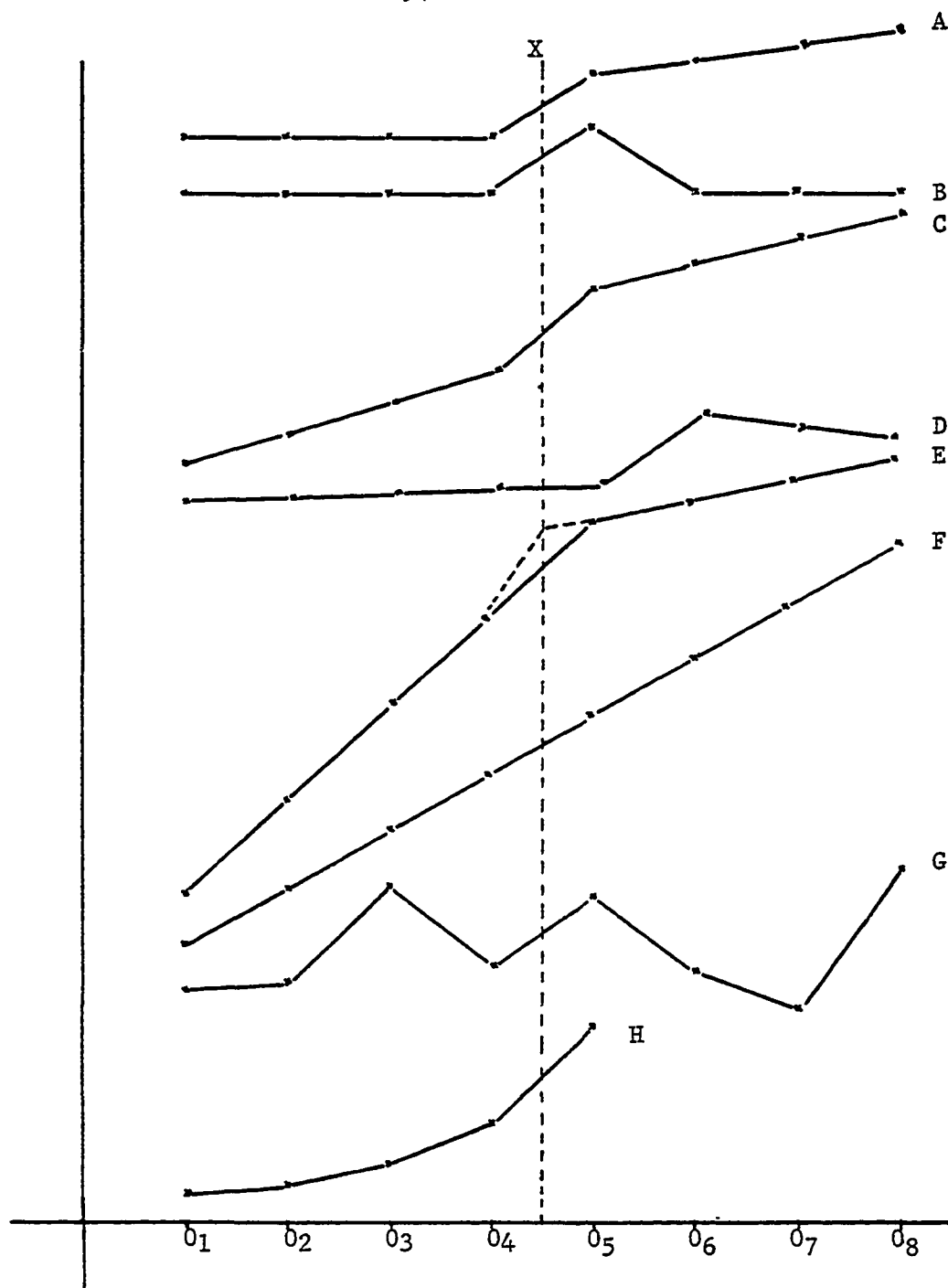


Figure 14: Some possible outcome patterns from the introduction of an experimental variable at point X into a Time Series of Measurements, 0<sub>1</sub> - 0<sub>8</sub>. (From D. T. Campbell & J. C. Stanley, Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.) (p. 38).

From Figure 14, a gain is seen from the beginning to the ends of the program ( $O_4 - O_5$ ) in most of the time series except for D. It seems that the program had some effect in time series with the outcomes such as A and B and possibly C, D, and E but not as F, G, and H.

History can be controlled by the evaluator if he can find a similar group or institute and take periodic measurement of it over the same time span. It can be diagrammed thus:

$O_1$	$O_2$	$O_3$	$O_4$	X	$O_5$	$O_6$	$O_7$	$O_8$	Experiment
$O_1$	$O_2$	$O_3$	$O_4$		$O_5$	$O_6$	$O_7$	$O_8$	Control

The Nonequivalent Control Group Design.

$O_1$	X	$O_2$	Experimental
<hr/>			
$O_3$		$O_4$	Control

This design is similar to the Pretest-Posttest Control Group Design but the subjects in the experimental and the control group are not randomly assigned. This design is well worth using when the random assignment to treatments is not possible. The subjects in the experimental and the control group will be assigned by the procedure of matching. But the problem of matching is that we often cannot define the characteristics with which people should be matched. We often times do not know which characteristics will effect the benefits the person receives from the program. But it is agreed that having a control group is better than having nothing. Campbell and Stanley (1963) suggest that using analysis of covariance with this design will produce the stronger results.

Collecting and Analyzing Data. In the processes of collecting and analyzing data in summative evaluation stage, the evaluator has to face the following questions. From whom should the data be collected? How should it be collected? When should it be collected? Who should collect it? How should the data be analyzed?

First of all, the data should be collected from all the people who are associated with the program. Such as with the Field Experience Program, students, public school (teachers, principals, superintendents, and students), and College of Education (faculty and administrators) are associated with the program. The data should be collected from these groups of people.

There are so many measuring instruments that can be used to collect the data such as interviews, questionnaires, rating scales, observation, teacher made tests, standardized achievement tests, and psychometric tests (attitudes, values, personality, preferences, beliefs), etc.. Selection of instruments should be based on the goals and objectives of the program. Before the right measuring instruments can be selected, the evaluator has to understand the goals and the objectives of the program well. This point is contradicted by Scriven. Scriven (1972b) does not think that it is necessary for the evaluator to know about the goals of the program because he believes that the more the evaluator knows about the goals of the program the more tunnel-vision will develop. By doing that he would tend to overlook the side-effects of the program which sometimes are more important than the goals of the program themselves. With this idea in mind, he has developed alternative evaluation

technique called Goal-Free Evaluation (GFE). The idea of GFE is good but it may be unrealistic. If the evaluator does not know what the program goals are, it is impossible to select the right measuring instruments. Kneller (1972) suggests two alternatives: first, by training the evaluators to observe both goals (and outcomes) and side-effects; second, by the researcher's specifying as many likely side-effects as possible within the original research design (p. 7).

The time for collecting data depends on the design of the evaluation. If the evaluator uses Pretest-Posttest Control Group Design, then he has to collect the data before the program starts and after the program ends.

The evaluator's staff, separate from the PPC, should collect the data. The reason that the evaluator's staff should collect the data is because they have been trained and know the techniques of the measuring instruments. By using those people to collect the data it can be assured that the data are valid and reliable. The valid and reliable data will have an effect on the accuracy of interpreting the outcomes of the program.

After the data have been collected, they will be analyzed. The statistical techniques will be selected according to the nature of the data. The t-test, analysis of variance, analysis of covariance, regression analysis, factor analysis, and chi-square are the most common techniques used in recent evaluation studies.

Cost-Benefit Analysis. After the program has been completed and the results have been analyzed, a cost-benefit analysis should be



done. The concept of cost-benefit analysis has already been presented in the program planning stage. The difference is that in the program planning stage, cost-benefit analysis is in the projective stage, but cost-benefit analysis in the summative evaluation stage is in the actual stage.

Reporting the Results. The evaluation results will be reported directly to the PPC. The report should be written in a formal style. Guidelines for technical reports in the social sciences can be used. Mostly such reports will include complete accounts of the program, the evaluation procedures, and the outcomes. The outcomes of the program should report on both the area of the program success and the area of the program failure. The side-effects of the program should also be reported. In the report, the evaluator should also judge the value of the outcomes and recommend the alternative direction to the PPC in the case that the program has failed.

Timing is a crucial point for the report of the outcomes of the program also. The report should be sent to the PPC before they recommend whether to continue or discontinue the program to the dean of the college. If the report come after the decision has been made, then it is a waste of time to conduct an evaluation.

#### Stage 7: Follow-up Study.

The last stage of the CEM is the follow-up study. The purpose of the follow-up study is to observe and record the effects of the program upon the performance of the graduates after they have left the program. This purpose is important to the improvement of the teacher preparation programs. Eventhough the results from the summative evaluation will show the effectiveness of the program, it cannot be guaranteed that the graduates

will be the effective teachers. The follow-up study will link the results of summative evaluation and the performance of the graduates in the public schools.

The follow-up study also can be used for other purposes, such as to evaluate the effectiveness of the programs. Wass and Combs (1973) conducted a follow-up study with thirty-five (35) teachers who graduated from a new elementary program and with thirty (30) teachers who graduated from the regular elementary program. They found that the teachers who graduated from the new elementary program performed in the classroom better than the teachers who graduated from the regular elementary program. Central Michigan University also used a follow-up study for the purpose of evaluating teacher education programs (Wotring, 1972). A questionnaire had been developed with 11 criteria: planning and organization, methods and materials, motivation, evaluation, management, overall classroom effectiveness, professionalism, community skills, academic preparation, personnel qualities, and human relations. The graduates and the principals participated in this study. The follow-up study was used as a teacher evaluation model at the Tennessee Technological University (Ayers, 1974).

The follow-up study can also be used to provide information on the opinion of the graduates about the training programs which can be used for program improvement. Beaty (1969) perceived this concept and conducted a follow-up study with the graduates of the undergraduate teacher education program for the class of 1964 at the Middle Tennessee State University to find out whether the teacher education program provided sufficiency experience for the graduates to work in the school. Ballantine et al. (1966) sent the questionnaires to the teachers who graduated from the elementary

teacher program in San Diego State College in order to see how they perceived their college preparation programs. Young (1977) conducted a follow-up study with first year graduates (1976) of the secondary teacher education program at the University of Oregon. The purpose of the study was to provide information for improving the secondary teacher education program.

In planning a follow-up study, the evaluator has to answer these questions: What types of information should he look for? How the information should be collected? Who should be the respondents?

The information that the evaluator has to look for deals with questions about the program such as does the program provide adequate skills and knowledge? Are the graduates satisfied with the programs? How do the graduates perform in the schools? What are the perceptions of the principals toward the graduates who are working with them? What are the perceptions of the students toward their teachers? This type of information should be sought in the follow-up study.

The measuring instruments that should be used and are most often used in follow-up study are questionnaires, interviews, and classroom observations. In the study, the interviewers and classroom observers should be well trained. The results from the interviews and observations should be used to cross-validate with the results from the questionnaires. That will make the results valid and reliable.

The respondents in the follow-up study will be composed of the graduates, the principals and superintendents, the students, and the students' parents.

The problem that the evaluator will be faced with the follow-up

study is to locate the graduates. Because of the problem of locating the graduates, the bias due to nonresponsiveness may be served. Most of the graduates who have success may well be eager to respond while those who have failed may well be less responsive (Cain & Hollister, 1972).

After the data have been collected and analyzed, the evaluator has to make a judgment on the outcomes and provide some suggestions for improving the program. Then, he presents the report to the PPC for discussion and lets the PPC make a decision on whether the program should be improved. After the decision has been reached, another report from the PPC will be presented to the dean of the college for the final decision.

## CHAPTER VIII

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

#### Summary

The purpose of this study is to answer the question of whether an evaluation model can be devised to be used with teacher preparation programs. In order to answer this question, two evaluation models were developed. The framework of these two evaluation models was drawn from the previous studies on evaluation models, evaluation studies, selecting and developing measuring instruments, research design, teacher preparation programs, and related fields, and synthesized in ways which will be precise, specific, and verifiable.

An extensive search for information had been done through the GIPSY, Books in Print, and Government reports.

1. General Evaluation Model (GEM). This model is composed of four stages: setting goals and objectives, planning the program, construction of measuring instruments, and collecting and analyzing data. These four stages are the basic foundation of most of the evaluation models that have been developed recently.

2. College of Education Model (CEM). This model was developed by expanding the concepts from the GEM for the purpose of applying it to the teacher preparation programs. This model is composed of seven stages:

needs assessment, setting goals and objectives, program planning, program implementation, program improvement, summative evaluation, and follow-up study. The discussion of the concepts, evaluation procedures, measuring instruments, and roles of the evaluator and the administrator were presented in each stage.

### Conclusions

Teacher preparation programs in the Southwest area have provided a variety of activities with both direct and indirect experiences for training prospective teachers. Usually, it is not appropriate to use one of the many evaluation models available as a guide for assessing the effectiveness of the programs without any changes or modifications. It would be better if the Colleges of Education would develop their own evaluation models. The models should be flexible enough to be modified and used as a guideline for assessing the effectiveness of the program. And they should be precise, specific, and verifiable. The models also should be oriented toward providing information for decision making.

The College of Education Model (CEM) was developed to serve those purposes. The model provided seven stages: needs assessment, setting goals and objectives, program planning, program implementation, program improvement, summative evaluation, and follow-up study.

Before planning a new program or modifying an existing program, information from the groups of people who will be affected or have been affected should be sought. Needs assessment is the process whereby the evaluator can gather this information. Needs assessment not only provides information on the needs of the groups of people but also provides in-

formation on the effectiveness of the existing program. Basically, the groups of people should be composed of the community members (teacher representatives, principals or superintendents, and the members of local boards of education), the faculty, and the students. The priority needs perceived by the groups of people should be set as priority goals to be pursued. Then need areas have to be transferred into goal statements that should be stated in clear, specific, and measurable terms. Objectives also have to be stated to contribute to the desired ultimate goals. Objectives should be stated in behavioral terms.

When the goal statements have been set, the program can be planned. In the process of program planning, the PPC will provide a variety of alternative instructional methods. The evaluator has to find out whether these alternatives have been studied or used in other institutions before. He also has to predict the likelihood of the success of the program if each alternative is selected. The cost and the benefit of each instructional method also has to be analyzed. In the process of program planning, the PPC has to realize that the prerequisite of planning is forecasting, in which the likelihood of the success is uncertain. Then, the plan should be flexible and be ready to be changed or modified. The success of the program is not only dependent on good planning but also on the process of implementing the program. The program should be implemented according to the plan. If some parts of the plan cannot be implemented, the evaluator has to find out why and provide alternative instructional methods that are comparable to the previous one. If the program has been implemented according to the plan but does not reach the level of expectation, then the necessary changes or modifications have to be made. This

process is called program improvement. The purpose of the program improvement stage is to provide information on which parts of the program need to be changed or modified and how it can be changed or modified in order to make sure that the program is more likely to be successful.

After the program has been completed, a summative evaluation will provide information on whether or not the program has been successful. The information from the summative evaluation will influence the decision of whether the program should be improved, modified, or terminated. It is important that the results from the summative evaluation are valid and reliable. True experimental designs and quasi-experimental designs are suggested for use in summative evaluation. Both of the control well for internal and external validity. Quasi-experimental designs should be used when the situation does not permit randomized assignment. Measuring instruments also have to be selected. Reliability and validity of the measuring instruments are the prime considerations in selecting the instruments. Because results of the summative evaluation have an effect on the survival of the program, the professional evaluator should be hired for conducting the summative evaluation.

The effectiveness of the program cannot be completely concluded when the program had been finished. The college has to find out whether or not the graduates are effective teachers in the schools. The process of a follow-up study will provide the answer to this question. The follow-up study also can be used to substitute for needs assessment.

It is important to mention that the evaluator need not follow through every step of these seven stages, but he has to determine which steps should be followed or modified so they can be suited to the situation



and program.

Unfortunately, the CEM has not yet been used to develop an evaluation plan, but with extensive research on the previous evaluation studies, it is certain that the CEM has the potential of providing a certain degree of success.

### Recommendations

1. Before the CEM can be implemented into the evaluation plan for use in providing guidelines in evaluating the effectiveness of the teacher preparation programs, it should be tried on a limited scale within the program first. The model should have this trial period before it is used for the entire program in order to check the validity of the model. The limited scale use is recommended because it is easier to manipulate, and if the model needs to be changed or modified, it can be done easily.

2. The time for conducting each evaluation stage within the CEM should be set and fixed such as the following examples: needs assessment should be done at the middle of the spring semester, program planning should be done at the beginning of summer, and etc.. Fixing the time schedual for conducting each evaluation stage is important because it permits comparison of outcomes of each stage with the previous outcomes done at the same time of the year. These comparisons will show the discrepancy of the program's effectiveness across time and people. The norm and standard of the program also can be set. It also is helpful to the new PPC to follow the works of the previous PPC.

3. It is assumed that the concepts, evaluation procedures, and measuring instruments provided in the CEM are contemporary, and because

they are contemporary they will need changes and modifications as time goes by. Then, the evaluators who use this model should keep the model up-to-date by implementing new concepts, new evaluation procedures, and new measuring instruments if they are appropriate (precise, valid etc.).

## BIBLIOGRAPHY

- Adams, S. G. Measurement and Evaluation in Education, Psychology, and Guidance. New York: Holt, Rinehart and Winston, 1964.
- Airasian, P. W. Designing Summative Evaluation Studies at the Local Level. In J. Popham (Ed.), Evaluation in Education. Berkeley: McCutchan, 1974.
- Airasian, P. W., Madaus, F. G., & Rakaw, A. E. An Introductory Evaluation Monitoring System. In J. Weiss (Ed.), Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972.
- Alkin, M. C. Evaluating the Cost-Effectiveness of Instructional Programs. CSE Report No. 25, Los Angeles: Center for the Study of Evaluation, 1969. (a)
- \_\_\_\_\_. Evaluation Theory Development. Evaluation Comment, 1969, 2, 2-7. (b)
- \_\_\_\_\_. Wider Context Goals and Goal-Based Evaluation. Evaluation Comment, 1972, 3, 5-6.
- Alkin, M. C., & Fink, A. Evaluation Within the Context of Product Development. In G. Borich (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974.
- Alkin, M. W. The Story of the Eight-Year Study (Vol. 1). New York: Harper and Brothers, 1942.

- Anastasi, A. Psychological Testing (4th ed.). New York: Macmillan, 1976.
- Anderson, B. S., Ball, S., Murphy, T. R., & Associates. Encyclopedia of Educational Evaluation. San Francisco: Jossey-Bass, 1975.
- Anton, T. A. Planning, Programming, Budgeting System for Hillsborough City School District. In L. H. Browder, Jr. (Ed.), Emerging Patterns of Administration Accountability. Berkeley: McCutchan, 1971.
- Apple, W. M., Subkoviak, J. M., & Lufner, S. H. Jr. Educational Evaluation: Analysis and Responsibility. Berkeley: McCutchan, 1974.
- Atkin, J. M. Behavioral Objectives in Curriculum Design. The Science Teacher, 1968, 35, 27-30.
- Ayers, J. B. Tennessee Technological University Teacher Evaluation Model. Cookeville, Tennessee: Tennessee Technological University, 1974. (ERIC Document Reproduction Service No. ED 095 169)
- Ayres, L. P. History and Present Status of Educational Measurements. In G. M. Whipple (Ed.), The Measurement of Educational Products. Bloomington, Illinois: School Publishing, 1918.
- Baer, G. T., & Foster, W. S. Teacher Preparation--What Graduates Tell Us. 1974. (ERIC Document Reproduction Service No. ED 103 402)
- Baker, E. Parents, Teachers, and Students as Data Sources for the Selection of Instructional Goals. American Educational Research Journal, 1972, 9, 403-411.
- \_\_\_\_\_. The Role of the Evaluator in Instructional Development. In G. D. Borich (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974. (a)
- \_\_\_\_\_. Formative Evaluation of Instruction. In J. Popham (Ed.), Evaluation in Education. Berkeley: McCutchan, 1974. (b)
- Ballantine, F., Rowland, M., & Wetherill, W. Perceptions of Elementary

- School Teachers Trained at San Diego State College with Respect to the Value of Several Aspects of Their College Preparation. The Journal of Teacher Education, 1966, 17, 218-223.
- Barr, A. S., & Singer, A., Jr. Evaluative Studies of Teacher Education. The Journal of Teacher Education, 1953, 4, 65-72.
- Beaty, E. Follow-up of Teacher Education Graduates as a Basis for Institutional Improvement. Peabody Journal of Education, 1969, 46, 298-302.
- Bellack, A., & Kliebard, M. H. Curriculum and Evaluation. Berkeley: McCutchan, 1977.
- Bennett, C. A., & Lumsdaine, A. A. Evaluation and Experiment. New York: Academic Press, 1975.
- Berelson, B. Content Analysis. In G. Lindzey (Ed.), Handbook of Social Psychology (Vol. 1). Reading, Mass.: Addison-Wesley, 1954.
- Berrie, P. J. Needs Assessment. Ankeny, Iowa: Area 11 Education Agency, 1976. (ERIC Document Reproduction Service No. ED 128 947)
- Bicknell, J. E., & Others. Formative Studies of First Year of the Competency-Based Field Center for Teacher Education at Hamburg, New York. Fredonia, N. Y.: Fredonia College, 1974. (ERIC Document Reproduction Service No. ED 093 853)
- Bishop, J. L. Staff Development and Instructional Improvement Plans and Procedures. Boston: Allyn and Bacon, 1976.
- Bloom, B. S. (Ed.). Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain. New York: David McKay, 1956.
- Bloom, B. S. Some Theoretical Issues Relating to Educational Evaluation. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Chicago: University of Chicago, 1969.

- Bloom, B. S., Hasting, T. J., & Madaus, F. G. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.
- Borich, G. D. Expanding the Stake Model to Increase Information Yield about New Educational Products. Educational Technology, 1971, 11, 21-23.
- Borich, G. D. (Ed.). Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974.
- Borich, G. D., & Drezek, S. F. Evaluating Instructional Transactions. In G. D. Borich (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974.
- Brantner, S. T. Follow-up Studies: Who Benefits. American Vocational Journal, 1975, 50, 26-27.
- Brieve, F. J., & Johnston, A. P. Meaning and Problems of Planning. School Management, 1973, 17, 34; 36; 38.
- Brittingham, B. E., & Netusil, J. A. The Reliability of Goal Ratings in a Needs Assessment Procedure. The Journal of Educational Research, 1976, 69, 184-188.
- Browder, L. H., Jr. Emerging Patterns of Administrative Accountability. Berkeley: McCutchan, 1971.
- Brown, S. R., & Brenner, D. Science, Psychology, and Communication. New York: Teacher College Press, 1972.
- Burns, R. W. Behavioral Objectives: A Selected Bibliography. Educational Technology, 1969, 9, 57-58.
- Butman, J. W., & Fletcher, J. L. The Role of the Evaluator and Developer In Educational Research and Development. In G. D. Borich (Ed.),

- Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974.
- Cain, G. G., & Hallister, R. G. The Methodology of Evaluating Social Action Programs. In P. H. Rossi & W. Williams (Eds.), Evaluating Social Programs. New York: Seminar Press, 1972.
- Campbell, D. T., & Stanly, J. C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1971.
- Case, C. M. The Application of PERT to Large-Scale Educational Research and Evaluation Studies. Educational Technology, 1969, 9, 79-83.
- Center for Community Needs Assessment. Educational Needs Assessment. Gainesville, Florida: University of Florida, 1973. (a) (ERIC Document Reproduction Service No. ED 095 654)
- \_\_\_\_\_. A Systematic Approach to Educational Needs Assessment. Gainesville, Florida: University of Florida, 1973. (b) (ERIC Document Reproduction Service No. ED 099 024)
- Churchman, C. W. The System Approach. New York: Delacorte Press, 1968.
- Cleland, D. I., & King, W. R. System Analysis and Project Management. New York: McGraw-Hill, 1968.
- Coffing, R. T., & Hutchinson, T. E. Needs Analysis Methodology. Chicago: American Educational Research Association, 1974. (ERIC Document Reproduction Service No. ED 095 654)
- Cook, T. D., Cook, F. L., & Mark, M. M. Randomized and Quasi-Experimental Designs In Evaluation Research. In L. Rutman (Ed.), Evaluation Research Methods. Beverly Hills: Sage, 1977.
- Cooley, W. W., & Lohnes, P. R. Evaluation Research in Education. New York: Irvington, 1976.

- Cronbach, L. J. Course Improvement Through Evaluation. Teacher College Record, 1963, 64, 672-683.
- Cyphert, F. R., & Gant, W. L. The Delphi Technique: A Tool for Collecting Opinions in Teacher Education. The Journal of Teacher Education, 1970, 21, 417-425.
- \_\_\_\_\_. The Delphi Technique: A Case Study. Phi-Delta Kappan, 1971, 82, 272-273.
- Dalkey, N., & Helmer, O. An Experimental Application of the Delphi Method to the Use of Experts. Management Science, 1963, 9, 458-467.
- Davies, I. K. Objectives in Curriculum Design. New York: McGraw-Hill, 1976.
- Davis, J. R., & Morrall, J. F. Evaluating Educational Investment. Lexington, Mass.: D. C. Heath, 1974.
- Downey, L. W. The Task of Public Education. Chicago: University of Chicago, 1960.
- Draayer, F. G. Toward Program Accountability. The Journal of Economic Education, 1974, 5, 97-102.
- Dressel, L. P. Handbook of Academic Evaluation. San Francisco: Jossey-Bass, 1976.
- Educational Psychology. Delma, California: Communications Research Machines, 1973.
- Eidell, T. L., & Kitchel, J. M. Knowledge Production and Utilization in Educational Administration. Eugene, Oregon: The University of Oregon, 1967.
- Eiss, F. A. Evaluation of Instructional Systems. New York: Gordon and Breach, 1970.



- English, F. W., & Kaufman, R. A. Needs Assessment: A Focus for Curriculum Development. Washington, D. C.: ASCD, 1975.
- Engman, B. D. Behavioral Objectives: Key to Planning. The Science Teacher, 1968, 35, 86-87.
- Eurich, A. C. Evaluation of General Education in Colleges. Journal of Educational Research, 1942, 35, 502-516.
- Finch, A. Growth In-Service Education Programs That Work. Englewood Cliffs, N. J.: Prentice-Hall, 1969.
- Findlay, D. C. Application of the CIPP Evaluation Model to a Center with Multiple Program Areas and Levels. Educational Technology, 1971, 11, 43-47.
- Fine, T. W. Implementing a Needs Assessment Program. Educational Technology, 1969, 9, 30-31.
- Fink, A., & Kosecoff, J. An Evaluation Primer. Washington, D. C.: Capitol Publications, 1978.
- Finn, J. D. Evaluation of Instructional Outcomes. In J. Weiss (Ed.), Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972.
- Fitzgerald, P. Assessing the Perceived Educational Needs of Students. Education, 1972, 92, 13-14.
- Fitzgerald, T. P., & Clark, R. M. Process Evaluation for Inservice Training. Kansas City: American Educational Research Association, 1974.  
(ERIC Document Reproduction Service No. ED 102 507)
- Fitz-Gibbon, T. C., & Morris, L. L. Theory-Based Evaluation. Evaluation Comment, 1975, 5, 1-7.

- Flanagan, J. C. The Uses of Educational Evaluation in the Development of Programs, Courses, Instructional Materials and Equipment, Instructional and Learning Procedures, and Administrative Arrangements. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Chicago: University of Chicago, 1969.
- Foley, J. W. The Future of Administration and Educational Evaluation. Educational Technology, 1970, 10, 20-25.
- Foster, W. S. Teachers' Opinions: Their Implications for In-Service Education. The National Elementary Principal, 1966, 45, 49-51.
- Frankel, E. Evaluation of Lehman College Undergraduate Education Program By Teacher Alumni. Bronx, N. Y.: Lehman College, 1974. (ERIC Document Reproduction Service No. ED 092 556)
- Freeman, H. E., & Sherwood, C. C. Social Research and Social Policy. Englewood Cliffs, N. J.: Prentice Hall, 1970.
- Gagne', R. M. The Conditions of Learning. New York: Holt, Rinehart and Winston, 1965.
- Gallagher, J. J., Nuthall, G. A., & Rosenshine, B. Classroom Observation. Chicago: Rand McNally, 1970.
- Galvin, K., & Others. Educational and Retraining Needs of Older Adults. Los Angeles: Institutional Research Association, 1975. (ERIC Document Reproduction Service No. ED 110 132)
- Garvey, J. F. The What and Why of Behavioral Objectives. The Instructor, 1968, 77, 127.
- Georgia State Department of Education. Needs Assessment Package. Atlanta: 1974. (ERIC Document Reproduction Service No. ED 107 695)

- Gephart, W. J. The Phi Delta Kappa Committee's Evaluation Model: One Member's View. In J. Weiss (Ed.), Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972.
- Gilli, A. C., Sr. Follow-up Means Feedback. American Vocational Journal, 1975, 50, 25.
- Glass, G. V. Evaluation Studies Review Annual (Vol. 1). Beverly Hills: Sage, 1976.
- Glass, L. W., & Keith, P. M. The Teacher Education Program--A Liberal Arts Option with a Vocational Choice. College Student Journal, 1975, 9, 319-321.
- Glick, I. D., & Others. Preventing a Second Orthodoxy: A Formative Evaluation of a Competency-Based Teacher Education Program. Toledo, Ohio: Toledo University, 1974. (ERIC Document Reproduction Service No. ED 103 368)
- Gordon, T., & Helmer, O. Report on a Long-Range Forecasting Study. In O. Helmer (Ed.), Social Technology. New York: Basic Books, 1966.
- Gottman, J. M., & Clasen, R. E. Evaluation in Education. Itasca, Illinois: Peacock, 1972.
- Grobman, H. Evaluation Activities of Curriculum Projects. Chicago: Rand McNally, 1968.
- Gross, E., & Grambach, P. V. University Goals and Academic Power. Washington, D. C.: American Council on Education, 1968.
- Grotelueschen, A. D., & Gooler, D. D. Evaluation in Curriculum Development. In J. Weiss (Ed.), Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972.

- Guba, E. G. Development, Diffusion, and Evaluation. In T. L. Eidell & J. M. Kitchel (Eds.), Knowledge Production and Utilization in Educational Administration. Eugene, Oregon: The University of Oregon, 1967.
- \_\_\_\_\_. The Failure of Educational Evaluation. Educational Technology, 1969, 9, 29-38.
- Hage, J., & Aiken, M. Social Change in Complex Organizations. New York: Random House, 1970.
- Harris, C. W. The Appraisal of a School--Problem for Study. Journal of Educational Research, 1947, 40, 172-182.
- Hayman, J. L., Jr., & Napier, R. N. Evaluation in the Schools. Belmont, California: Brooks/Cole, 1975.
- Herbert, J. A Research Base for the Accreditation of Teacher Preparation Programs. Minneapolis, Minnesota: American Educational Research Association, 1970. (ERIC Document Reproduction Service No. ED 040 124)
- Hillison, J. The Development and Try-out of a System of Evaluating Programs of Vocational Teacher Education. Frankfort, Kentucky: 1974. (ERIC Document Reproduction Service No. ED 092 586)
- Hinrichs, H. H., & Taylor, G. M. Program Budgeting and Benefit-Cost Analysis. Pacific Palisades, California: Goodyear, 1969.
- House, R. E. School Evaluation: The Politics and Process. Berkeley: McCutchan, 1973.
- Hovey, H. A. The Planning-Programming-Budgeting Approach to Government Decision-Making. New York: Frederick A. Praeger, 1968.

- Immegart, G. L., & Pilechi, F. J. An Introduction to Systems for the Educational Administrator. Reading, Mass.: Addison-Wesley, 1973.
- Jackson, D. M., & Bidwell, C. E. A Modification of Q-Technique. Educational and Psychological Measurement, 1959, 19, 221-232.
- Jarolimek, J. The Taxonomy: Guide to Differentiated Instruction. Social Education, 1962, 26, 445-447.
- Jones, A. Distributions of Traits in Current Q-Sort Methodology. Journal of Abnormal and Social Psychology, 1956, 53, 90-95.
- Katz, D. S., & Morgan, R. L. A Holistic Strategy for the Formative Evaluation of Educational Programs. In G. D. Borich (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974.
- Kaufman, R. A. Accountability, A System Approach and the Quantitative Improvement of Education--An Attempted Integration. Educational Technology, 1971, 11, 21-25. (a)
- \_\_\_\_\_. A Possible Integrative Model for the Systematic and Measurable Improvement of Education. American Psychologist, 1971, 26, 250-255. (b)
- \_\_\_\_\_. Educational System Planning. Englewood Cliffs, N. J.: Prentice-Hall, 1972.
- Kerlinger, F. N. Q Methodology in Behavioral Research. In S. R. Brown & D. J. Brenner (Eds.), Science, Psychology, and Communication. New York: Teacher College Press, 1972.
- \_\_\_\_\_. Foundations of Behavioral Research (2nd ed.). New York: Holt, Rinehart and Winston, 1973.

- King, M. C., & Others. A Model Process. Los Angeles: University of California, 1974. (ERIC Document Reproduction Service No. ED 097 069)
- Kirt, M. W. The Rise and Fall of PPBS in California. Phi Delta Kappan, 1975, 25, 535-538.
- Klein, S., Fenstermacher, G., & Alkin, M. C. The Center's Changing Evaluation Model. Evaluation Comment, 1971, 2, 9-12.
- Kneller, G. F. Goal-Full Evaluation. Evaluation Comment, 1972, 3, 7-8.
- Knox, A. B. Continuous Program Evaluation. In C. S. Nathan (Ed.), Administration of Continuing Education. Washington, D. C.: National Association for Public School Education, 1969.
- Krathwohl, D. R. The Taxonomy of Educational Objectives--Its Use in Curriculum Building. In C. M. Lindvall (Ed.), Defying Educational Objectives. Pittsburgh: University of Pittsburgh, 1964.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. Taxonomy of Educational Objectives, Handbook 2: Affective Domain. New York: McKay, 1964.
- Larsen, C. J. Summary and Conclusions of a Survey of Residents Residing within a Three-Mile Radius of the Downtown Center--Bakersfield College. Bakersfield, California: Bakersfield College, 1973. (ERIC Document Reproduction Service No. ED 104 503)
- Lewy, A. Handbook of Curriculum Evaluation. New York: Longman, 1977.
- Lindvall, C. M. (Ed.). Defying Educational Objectives. Pittsburgh: University of Pittsburgh, 1964.
- Lindvall, C. M., & Cox, R. C. The IPI Evaluation Program. Chicago: Rand McNally, 1970.
- Loomis, C. P. Social System. Princeton: Van Nostrand, 1960.

- Louisiana State University. LSU General Catalog, 1977-78 Issue.  
Baton Rouge, Louisiana: Louisiana State University, 1977.
- Lutterrodt, S. A. A Systematic Approach to Curriculum Evaluation. Journal of Curriculum Studies, 1975, 7, 135-150.
- McClurg, R. Needs for Integrating Career Education Services Within the Des Moines Area Community College District. Ankeny, Iowa: Des Moines Area Community College, 1975. (ERIC Document Reproduction Service No. ED 122 072)
- McCullough, J. Cost Analysis for Planning-Programming-Budgeting Cost-Benefit Studies. Santa Monica, California: Rand Corporation, 1966.
- McLanghlin, M. Implimentation as Mutual Adaptation. In W. Williams & R. F. Elmore (Eds.), Social Program Implementation. New York: Academic Press, 1976.
- Mager, R. F. Goal Analysis. Belmont, California: Fearon, 1972.  
\_\_\_\_\_. Preparing Instructional Objectives (2nd ed.). Belmont, California: Fearon, 1975.
- Marshall, H. H. Evaluation of Teacher Education Programs. Berkeley: University of California, 1974. (ERIC Document Reproduction Service No. ED 102 159)
- Merriman, H. O. From Evaluation Theory into Practice. Journal of Research and Development in Education, 1970, 3, 48-58.
- Merritt, D. L. Performance Objectives for Student Teaching. Terre Haute, Indiana: Indiana State University, 1972. (ERIC Document Reproduction Service No. ED 065 475)
- Messick, S. The Criterion Problem in the Evaluation of Instruction. In Research Bulletin. Princeton, N. J.: Educational Testing Service, 1969.

- Miller, R. W. Schedule, Cost, and Profit Control with PERT. New York: McGraw-Hill, 1963.
- Montague, E. J., & Butts, D. P. Behavioral Objectives. The Science Teacher, 1968, 35, 33-35.
- Moore, G. R. An Assessment of Community Needs. Salem, Oregon: Chemeketa College, 1975. (ERIC Document Reproduction Service No. ED 121 374)
- Mullen, D. J. Involving Parents, Students, and Staff in Determining a Needs Assessment of Educational Priorities. Anaheim, California: 1974. (ERIC Document Reproduction Service No. ED 093 036)
- Musella, D. Improving Teacher Education. The Journal of Teacher Education, 1970, 21, 15-21.
- New Mexico State University. Bulletin: Undergraduate Courses and Degrees 1976-77. Las Cruces, New Mexico: New Mexico State University, 1976.
- Niemi, J. A. Programs for Culturally Different Adults. 1976. (ERIC Document Reproduction Service No. ED 117 427)
- Northwestern State University. General Catalogue 1977-78. Natchitoches, Louisiana: Northwestern State University, 1976.
- Norton, P. D. Governors State University Needs Assessment Survey. Evanston, Illinois: Governors State University, 1970. (ERIC Document Reproduction Service No. ED 078 014)
- Nunnally, J. C. Educational Measurement and Evaluation (2nd ed.). New York: McGraw-Hill, 1964.
- Office of Information Systems Programs. General Information Processing System (Vol. 2). Norman, Oklahoma: The University of Oklahoma, 1977.



- Ojemann, R. H. Should Educational Objectives Be Stated in Behavioral Terms?. The Elementary School Journal, 1968, 68, 223-231.
- Oklahoma State University. Catalog 1976-77. Stillwater, Oklahoma: Oklahoma State University, 1976.
- Pace, R. C. Evaluation Perspectives: 1968. Los Angeles: Center for the Study of Evaluation, 1968.
- Pallante, J. J. The Delphi Technique for Forecasting and Goal-Setting. NASSP Bulletin, 1976, 60, 86-89.
- Parten, M. Surveys, Polls, and Samples. New York: Harper and Brothers, 1950.
- Patton, M. Q. Utilization Focused Evaluation. Beverly Hills: Sage, 1978.
- Peterson, R. E. The Crises of Purpose: Definition and Uses of Institutional Goals. Princeton: Educational Testing Service, 1970.
- Pffeiffer, J. New Look at Education. New York: Odyssey, 1968.
- Popham, W. J. Educational Needs Assessment. In J. Weiss (Ed.), Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972. (a)
- \_\_\_\_\_. Results Rather than Rhetoric. Evaluation Comment, 1972, 3, 6-7. (b)
- \_\_\_\_\_. Objectives and Instruction. In B. R. Worthen & J. R. Sanders (Eds.), Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, 1973.
- \_\_\_\_\_. Evaluation in Education. Berkeley: McCutchan, 1974.
- \_\_\_\_\_. Educational Evaluation. Englewood Cliffs, N. J.: Prentice Hall, 1975.

- Provus, M. Evaluation of Ongoing Programs in the Public School System.  
In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Chicago: University of Chicago, 1969.
- \_\_\_\_\_. Discrepancy Evaluation. Berkeley: McCutchan, 1971.
- Pyatte, J. A. Functions of Program Evaluation and Evaluation Models in Education. The High School Journal, 1970, 53, 385-398.
- Quade, E. S. Analysis for Public Decisions. New York: American Elsevier, 1975.
- Randall, R. S. An Operational Application of the CIPP Model for Evaluation. Educational Technology, 1969, 9, 40-44.
- Revie, V. A. The Effect of Psychological Case Work on the Teacher's Concept of the Pupil. Journal of Counseling Psychology, 1956, 3, 125-129.
- Rhode Island State Department of Education. Needs Assessment: A Manual for the Local Educational Planner. Providence, Rhode Island: 1975. (ERIC Document Reproduction Service No. ED 106 978)
- Riecken, H. W., & Boruch, R. F. Social Experimentation. New York: Academic Press, 1974.
- Rippey, R. M. Studies in Transactional Evaluation. Berkeley: McCutchan, 1973.
- Rittel, H., & Webber, M. M. Dilemmas in a General Theory of Planning. Policy Sciences, 1973, 4, 155-169.
- Rivlin, H. N., & Robbins, I. Evaluating a Teacher Education Program. The Journal of Teacher Education, 1956, 7, 371-374.
- Rookey, T. J. Needs Assessment: Needs and Goals--Model. East Stroudsburg, Pennsylvania: East Stroudsburg College, 1975. (ERIC Document

Reproduction Service No. ED 106 989)

Rossi, P. H., & Williams, W. Evaluating Social Programs. New York: Seminar, 1972.

Rubin, L. Curriculum Handbook. Boston: Allyn and Bacon, 1977.

Rutman, L. Evaluation Research Methods. Beverly Hills: Sage, 1977.

Sanders, J. R., & Cunningham, D. J. A Structure for Formative Evaluation in Product Development. Review of Educational Research, 1973, 43, 217-235.

\_\_\_\_\_. Formative Evaluation: Selecting Techniques and Procedures. In G. D. Borich (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N. J.: Educational Technology, 1974.

Sax, G. Principles of Educational Measurement and Evaluation. Belmont, California: Wadsworth, 1974.

Scriven, M. The Methodology of Evaluation. In R. Tyler, R. Gagne', & M. Scriven (Eds.), Perspectives of Curriculum Evaluation (AERA Monograph 1). Chicago: Rand McNally, 1967.

\_\_\_\_\_. Evaluation: Noble Profession and Pedestrian Practice. In J. Weiss (Ed.), Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972. (a)

\_\_\_\_\_. Prose and Cons about Goal-Free Evaluation. Evaluation Comment, 1972, 3, 1-8. (b)

\_\_\_\_\_. Evaluating Program Effectiveness or, If the Program is Competency-Based, How Come the Evaluation is Costing So Much?. 1974. (ERIC Document Reproduction Service No. ED 093 866)

Scully, M. G. How to Evaluate Graduate Programs. The Chronicle of Higher Education, 1976, 13, 1; 11.

- Selltiz, C., Wrightsman, L. S., & Cook, S. W. Research Methods in Social Relations (3rd ed.). New York: Holt, Rinehart and Winston, 1976.
- Senf, R. Follow-up Study of 1966 Summer Institutes for Teachers of Disadvantaged Children. New York: Center for Urban Education, 1967.  
(ERIC Document Reproduction Service No. ED 025 476)
- Skelthorne, J. M. The Adult Learner in the University. Toronto, Canada: The Ontario Institute for Studies in Education, 1975. (ERIC Document Reproduction Service No. ED 118 817)
- Smith, C. B., & Farr, R. Evaluation Training: Simulation Exercises. Bloomington, Indiana: The University of Indiana, 1971. (ERIC Document Reproduction Service No. ED 054 917)
- Smith, E. R., & Tyler, R. W. Adventure in American Education (Vol. 3). New York: Harper and Brothers, 1942.
- Sommer, R. "No, Not Research. I Said Evaluation.". APA Monitor, 1977, 8, 1; 11.
- Sorenson, G. Evaluation for the Improvement of Instructional Programs. Evaluation Comment, 1971, 2, 13-17.
- Spooner-Smith, L. Applying Empirical Needs Assessment Procedures to University-Level Instruction in Composition. San Francisco: American Educational Research Association, 1976. (ERIC Document Reproduction Service No. ED 122 294)
- Stake, R. E. The Countenance of Educational Evaluation. Teacher College Record, 1967, 68, 523-540.
- \_\_\_\_\_. Objectives, Priorities, and Other Judgment Data. Review of Educational Research, 1970, 40, 181-212.

- Steele, S. M. Contemporary Approaches to Program Evaluation. Washington, D. C.: Capital Publications, 1977.
- Stephenson, W. The Study of Behavior. Chicago: University of Chicago, 1953.
- Stufflebeam, D. L. A Depth Study of the Evaluation Requirement. Theory into Practice, 1966, 5, 121-133.
- \_\_\_\_\_. The Relevance of the CIPP Evaluation Model for Educational Accountability. Atlantic City: 1971. (ERIC Document Reproduction Service No. ED 062 385)
- \_\_\_\_\_. Should or Can Evaluation Be Goal-Free?. Evaluation Comment, 1972, 3, 4-5.
- \_\_\_\_\_. Alternative Approach to Education Evaluation. In W. J. Popham (Ed.), Evaluation in Education. Berkeley: McCutchan, 1974.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R., Merriman, H. O., & Provus, M. M. Educational Evaluation Decision Making. Itasca, Illinois: Peacock, 1971.
- Sundland, D. M. The Construction of Q-Sorts. Psychological Review, 1962, 69, 62-64.
- Sweigert, R. L., Jr. The Discovery of Need in Education. Journal of Secondary Education, 1968, 43, 345-348.
- \_\_\_\_\_. The First Step in Educational Problem Solving--A Systematic Assessment of Student Benefits. San Dimas, California: California State Department of Education, 1969. (ERIC Document Reproduction Service No. ED 038 740)
- Taba, H. Curriculum Development Theory and Practice. New York: Harcourt, Brace and World, 1962.

- Taba, H., & Sawin, E. A Proposed Model of Evaluation. Educational Leadership, 1962, 20, 57-59.
- Tanner, D. Evaluation and Modification of the Comprehensive Curriculum. The High School Journal, 1971, 54, 313-319.
- Tanner, J. R. The Focus of the Teacher Center in an Educational Renewal Site. 1972. (ERIC Document Reproduction Service No. ED 086 675)
- Thompson, R. H. Where Teacher Education Programs Fail. The Journal of Teacher Education, 1970, 21, 264-269.
- Texas Tech University. Undergraduate Catalog, 1977-78. Lubbock, Texas: Texas Tech University, 1977.
- Torkelson, G. M. An Experimental Study of Patterns for Improving the Preparation of Pre-Service Teachers--In the Use of Audiovisual Materials and of Effects on Pupils. University Park, Pennsylvania: Pennsylvania State University, 1965. (ERIC Document Reproduction Service No. ED 003 513)
- Tripodi, T. T., Fellin, P., & Epstein, I. Social Program Evaluation. Itasca, Illinois: Peacock, 1971.
- Trow, C. Behavioral Objectives in Education. Education Technology, 1967, 7, 6-10.
- Tyler, R. W. Constructing Achievement Tests. Columbus, Ohio: Ohio State University, 1934.
- \_\_\_\_\_. General Statement on Evaluation. Journal of Educational Research, 1942, 35, 492-501.
- \_\_\_\_\_. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago, 1950.

Tyler, R. W. (Ed.). Educational Evaluation: New Roles, New Means.

Chicago: University of Chicago, 1969.

Tyler, R. W., Gagne', R. M., & Scriven, M. (Eds.). Perspectives of Curriculum Evaluation (AERA Monograph 1). Chicago: Rand McNally, 1969.

Uhl, N. P. Identifying Institutional Goals. Durham, North Carolina: National Laboratory for Higher Education, 1971.

The University of Arkansas. College of Education Catalog for 1977-78. Fayetteville, Arkansas: The University of Arkansas, 1977.

The University of New Mexico. General Issue 1975-76, 1976-77. Albuquerque, New Mexico: The University of New Mexico, 1975.

The University of Oklahoma. College of Education Bulletin, 76-77. Norman, Oklahoma: The University of Oklahoma, 1977.

The University of Texas. The College of Education: 1976-1978. Austin, Texas: The University of Texas, 1976.

Valine, W. J. A Four-Year Follow-up Study of Underachieving College Freshmen. Journal of College Student Personnel, 1976, 17, 309-311.

Walberg, J. H. Evaluating Educational Performance. Berkeley: McCutchan, 1974.

Wass, H., & Combs, A. W. A Follow-up Study of On-the Job Behavior of Elementary Teachers Trained in an Innovative and a Traditional Program of Teacher Preparation. Gainesville, Florida: The University of Florida, 1973. (ERIC Document Reproduction Service No. ED 095 170)

- Weaver, W. T. The Delphi Forecasting Method. Phi Delta Kappan, 1971, 52, 267-273.
- Wechsler, H., Reinherz, H. Z., & Dobbin, D. D. (Eds.). Social Work Research in the Human Services. New York: Human Sciences, 1976.
- Weikart, D. P., & Banet, B. A. Planned Variation from the Perspective of a Model Sponsor. In W. Williams & R. F. Elmore (Eds.), Social Program Implementation. New York: Academic Press, 1976.
- Weiss, C. H. (Ed.). Evaluating Action Programs. Boston: Allyn and Bacon, 1972. (a)
- Weiss, C. H. Evaluation Research. Englewood Cliffs, N. J.: Prentice Hall, 1972. (b)
- \_\_\_\_\_. Alternative Models of Program Evaluation. Social Work, 1974, 19, 675-681.
- Weiss, J. (Ed.). Curriculum Evaluation. Toronto, Canada: The Ontario Institute for Studies in Education, 1972.
- Weiss, R. S., & Rein, M. The Evaluation of Broad-Aim Programs. Administrative Science Quarterly, 1970, 15, 97-109.
- Westbury, I. Curriculum Evaluation. Review of Educational Research, 1970, 40, 239-260.
- Whiting, J. F. Needs, Values, Perceptions, and the Nurse-Patient Relationship. Journal of Clinical Psychology, 1959, 15, 146-150.
- Whritner, J. A., & Austin, A. P. Program Planning-Using Management-by-Objectives in School Administration. Englewood Cliffs, N. J.: Prentice-Hall, 1972.
- Wildavsky, A. If Planning is Everything, May be It's Nothing. Policy Science, 1973, 4, 127-153.



- Wiley, D. E., & Boch, R. D. Quasi-Experimentation in Educational Setting. The School Review, 1967, 75, 353-366.
- William, J. B. Formative Evaluation Study for AESP Diagnostic and Pre-scriptive Reading Course. Lexington, Kentucky: 1974. (ERIC Document Reproduction Service No. ED 103 008)
- Williams, W., & Elmore, R. F. Social Program Implementation. New York: Academic Press, 1976.
- Williams, W. G., & Snyder, F. A. The Status of Community College Follow-up. American Vocational Journal, 1974, 49, 40-43.
- Wittenborn, J. R. Contribution and Current Status of Q Methodology. Psychological Bulletin, 1961, 58, 132-142.
- Wittrock, M. C., & Wiley, D. E. The Evaluation of Instruction. New York: Holt, Rinehart and Winston, 1970.
- Woody, C. Nature of Evaluation. Journal of Educational Research, 1942, 35, 481-491.
- Worthen, B. R., & Sanders, J. R. Educational Evaluation: Theory and Practice. Worthington, Ohio: Chales A. Jones, 1973.
- Wotring, J. C. A Pilot Study to Evaluate Teachers Educated at Central Michigan University. Mount Pleasant, Michigan: Central Michigan University, 1972. (ERIC Document Reproduction Service No. ED 064 230)
- Wrightstone, W. J., Justman, J., & Robbins, I. Evaluation in Modern Education. New York: American Book, 1956.
- Yee, A. H., Shores, J., & Skuldt, K. Systematic Flowcharting of Educational Objectives and Process. Audio-Visual Communication Review, 1970, 18, 72-83.

Young, S. Evaluation Procedures for Follow-up of 1st Year Graduate (1976).

Unpublish Manuscript, 1977. (Available from College of Education,  
the University of Oregon)

Yuskievicz, V. D. Educational Needs Assessment: A Systematic Method for  
Determining Educational Need for Compensatory Programs. Scranton,

Pennsylvania: Northeastern Educational Intermediate Unit, 1975.

(ERIC Document Reproduction Service No. ED 120 304)