



1981

An Analysis of the Practice of Educational Program Evaluation in Terms of the CIPP Evaluation Model

Kay M. Smith
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss



Part of the [Curriculum and Instruction Commons](#)

Recommended Citation

Smith, Kay M., "An Analysis of the Practice of Educational Program Evaluation in Terms of the CIPP Evaluation Model" (1981). *Dissertations*. 1997.
https://ecommons.luc.edu/luc_diss/1997

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](#).
Copyright © 1981 Kay M. Smith

AN ANALYSIS OF THE PRACTICE OF
EDUCATIONAL PROGRAM EVALUATION
IN TERMS OF THE CIPP EVALUATION MODEL

BY

KAY M. SMITH

A Dissertation Submitted to the Faculty of the Graduate
School of Loyola University of Chicago in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

October

1980

VITA

The author, Kay Monroe Smith, is the daughter of Charles Monroe and Bess (Murdock) Monroe. She was born February 24, 1942, in Racine, Wisconsin. She is married to Robert Smith and has two sons, Christian and Garrett.

Her elementary education was obtained in the schools of Racine, Wisconsin, and secondary education at the Lutheran High School in Racine, where she graduated in 1960. Her undergraduate work was done at Concordia College, River Forest, Illinois, where, in 1964, she received the degree of Bachelor of Science with a major in education. She completed her master's degree work at Northwestern University, Evanston, Illinois, where, in 1970, she was awarded the degree of Master of Arts in education.

From 1964 to 1970, she taught in the Headstart Program and in elementary schools in Chicago and Riverside, Illinois. In 1979 she taught in the Early Childhood Education Program at Concordia College, River Forest, Illinois. From 1979 to 1980, she served as Coordinator of Clinical Experience and Director of Elementary Student Teaching at Rosary College, River Forest, Illinois. From late 1980 she has taught in the Curriculum Department of Loyola University of Chicago, and in the Education Department of Elmhurst College, Elmhurst, Illinois.

TABLE OF CONTENTS

| | |
|---|----|
| CHAPTER I - OVERVIEW OF THE PROBLEM | 1 |
| Introduction | 1 |
| Statement of the Problem | 2 |
| Purpose of the Study | 4 |
| Significance of the Study | 7 |
| Assumptions | 8 |
| Limitations | 8 |
| CHAPTER II - REVIEW OF THE LITERATURE | 10 |
| Introduction | 10 |
| An Analysis of Educational Program Evaluation in an Historical Perspective | 11 |
| Evaluation by Measurement | 11 |
| Evaluation by Goals | 15 |
| Evaluation by Accreditation | 17 |
| A Transition: The Elementary and Secondary Education Act of 1965 | 18 |
| Contemporary Formal Evaluation Theories | 21 |
| Contemporary Formal Evaluation Models | 22 |
| Goal Attainment | 22 |
| Judgment | 23 |
| Decision Facilitation | 24 |
| An Analysis of the CIPP Evaluation Model | 25 |
| Applications of the CIPP Evaluation Model | 34 |
| Theoretical Applications | 34 |
| Educational Applications | 35 |
| Development of New Models | 39 |
| A Rationale for Examining Educators' Practice of Educational Program Evaluation through the CIPP Evaluation Model | 40 |

| | |
|---|----|
| CHAPTER III - METHODOLOGY | 46 |
| Participants | 46 |
| Data Sources | 48 |
| The Research Instrument | 49 |
| The Validation Procedure | 49 |
| Data Collection | 51 |
| Phase 1 | 51 |
| Phase 2 | 51 |
| Phase 3 | 52 |
| Coding of the Research Instrument | 52 |
| Hypotheses to be Tested and Research Questions to be Addressed | 53 |
| Questionnaire | 53 |
| Interview | 54 |
| Report | 54 |
| Questionnaire, Interview, and Report | 54 |
| CHAPTER IV - RESULTS OF THE STUDY | 55 |
| Introduction: Data Presentation and Description of Sample Districts | 55 |
| Analysis of Questionnaire Results | 58 |
| Introduction | 58 |
| Comparisons within Each District - Null Hypothesis I | 58 |
| Comparisons Within the Group of Districts Null Hypothesis II | 65 |
| Ranking the Evaluation Types - Null Hypothesis III | 70 |
| Summary of Questionnaire Results | 72 |
| Analysis and Summary of Interview Results | 72 |
| Analysis of Report Results | 80 |
| Discussion and Summary of Report Results | 81 |
| Summary Analysis of Combined Questionnaire, Interview and Report Results (QIR) | 81 |

| | |
|--|-----|
| Quantitative Analysis of Questionnaire, Interview, and Report Results Null Hypothesis IV | 82 |
| Qualitative Analysis of Questionnaire, Interview and Report Results | 86 |
| Summary of Results | 91 |
| CHAPTER V - CONCLUSIONS | 93 |
| REFERENCES | 97 |
| APPENDIX A | 103 |
| Research Instrument: First Draft | 104 |
| Research Instrument: Second Draft | 107 |
| Instrument Validation: Background Summary of Curriculum and Instruction Doctoral Seminar Group | 112 |
| APPENDIX B | 113 |
| Research Instrument: Third Draft | 114 |
| Cover Letter for Districts | 118 |
| Final Research Instrument | 119 |
| APPENDIX C | 123 |
| Description of Proposed Research Study | 125 |

LIST OF TABLES

| Table | | Page |
|-------|--|------|
| 1 | Means and Grand Means for CIPP in all Districts | 59 |
| 2 | One-Way Analysis of Variance for Questionnaire Data | 61 |
| 3 | Comparisons of Means of Evaluation Types Reported on Questionnaire | 62 |
| 4 | Number of Evaluation Types Completed by all Districts as Reported on Questionnaire | 66 |
| 5. | Two-Way Analysis of Variance for Questionnaire Data from all Four Districts | 67 |
| 6 | Means and Grand Means for CIPP in all Districts | 68 |
| 7 | Ranking of Evaluation Types for Each District . | 71 |
| 8 | Number of Evaluation Types Completed by all Districts as Reported in the Interview | 73 |
| 9 | Number of Evaluation Tasks Completed by all Districts as Reported in the Written Evaluation Report | 81 |
| 10 | Coefficients of Completion for CIPP in all Districts | 82 |
| 11 | One-Way Analysis of Variance for Coefficient of Completion as Reported in Questionnaire Interview and Written Report | 84 |
| 12 | Comparisons of Means of Evaluation Types as Reported in Questionnaire, Interview and Written Report | 85 |

LIST OF CHARTS

| Chart | | Page |
|-------|------------------------------------|------|
| 1 | Summary of Interview Results | 75 |

CHAPTER I

OVERVIEW OF THE PROBLEM

Introduction

Education in America today is a vast enterprise, involving three-fourths of the nation's population in teaching, learning, and administering the affairs of the school (Ornstein, 1976). The operation of this massive venture is continuously influenced by accelerating social changes.

Education is highly valued as a means for meeting the social, economic, technological, and scientific needs of society as well as the intellectual needs of citizens. . . . Education thus has a most difficult charge which requires the initiation of many innovative programs (Guba & Stufflebeam, 1970, p. 7).

The number of innovative programs developed in response to the social changes of the last decade surpasses the number in operation during the previous five decades. To finance these educational changes, society annually provides billions of dollars through federal and state programs to educational agencies at all levels.

Along with the opportunities for change goes a responsibility to evaluate new programs. Evaluation is not a new concept. The history of formal evaluation goes back to 2000 B.C. when Chinese officials were conducting social service examinations. Program evaluation became a matter of serious consideration in the post-Sputnik era of the late 1950's and the early 1960's. During this period, this country responded to the Soviet penetration of space by launching numerous curriculum reform projects, and with these innovations came the need for new evaluation procedures. In this context, the United States Congress enacted the Elementary and Secondary

Education Act of 1965 (ESEA). This comprehensive educational legislation authorized the expenditure of vast amounts of federal monies for education, with the proviso that each project under Title I and Title III of ESEA be evaluated and a report of that evaluation submitted to the federal government. These limited efforts signaled the beginning of formal program evaluation. However, only a tiny proportion of the country's educational programs were evaluated.

The public cry for accountability in education continued to increase in the 1960's. Federal requirements for program evaluation caused professional associations to establish study commissions to give serious consideration to program evaluation. However, while the need for evaluation was acknowledged, most educational programs continued to function without benefit of genuine evaluation (Worthen, 1973).

The 1970's brought continuing demands for accountability. These demands centered on minimum competency testing, competency-based teacher education, the "back-to-basics" movement, and national concerns over declining enrollment and decreasing revenues. Increasingly, educational agencies were being required to evaluate their programs and report findings in both the academic and financial arenas.

Statement of the Problem

The thrust for fiscal accountability on the national, state, and local level, and the concern for producing competent graduates, have resulted in a national need for program evaluation. It is of paramount importance for schools to determine and report what is being done and how well it is being done. There is an urgent need for program evaluation to accomplish the

assessment and the improvement of educational programs.

Such requirements for evaluation are reasonable and long overdue. Funding agencies and the public have a right to know whether their huge expenditures for education are producing the desired effects. And, educators themselves need evaluative information to be sure the changes they induce are in fact improvements (Guba & Stufflebeam, 1970, p. 7).

In the years following the passage of ESEA, program evaluation was established as an area of inquiry, and in the sixties and the seventies became a sophisticated, well-defined discipline. Evaluation specialists developed evaluation theory and methodology in response to the obvious needs of the times and in order to insure the relevance of future educational programs (Good, Biddle & Brophy, 1975). Researchers in the public and private sectors created the requisite philosophical background for the study of evaluation; prepared definitions and purposes of evaluation, classified evaluation types, and delineated the place of evaluation in the educational program.

Thus, the literature and technology of evaluation proliferated. Broad theories of evaluation evolved from the work of researchers such as Bloom and Scriven. Provus' 'Stakes', and Stufflebeam's contributions were well-developed models designed to encourage evaluation in a variety of settings. More narrowly defined applications of evaluation, such as Planning-Programming-Budgeting System (PPBS), emerged to meet the specific needs of particular types of educational programs.

That there is a need for educational program evaluation as an integral part of educational programming is evident. Further, it is abundantly clear from educational literature that the theoretical sophistication necessary to such evaluation exists. The question remains: Have educators employed this

new knowledge in meeting their increased evaluation responsibilities?

Evaluation experts report that educators have responded in large numbers to requirements for evaluation. The multitude of evaluation reports now available from schools, state departments of education, regional educational laboratories, and educational industries indicates dramatically the significant expenditures of time, effort, and money for the evaluation of educational programs. However, increased activity by itself does not meet the need for effective program evaluation. Many evaluation reports contain only impressionistic information and thus important educational decisions are based on speculation only, rather than on measurable data (Guba & Stufflebeam, 1970).

What is the explanation for this situation? Despite the fact that the conceptual basis for program evaluation exists, why are educators failing to use it and thereby also failing to provide evaluations which are at the same time useful and statistically verifiable? An apparent discrepancy exists between the sophisticated theories of evaluation which became available during the 1960's and 1970's and the actual program evaluation in the local school system. If the potential benefit of program evaluation is to be fully realized, educators must gain an understanding of the nature of the discrepancy between knowledge and practice, and learn to bridge that gap.

Purpose of the Study

The purpose of this study is to gather information on the nature of the gap that exists between the sophisticated theories of program evaluation and the actual practice of program evaluation with the hope that this information will contribute to the literature emerging in the field of educational

program evaluation.

In order to use evaluation theory to analyze the practice of evaluation, the study will have a specific focus. Thus the study will be defined within evaluation itself using a classic evaluation model more narrowly limited to educational programming, and finally confined to a contemporary time and geographic locality. Stufflebeam has made a major contribution to program evaluation theory and practice, and therefore his well-known Context-Input-Process-Product Model (CIPP) will serve as the framework for this evaluation study. Because this research is designed substantively to address the gap between theory and practice, CIPP will be used to assess the actual practice of in-place, completed program evaluations. Further, since a practical emphasis is crucial to the very nature of the study, it will analyze evaluations in which teachers were involved in the process of adopting a textbook series in reading. This scheme provides a link between the philosophical and the practical aspects of evaluation. Because it has a basis in both educational literature and in real world evaluation, the study promises to provide information linking theory to practice.

The CIPP Model is an ideal choice for analyzing program evaluation. Developed in 1971 by the Phi Delta Kappa Committee on Evaluation, it is based on a comprehensive theory that requires a formal, systematic evaluation procedure. It is also useful for this study because it is a practitioners' model - developed by practitioners, for practitioners.

The present plan for CIPP-oriented evaluation research is based on the new definition of evaluation prepared by the PDK Study Committee on Evaluation: "Evaluation is the process of delineating, obtaining, and providing:

useful information for judging decision alternatives" (Stufflebeam, Daniel, Foley, Gephart, Guba, Merriman & Provus, 1971). This definition emphasizes program description, and thus forms the basis for a unique theory of evaluation conceived of as a cyclic, continuing process which must necessarily be implemented through a systematic program. The PDK committee saw the need for a model which would meet the regular information requirements of a system, and at the same time be responsive to the emergent needs for idiosyncratic data. In fulfillment of this need and in keeping with their definition of evaluation as providing information, the committee formulated this total evaluation model.

The specific purpose of this study, then, is to investigate the actual practice of program evaluation. The study analyzed the process of program evaluation conducted by evaluation committees in four elementary school districts. The evaluation committee members were selected from districts that recently undertook a major reading program evaluation, resulting in the adoption of a new basal reading program. These educators were asked to report on their completed program evaluations in terms of the CIPP Model's four evaluation types and corresponding tasks. Information was gathered from the committee through the use of a questionnaire and a group interview. Further, an analysis was made of the final written report prepared by the evaluation committee to determine what evaluation types and tasks the committee chose to report to the board of education. A companion study was undertaken to investigate the perception of program evaluation by curriculum specialists and teachers in the same school districts according to the CIPP framework.

Significance of the Study

This study responds to the need to investigate formal program evaluation by analyzing real-world evaluations. Greenbaum, in his treatise, Measuring Educational Progress, (1977), justifies studying formal program evaluation:

In recent decades there has been an increased interest in establishing systems of social accounting and evaluation to measure progress or its lack in policy areas such as education and health. If society is to become more self-aware and prepare itself through the use of social accounting and evaluation, it must first become more self-conscious about the processes themselves. The potentials, limits, and negative side effects of these processes must be fully understood, for the sake of both those who engage in such work and those who enjoy or suffer its consequences (p. 30).

According to Greenbaum, there is a clear mandate to assess the state of the art in evaluation. This investigation of in-place evaluations answers his call to attend to the needs of consumers and also producers of educational programs.

Other evaluation specialists agree with Greenbaum that analyses of systems of evaluation must be based on an accepted evaluation theory and model. Rose and Nyre, in The Practice of Evaluation, present this compelling premise: "Real world evaluations must be studied in the context of recommended evaluation models and designs in order to facilitate the transition of contemporary evaluation theory and methodology into practice" (Rose & Nyre, 1977).

While school districts do not ordinarily prepare formal evaluation plans according to some designated conceptual framework, the districts considered in this study did conduct formal systematic program evaluations. From the setting of program goals to the final selection of the new basal

reading program, they followed a comprehensive evaluation procedure.

Therefore, this study provides a means for using the classic CIPP model as the measure for assessing completed program evaluations. The combined analysis of this study and the companion study describes educational practitioners' understanding and practice of educational program evaluation, and thus will help us understand the discrepancy between educational evaluation theory and practice.

Assumptions

The following assumptions have been formulated after a review of the literature. These serve to limit the hypotheses for this study. It was assumed that:

1. There is a need for data to describe and delineate the actual practice of program evaluation.
2. A theoretical body of literature exists which can measure the practice of educational program evaluation.
3. Educational program evaluation is a useful means for assessing current practice and for planning educational change.

Limitations

1. The CIPP Evaluation Model has been chosen as the only comprehensive evaluation model to refer to in studying the practice of program evaluation.
2. Program evaluation is limited in this study to the formal process by which a school district selects a new basal reading program. Such a limitation focuses the study and allows comparison across districts.
3. The study will concentrate on the practice of evaluation rather

than illustrating the application of the CIPP Model.

CHAPTER II
REVIEW OF THE LITERATURE

Introduction

The review of the literature relevant to this study is divided into four major areas:

1. An analysis of educational program evaluation in an historical perspective.
2. An analysis of the CIPP Evaluation Model.
3. Applications of the CIPP Evaluation Model.
4. A rationale for examining educators' practice of educational program evaluation through the CIPP Evaluation Model.

The following resources were consulted in searching out current literature:

1. The computerized searches of ERIC; Dissertation Abstracts; and Psychological Abstracts.
2. Research in Education.
3. Education Index.
4. Professional books, journals, and papers related to the topic.

An Analysis of Educational Program Evaluation
In An Historical Perspective

Three great traditions emerged in early evaluation, and have continued to this day. These three methods are: measurement, goal-assessment and accreditation. The history of evaluation has been a constant give and take between these three methods, with one and then another predominating. Even now, when program evaluation has become formal and systematic, we still weave these methods into our theory and practice.

The measurement tradition, which characterized the earliest evaluations, persisted until the 1930's when goal-centered evaluation emerged. During the 1930's, evaluation by accreditation also came into prominence. Not until the mid-1960's, with the advent of ESEA, did formal educational program evaluation become a discipline. Only in the past fifteen years have modern evaluation theory and methodology been in use.

Evaluation by Measurement

The longest period in educational evaluation was the time when the measurement tradition was the predominant theme. From its beginning in China in 2200 B.C., with the administration of civil service examinations, until the widespread introduction of testing in the early 20th century in America, measurement was the major focus of evaluation studies. For thousands of years, then, educational evaluation was defined in terms of emerging measurement technology (Du Bois, 1970).

Chinese officials were examined by the emperor every third year to determine their fitness for continuing in public office. Later, under the Han dynasty, (202 B.C. - 200 A.D.) local authorities introduced written

examinations covering the classical areas of scholarship. Since no university or public school system existed, the civil service examination served as the sole determinant for appointing individuals to public positions (Du Bois, 1970).

The beginnings of formal program evaluation by measurement in the western world emerged in Europe in the middle ages. A system of competitive examinations was introduced in France by Napoleon in 1791, and used in England in 1833 to select trainees for the Indian Civil Service. The successful experience in these European countries led to the establishment of the Civil Service Act in America in 1883. This tradition of administering rigorous civil service examinations as a means for entry into government service has continued into the twentieth century.

School and university examinations began in Europe in the 13th century with the administration of oral law examinations at the University of Bologna. This tradition of using oral examinations as a means for determining eligibility for a degree continued for several centuries, until the introduction of paper and the development of written examinations. In the 16th century the Jesuits pioneered the systematic use of written tests, both for the placement of students and for their evaluation after instruction. During the 1800's, the English universities of Oxford and Cambridge successfully combined the use of oral and written examinations as prerequisites for the BA and MA degrees. By the middle of the 19th century, the use of written examinations was also recognized in the United States as an appropriate basis for such important decisions as: Who should be awarded degrees? Who should be permitted to follow a profession? Who should serve in a government

post? (Du Bois, 1970).

Educational program evaluation by measurement in America followed this tradition of evaluating individuals using written examinations. Joseph Mayer Rice's landmark study of 1887 stood in contrast to the simplistic evaluation studies which preceded it. Rice conducted a comparative study of the spelling performances of 33,000 students in a large metropolitan school system, concluding that student achievement in spelling had no relationship to the amount of time students spent in repetitious spelling drills. Rice's study became a model of educational measurement for the next quarter century (Rose & Nyre, 1977).

Educational program evaluation took on new sophistication in the first two decades of the 20th century. Robert Thorndike, the father of the educational testing movement, helped convince educators that the measuring of human change was important. This new concept of evaluation caught on because the measurement technology for assessing human abilities was just being developed. Standardized intelligence tests and instruments for personality assessment were used by schools, the military, and industry to evaluate students, recruits, or applicants. These sophisticated instruments became the basis for later evaluation studies, (Du Bois, 1970).

The practice of measurement as educational evaluation is evident today in the work of measurement specialists Thorndike and Hagan (1969, 1977) and Ebel (1965). In their writings these researchers contend that measurement is the primary process for securing data for educational decision making. Further, they maintain that the measurement tradition is based on a rigorous, efficient, scientific methodology which yields objective, reliable results

(Popham, 1974).

The measurement tradition has characterized evaluation efforts of school systems, for they have traditionally based judgments of student achievement and program effectiveness on the results of testing information (Popham, 1975 and Womer, 1970). The minimum competency testing movement is further evidence that this tradition remains a central focus for educational decision making. Educational criteria are established and test performance is used to judge individual learners (Pipho, 1978).

Evaluation literature has also set limits for adopting a measurement orientation to evaluation. In this tradition, evaluation is synonymous with measurement, and the role of the evaluator is often confused with that of the psychometrist or test administrator (Borich, 1974). This concept of evaluation is evident in the writings of Thorndike and Hagan (1969, 1971), and Ebel (1965). A broader definition of evaluation is presented by Cooley and Lohnes (1976) and Worthen and Sanders (1973). They maintain that evaluation is a process by which relevant information is collected and analyzed for use in decision making. They contend that evaluation transcends the limits set by research and data collecting.

Several evaluation specialists have criticized the practice of limiting evaluation to measurement. Stufflebeam suggests that measurement is too narrow and inflexible to meet the wide range of needs in evaluation (Guba & Stufflebeam, 1970). Measurement gives evaluation an instrumental focus by which judgements and the criteria for making them are obscured (Popham, 1974). Further, this limited focus ignores the fact that value judgements are a necessary component in all evaluations. Finally, measurement as

evaluation is limited to those variables for which the science of measurement has instruments, ignoring many intangible variables. These factors include sociological, cultural, economic, sociometric, and philosophical influences. The limits of the instrumentation become the real limits of evaluation, resulting in evaluation that is too narrow in its focus and too mechanistic in its approach (Stufflebeam, et al., 1971).

The above position was endorsed by Cronbach in 1963. Focusing evaluation on measurement over-emphasizes testing and ignores many sources of information necessary to evaluation. The principles pertinent to test construction thereby become the principles of evaluation (Cronbach, 1978). Finally, Stake maintains that the value of test data depends upon the professional experience and intuition of the educator using them. He contends that measurement offers useful but limited contributions to a comprehensive program evaluation scheme (Stake, 1978).

Evaluation of Goals

A second evaluation theme, goal-entered evaluation, emerged in the 1930's with the Tyler and Smith Eight Year Study. This study broadened the scope of evaluation to include the use of a variety of data and systematic processes of assessment. The focus of evaluation was on the achievement of program objectives defined in behavioral terms. In the Eight Year Study, Tyler and Smith used a wide variety of tests, scales, inventories, questionnaires, and other measures to gather information about the achievement of curricular objectives. Tyler and Smith's evaluation approach influenced program evaluations for the next 30 years, and even today the Tylerian rationale is evident in program planning and evaluation (Rose &

Nyre, 1977).

This model also has contemporary adherents. Tyler expanded on his theory in Basic Principles of Curriculum and Instruction (1950), presenting a curriculum design model which stresses rationality. Tyler's model is based on the logical relationship which exists among the stated objectives of a program, the means to achieve them, and the "ends" which are evaluated according to the objectives (Tyler, 1942, 1950, 1958, 1964).

More recently Taba (1962) elaborated on Tyler's curriculum-development rationale and included the same objective-based orientation to evaluation. Hammond's evaluation model is also directly based on this goal-centered theme. He developed his model to assess the effectiveness of current and innovative programs by comparing behavioral data with objectives. Today some major evaluation projects, such as the National Assessment of Educational Progress, are firmly rooted in the goal-centered concept of evaluation. Educational publishers and school systems have also adopted this approach in designing educational programs in behavioral terms (Popham, 1975).

A number of authors have specifically criticized the goal-centered approach to educational evaluation. Stufflebeam defines three major limitations in this approach. Initially, it does not attempt to assess the efficiency of objectives. Secondly, it provides only data related to stated objectives, and finally, it yields findings only at the end of the project term. Scriven contends that evaluators must judge both program goals and program results. He points out that if the goals aren't worth achieving, then it is useless to know how well they are achieved (Popham, 1974).

In a similar vein, Bloom (Bloom, Hastings and Madus, 1970) contends

that evaluation should begin with an assessment of the goals of a program. In his criticism of the goal-centered orientation, he claimed that it is not sufficient to evaluate goals against a single philosophy and psychology of education, nor is it sufficient to merely assess congruence between stated objectives and learning outcomes. Popham's criticism (1975) of this evaluation structure maintains that a product-oriented evaluation scheme emphasizes quantitative outcomes and ignores qualitative results.

Evaluation by Accreditation

Also during the 1930's, another major development occurred which has had a continuing impact on evaluation practices: evaluation by accreditation. Formal accrediting agencies and a quasi-evaluation process became a permanent force in American education, and is still the primary means for elementary and secondary school program appraisal (Glass, 1969).

The Educational Testing Service, established in 1947, has been an influential force in accreditation through its large-scale evaluation projects. Bloom's Taxonomy of Educational Objectives: Handbook I: Cognitive Domain (1956), and Krathwohl's taxonomy of the affective domain have provided educators with meaningful guidelines for preparing and evaluating instructional objectives (Krathwohl, 1964).

Large-scale descriptive studies were dominant in the 1950's and 1960's. The Coleman Study (1966) was a cross-sectional analysis of educational opportunities available to minorities across the country. A second study, Project Talent, was conducted by the American Institute for Research to determine the abilities associated with success in various careers. Finally, the National Assessment Program, headed by Ralph Tyler (1964), was designed to

provide information on assessing various procedures for use in evaluation studies (Ornstein, 1977).

Several authors have critiqued this form of evaluation. Stufflebeam maintains that this approach, which relies on professional judgment, is seriously lacking in rigor. The judgments it produces lack reliability and objectivity and are, therefore, not susceptible to ordinary scientific, prudential measures (Popham, 1974 and Stake, 1973). While this professional-judgment approach focuses on a potentially relevant variable, it does not exhibit technical adequacy in measuring this variable and in arriving at defensible judgments. Therefore it is difficult to generalize from these types of evaluations (Glass, 1969).

A Transition Period: The Elementary and Secondary Education Act of 1965

In 1965, the federal government became involved in financing public education with the passage of Public Law 89-10: The Elementary and Secondary Education Act. The ESEA, through its various titled programs, provided for thousands of grants to educational agencies throughout the country. The Act explicitly stated that each local project had to be evaluated in order to continue to receive federal funds (Carter, 1975). Annual evaluations of these funded programs were to be filed with the federal government to insure that federal funds were accomplishing their intended purposes. Thus money was provided for the specific purpose of evaluating educational programs.

What resulted was a massive demand for evaluation methodology to support this effort. Not surprisingly the educational community was not equipped to handle the large number of evaluations required by ESEA (Pasch, 1976). The need created for evaluation theory, technology, and personnel resulted in a

disciplined study of educational program evaluation, a study which is still evolving today.

Evaluation studies conducted as part of the ESEA occurred during a transitional period in educational evaluation. The evaluation requirements of these titled programs followed the three evaluation phases, and predated the beginning of formal program evaluation. In a summary analysis of Title I and Title III evaluations, Browder concluded that these evaluations created as many problems as they solved. The need to meet the evaluation requirements of these programs led educators to obtain the services of evaluators without considering the nature and scope of the evaluation task (Browder, 1973).

The major criticism of the early ESEA-mandated evaluations was that they were conducted as research projects. Guba and Stufflebeam, in a major position paper, specifically addressed this issue. They rejected the proposition that research is equivalent to evaluation, and that the same assumptions and methodologies hold for both fields. "The authors think that many researchers make wrong assumptions about what an evaluation study should accomplish and, based on these erroneous assumptions, researchers foist bad advice upon unsuspecting and unsophisticated practitioners" (Guba & Stufflebeam, 1970, p. 6).

Guba, (1967) writing independently, outlined a series of problems that plagued these early ESEA efforts. The lack of trained personnel was a major reason that the final reports were of poor quality. Further, the United States Office of Education did not provide adequate guidelines for local evaluators pertaining to the type of evaluation needed or the methodology

required. Quoting from Guba: "The present guidelines [are] markedly inadequate: They do little more than to encourage sloppily conceived product evaluation" (Guba, 1967, p. 7). Finally, the deficiencies in ESEA evaluations were evidence that theoretical work in evaluation was almost non-existent. Other large-scale evaluation studies conducted by the federal government during this period were also criticized. In reviewing evaluations of Upward Bound and other programs administered by the United States Office of Education, Guba and Stufflebeam concluded that these evaluations did not provide the information needed to support decision-making related to the programs being evaluated. One major criticism was leveled against these evaluations. Many of the completed evaluations contained only impressionistic information and anecdotal accounts of project outcomes (Guba & Stufflebeam, 1970).

Popham theorizes that these large-scale evaluations were inadequate because they were inappropriately based on classic experimental models. Methodologically, null hypotheses were created, experimental and control groups were studied, and a great wealth of data was collected. However, the process of reporting quantitative data with illustrative, qualitative information resulted in reports which were of little use to the federal government (Rose & Nyre, 1977).

As a result of their investigation of the evaluation studies conducted by the federal government and their analysis of the state of the art of program evaluation in the late sixties, Guba and Stufflebeam concluded:

The lack of adequate evaluation information probably persists because of several fundamental impediments which must be removed before educators can improve their evaluations. These impediments include

the lack of trained evaluators and training programs, the lack of appropriate evaluation instruments and procedures, and the lack of adequate evaluation theory (Guba & Stufflebeam, 1970, p. 8).

Contemporary Formal Evaluation Theories

Along with the methods of evaluation discussed above, specific theories of evaluation have contributed to the field of evaluation literature. Their concepts have provided a theoretical foundation for a clarification of the models. Scriven (1964, 1967, 1974) has clarified the varied aspects of educational evaluation, providing a cohesive framework for the conduct of evaluation. Scriven distinguished between formative and summative evaluation, and between intrinsic and pay-off evaluation. His propositions on the roles and the goals of education were a major contribution to evaluation literature.

Popham (1975) has described evaluation as a holistic, systematic, and adaptive process. Further, he has clarified the role of measurement in evaluation and delineated the relationship of goals and objectives to evaluation.

Benjamin Bloom in his Taxonomy of Educational Objectives (1956) classified objectives in the cognitive, affective and psychomotor domains, and provided a foundation for curriculum implementation and evaluation. Bloom's most recent contribution to literature in this field is the evaluation component in his new theory of school learning, Human Characteristics and School Learning (1978), where he strongly advocates the use of performance standards to facilitate learning for mastery.

Elliott Eisner (1979) proposes that educators expand and broaden the conduct of evaluation. According to Eisner, "Conventional means of evaluation,

particularly the use of achievement tests, are designed to capture only a slender slice of educational life" (p. 20). He contends that evaluation must seek to secure the kinds of information that different measurements make possible. Descriptive information might be gained through the use of non-conventional devices, such as descriptive narratives, films, and interviews, stressing the need for a qualitative element along with the traditional quantitative element in evaluation studies.

Contemporary Formal Evaluation Models

In the years following ESEA, concern over educational evaluation grew, resulting in the development of sophisticated evaluation theory and methodology. Models and strategies were produced to create a new discipline of formal educational program evaluation. Popham (1975) devised a classification system to organize the evaluation models, goal-attainment models, judgmental models emphasizing intrinsic criteria or extrinsic criteria, and decision facilitation models.

Goal Attainment. Goal-attainment models define evaluation as the determination of the degree to which instructional programs' goals are achieved. This concept of evaluation is usually associated with the efforts of Ralph Tyler. According to Tyler's classic framework, educational goals are formulated according to three goal sources --the student, the society, and the subject matter, and two goal screens-- a psychology of learning, and a philosophy of education. Goals are then transformed into measurable behavioral objectives. Attained educational goals reflect the adequacies of the program.

A contemporary version of a goal attainment model is Hammond's

evaluation model. Hammond's study of goal attainment analyzed both institutional and instructional variables and their relationship to measured learner behavior (Popham, 1974).

Metfessel and Michael (1967) also developed a goal attainment model, involving the total community in the evaluation process. The strengths of this model are the different classes of criterion measures employed to provide a comprehensive assessment of factors influencing the goal-attainment of an educational program.

Judgment. Popham designates another class of models: those giving major attention to professional judgment. In these models the evaluator's judgment determines favorable or unfavorable evaluation results. This evaluation approach is subdivided into two categories, depending on whether the evaluator focuses on intrinsic criteria or extrinsic criteria. Intrinsic criteria are defined as process criteria which focus on the inherent nature of the program and extrinsic criteria are product criteria, referring to the effects of the program. Popham notes that judgmental approaches to evaluation emphasizing intrinsic criteria are common in education, but most are too haphazard to be classified as instances of systematic educational evaluation. An exception to this is the accreditation model of evaluation.

There are two major judgment models, both emphasizing extrinsic criteria, one developed by Scriven and one by Stakes. Scriven's model is called the Modus Operandi Method and is specifically designed for use in situations where a judgment must be made about an intervention without the use of experimental or quasi-experimental approaches. The strength of this model is its ability to help determine the characteristic causal chain in an evaluative situation.

Robert Stakes' (1967) Countenance Model is a second judgmental strategy emphasizing extrinsic criteria, systematically analyzing the relation between description and judgment, and focusing on antecedent, transaction, and outcome phases of evaluation. Both judgmental models are based on the presumption that a capable evaluator will be able to make subtle judgments about various aspects of educational programs (Stake, 1973, 1977).

Decision Facilitation. The purpose of decision facilitation models is to provide a framework for servicing the information needs of decision makers. The role of evaluators in this paradigm is to collect and present evaluation information to those individuals who determine the worth of a program.

1. The CSE Model developed by Marvin Alkin (1969) at the UCLA Center for the Study of Evaluation, is the first model in this category. The CSE is an adaptable model which provides for evaluation at any program stage. It is a systematic strategy for providing evaluation reports to decision makers. The CSE Model is implemented through the use of a wide range of evaluation materials and resources.

2. A second decision-facilitation model is the Discrepancy Model devised by Malcolm Provus (1971). The Provus model concentrates on the discrepancies between posited standards and actual program performance. Its purpose is to provide continuous communication between program and evaluation staff through the use of feedback loops, to determine whether to improve, maintain, or terminate a program. The model involves a team in an on-going process of formative and summative evaluation.

3. The CIPP Evaluation Mode, originated by the PDK National Study Committee on Evaluation, (Stufflebeam et al., 1971) approaches evaluation as

a collaborative process between evaluators and decision makers. The CIPP evaluation framework is premised on four decision settings, four evaluation decisions, and four evaluation types. Evaluation is to service the needs of decision making and accountability (Stufflebeam, 1971).

An Analysis of the CIPP Evaluation Model

The Elementary and Secondary Education Act provided an impetus for educational evaluation, an activity which has had an impact on education equal to, if not greater than, the act itself. Evaluation centers were established in various parts of the country to meet the needs created by the passage of this act. One of these centers was developed at Ohio State University under the direction of Daniel L. Stufflebeam.

Stufflebeam and his staff determined that there was a dearth of adequate evaluation information because of limited evaluation theory, inadequate evaluation methodology, and a lack of trained evaluators (Carter, 1975). The committee believed that three problems prevented the educational community from conducting effective evaluation:

1. Their lack of understanding of decision processes and information requirements in current programs of educational change.
2. The absence of a definition of educational evaluation pertinent to emergent requirements for that evaluation.
3. The lack of appropriate evaluation designs (Carter, 1975). To address these problems, the Phi Delta Kappa's Advisory Committee recommended the establishment of a National Study Committee on Evaluation. Committee members came from the Evaluation Center at Ohio State University, the Research and Development Center on Evaluation at UCLA, and EPIC, a Title II Center in

Tucson, Arizona (Gess, 1974). The committee members were: Daniel Stufflebeam (chairman), Walter Foley, William J. Gephart, Egon Guba, Robert Hammond, Howard Merriman and Malcolm Provus. The purpose of the committee's work was to define evaluation, and describe the process according to the conceptual and methodological needs of the evaluation field. The CIPP Evaluation Model was the result.

Understanding the conceptual basis for evaluation is central to this study. Evaluation theories and models emerging in the years following 1965 addressed the following basic needs:

1. The definition of educational evaluation.
2. The delineation of information requirements for educational evaluation.
3. The nature of the educational settings within which evaluations must be conducted.
4. The structure of evaluation systems.
5. The definition of criteria for judging evaluation. The CIPP Model is based on a new definition of evaluation, supported by an appropriate evaluation theory. Further, CIPP's evaluation framework is a two-way matrix describing the change settings and the information requirements for the evaluation. Finally, CIPP is a comprehensive evaluation model which provides a variety of evaluation designs appropriate for application to evaluation efforts in numerous educational settings. Thus, the CIPP model meets the basic needs listed above.

The CIPP evaluation process is based on five principles unique to the theory underlying the CIPP Model:

1. Evaluation provides information for decision making; to evaluate, therefore, it is necessary to know what decisions need to be made.
2. Different evaluation strategies are required to match different decision-making settings.
3. A widely usable evaluation model should take into account the types of decisions and evaluation designs appropriate to many settings.
4. While the content of evaluation designs varies, a single set of steps can be followed in the design of any sound evaluation.
5. Evaluation studies should answer questions posed by decision makers, and should be scientifically sound and practical (Guba & Stufflebeam, 1970).

The CIPP model defines evaluation as "the process of delineating, obtaining, and providing useful information for judging decision alternatives" (Stufflebeam, et al., 1971, p. 40). The primary emphasis in this model is providing information to decision makers. It allows for both decision making and accountability. By maintaining a "record of past decisions and of the information that was available to support them", evaluation aids decision makers in accounting for their past decisions and actions (Ewy & Chase, 1977, p. 3). To serve these needs for change and accountability, evaluation information must meet the scientific criteria of validity and reliability, and the practical criteria of relevance, significance, scope, credibility, timeliness, pervasiveness and efficiency (Wallace & Shavelson, 1970).

In actual application, CIPP uses a five-step evaluation procedure; (a) focusing the evaluation on the questions to be answered and the criteria for answering them; (b) collecting information; (c) organizing information;

(d) analyzing information; and (e) reporting information (Guba & Stufflebeam, 1970).

These steps answer the major evaluation questions: (a) What objectives should be accomplished? (b) What procedures should be followed to accomplish the objectives? (c) Are the procedures working properly? (d) Are the objectives being achieved?

The CIPP evaluation process, then, may be used to assess a wide variety of projects including educational programs, materials and institutions. These various evaluations can yield descriptive and judgmental information about the goals, design, implementation, and results of some specified project.

CIPP is a comprehensive, complex evaluation model, designed for adaptation to both small-and large-scale program evaluations. CIPP's well-grounded theoretical basis contains the following elements: the decision settings - homeostatic, incremental, neomobilistic, and metamorphic; the decision types - planning, structuring, implementing and recycling; and the evaluation types that form the model's name - context, input, process, and product (CIPP).

The first element, the decision-making settings, emerges directly from the authors' definition of evaluation. The extensiveness of an evaluation and the rigor with which it is conducted are determined by the importance of the decision to be made and the availability of information. The decision can range from "small to large change" while the information can range from "low to high". Large changes involve major restructuring in the educational programming of the school. Small changes, however, deal with relatively

inconsequential educational matters.

The four decision settings are called "homeostatic," "incremental," "neomobilistic," and "metamorphic" change, each referring to the extent of the intended change. "Homeostatic" decisions maintain the status quo. "Incremental" decisions concern developmental activities. "Neomobilistic" decisions indicate major innovative activities, and "metamorphic" decisions call for complete changes in school systems.

The second theoretical element in the CIPP model is the group of decision types: (a) planning decisions to determine objectives; (b) structuring decisions to design the procedures used to achieve the objectives; (c) implementing decisions to monitor and refine the procedures; and (d) recycling decisions to judge the outcomes or attainments of the project.

Planning decisions determine the major changes needed in a program. They are concerned with such questions as: What are the conditions preventing the objectives from being achieved? What priorities should the program serve? What new objectives would best serve the philosophy and general goals of the program?

Structuring decisions determine the means used to attain the program objectives. Structuring decisions specify action to implement a program. They prescribe program variables, including content, organization, personnel, schedules, and human and material resources (Gess, 1974).

Implementing decisions are concerned with the procedures for making an educational program work. These decisions deal with such questions as: Should the schedule be modified: Is effective use being made of resources? Are additional personnel needed?

Recycling decisions determine the congruence between the original objectives and the quality of the project attainments. Recycling decisions indicate whether a program should continue, change, be modified, or be terminated. They provide answers to such questions as: Are the students' needs being met through continuing program implementation? Are project problems being solved? Is the project worth the investment of time and money?

The next element, corresponding to each of these four decision types, is the group of evaluation types for which the model was named: context, input, process, and product. The decision-making settings and the evaluation types have an interdependent relationship.

Context evaluation is the most basic and most prevalent evaluation type in education. The major purpose of context evaluation is providing a rationale to justify a particular type of program. Context evaluation determines needs, specifies the population and sample of individuals to be served, and devises objectives to meet these needs. Context evaluation procedures include; (a) defining and describing the environment where the change should occur; (b) identifying unmet needs, along with necessary and available resources; (c) identifying sources of problems or deficiencies in meeting these needs; and (d) predicting future problems by considering the desirable, expected, possible and probable outcomes.

Stufflebeam suggests that context evaluation addresses these questions: (a) What unmet needs exist in the context served by a particular institution? (b) What objectives should be pursued in order to meet these needs? (c) What objectives will receive support from the community? (d) Which set of objectives is most feasible to achieve? Context evaluation thus delineates goals

and objectives in relation to the educational program of a given institution or agency.

Context evaluation includes two processes: the contingency process and the congruency process. The contingency process searches for opportunities and pressures outside of the system to implement improvement within the system, by collecting baseline data on the performance of a program. The congruency process compares actual and intended system performance, reporting discrepancy information concerning the school system's statement of goals, laws and general policies governing education (Gess, 1974).

Input evaluation determines how to use resources to meet the objectives of the program. It identifies and assesses the relevant capabilities of responsible individuals or agencies, strategies for achieving program goals, and designs for implementing a selected strategy. The final product of input evaluation is an analysis of the potential costs and benefits of alternative procedural designs. It is designed to be ad hoc and micro-analytic in comparison to context evaluation, which is essentially systematic and macroanalytic.

Stufflebeam suggests that input evaluation answers five questions: (a) Does a given project strategy provide a logical response to the specified objectives? (b) What potentially relevant strategies already exist for meeting previously established objectives? (c) What procedures and time schedules will be needed to implement a given strategy? (d) How do alternate strategies work under pilot conditions?

Input evaluation information is used to make decisions concerning specific materials, procedures, schedules, and facilities in order to attain

program objectives. In Stufflebeam's words: "It is diagnostic in detecting resource problems to be solved in implementing a selected strategy, yet, it is therapeutic in seeking a solution for a basic problem within the overall system" (Stufflebeam, et al., 1971, p. 224).

Process evaluation, the third of the evaluation types, provides continuing feedback to project directors on how the project is progressing during initiation and implementation. Process evaluation detects defects in the design or its implementation and monitors the various elements of the project, so that potential problems or sources of failure can be identified and remedied. These project elements include: interpersonal relationships among staff and students, communication channels, logistics and adequacy of the resources, physical facilities, staff, and time schedule.

Process evaluation, according to Stufflebeam, should answer the following questions: (a) Is the project on schedule? (b) Should the staff be reoriented or retrained before completing the present project cycle? (c) Are the facilities and materials being used adequately and appropriately? (d) What procedural barriers need to be overcome during the present cycle?

Process evaluation identifies, obtains, and reports information as often as project personnel require such information. In addition to providing feedback for continual program improvement, process evaluation yields a record of the project, which may prove valuable after the project is completed.

The final CIPP evaluation type, product evaluation, provides information about the degree to which goals and objectives have been achieved. It measures and interprets attainments as often as necessary during the project term and at the completion of the project cycle. Product evaluation includes:

(a) identifying congruencies and discrepancies between the intended objectives and the actual achievements, (b) identifying unintended outcomes, (c) providing for objectives that have not been met by recycling the program, and (d) providing appropriate information to decision makers about the future of the program - whether it should be continued, modified, or terminated.

All four evaluation types: context, input, process and product can be considered formative when they provide information for program improvement, and summative when they provide information for decisions regarding a program's future.

Not only does the CIPP model aid decision making, but it also provides a measure for accountability. In "The relevance of the CIPP Evaluation Model for Educational Accountability", Stufflebeam defines accountability:

"Accountability means the ability to account for past actions, the wisdom of those decisions, the extent to which they were adequately and efficiently implemented, and the value of their effects" (Stufflebeam, 1971, p. 13).

Each of the four evaluation types helps measure accountability. Context evaluation provides educators the information to identify their objectives and the rationale for those objectives. Context questions which pertain to accountability include: What are the objectives? What assumptions do they follow? Are they morally, socially, and scientifically valid? Input evaluation provides a record of chosen strategies and designs, as well as reasons for their choice: What kind of information was available? Were alternative designs and strategies explored? Why was a particular project design chosen? Process evaluation records the actual implementation process, answering two basic accountability questions: Was the design successful or not? If it

was not successful, was it because the project design was never implemented, or was it because the design, though implemented, was inadequate to achieve the desired outcomes? (Gess, 1974).

Finally, product evaluation records results and decisions about procedures. If a project was continued, on what basis was that decision made? Were modifications made during the implementation process? If the procedure was terminated, was the decision made on legal, moral, or educational grounds? (Stufflebeam, 1971). The CIPP Evaluation Model, then, provides a sound accountability system based on continuing efforts for change in an educational system (Stufflebeam, 1971).

Applications of the CIPP Evaluation Model

The theoretical sophistication presented in the CIPP Evaluation Model has been applied to a variety of educational programs, falling mainly into three categories: theoretical applications of the model, applications of the CIPP model to existing educational programs, and the development of new evaluation models from the CIPP model.

Theoretical Applications

In the process of analyzing its effectiveness in theoretical situations, researchers have at some times dealt with the full-scale model and at other times addressed selected, pertinent elements of the model. Dennis Hinkle's 1971 dissertation, entitled, "The Conceptualization of the Stufflebeam CIPP Evaluation Model in a Multivariate Context", applied the model to a hypothetical and complex educational change activity. He found that the thoroughness inherent in the model required a painstakingly careful application, concluding that using the model would yield information that was valid,

reliable, timely, pervasive, and credible.

In a study conducted at the Ohio State University Evaluation Center (1973), CIPP was used to develop a technical manual for conducting evaluation through the use of advocate teams. These teams used the framework for developing alternative evaluation strategies (Reinhard, 1973).

CIPP was the evaluation model for a systematic approach to sex-role stereotyping and evaluation prepared for the National Education Association. This study (1973) examined the usefulness of applying CIPP to school systems or projects where social change was needed. The study visualized a school system and examined areas where institutionalized stereotyping would occur (McClure, 1973).

Finally, CIPP was chosen as the paradigm to assess the evaluation priorities of students, teachers, and principals in a study conducted by Stufflebeam and Nevo. They attempted to identify these evaluation needs and to develop recommendations for an evaluation system within a school building. The study illustrated that while information on outcomes is the most available evaluative information, school people show a great concern for other kinds of evaluative information, namely that provided by context, input and process. The comprehensive nature of the CIPP framework made it ideally sensitive to these emerging evaluation attitudes and needs (Nevo & Stufflebeam, 1975).

Educational Applications

The CIPP model has also been applied to program evaluation in elementary and secondary education, post-secondary education, and related educational agencies which support the work of the schools.

In elementary and secondary program evaluations, the CIPP model was used in total or in part to facilitate evaluation of educational programs, including individual learning strategies, courses of instruction, curricular content areas, and the educational program of a school system. One example of the use of CIPP to evaluate individual learning strategies involved methods of teaching literature to twelfth grade students in a midwestern high school. CIPP was effectively used to conclude that a particular technique of teaching, rhetorical stance, has a positive effect on students' cognitive recall of factual knowledge in a literature course (Blakely, 1973). CIPP was also used to assess the merits of an educational innovation implemented on a school-wide level. The model served as the framework for examining the use of paraprofessionals at Parma (Ohio) Schaaf Junior High School. The study recommended improving the training system that assists instructional aides and the faculty persons who supervise them (Pasch, 1976). In 1976, CIPP was used to evaluate a course of instruction in the Appalachian Maryland Experience Based Career Education Project (EBCE). The study concluded that ECBE students showed impressive academic and attitudinal progress as a result of the project (Stead, 1977).

Also in the elementary and secondary levels, CIPP has shown evidence of both breadth and adaptability in evaluating curricular content areas at the system-wide level used to restructure the program evaluation plan in the Saginaw Public Schools. In this study, administrators reported that CIPP provided information to answer four basic questions: (a) What should we do? (b) How should we do it? (c) Are we doing it as planned? (d) Did the program work? (Taylor, 1974). CIPP was also applied to the educational program of

a state-wide school system to develop information materials for implementing local educational improvement plans. Evaluation packets were developed as part of Project Access to aid Colorado teachers, educational administrators, and school-board members in implementing evaluation plans as mandated by Colorado's Educational Accountability Act of 1971. In this large scale evaluation effort, CIPP was used to evaluate both the program and the resources. Further, CIPP proved to be a true "practitioner model," for the evaluation materials it engendered were designed for use by educators who were not formal evaluators (Ewy & Chase, 1977).

The CIPP model has also shown its practical utility in a wide variety of post-secondary educational settings. The program evaluations employing the CIPP design have been used at a community college, at the college and professional school level, and in assessing far-reaching state-wide programs.

The usefulness of CIPP was proven in evaluating an established career education program at Moraine Valley Community College in Illinois. The major objective of the evaluation was to establish procedures for enhancing the impact of evaluation results on institutional decision making. The study concluded that CIPP could be modified for evaluating a single program in a small school (Hecht, 1977).

Professional schools have also applied the CIPP model. The Loma Linda University School of Dentistry used it to evaluate a dental team training program involving the use of paraprofessionals. The Training in Expanded Auxiliary Management (TEAM) program found that CIPP provided a useful and viable evaluation method for implementing educational decisions (Reeves & Michael, 1973).

Two CIPP applications have been reported at the college and university level, both in the area of teacher education. A consortium of the state universities of Ohio designed a comprehensive undergraduate inservice model for the preparation of elementary school teachers to deal with the forces of societal and educational change. CIPP was especially beneficial in this study because it could be applied directly to classroom practice by both individuals and groups of learners (Summary of Educational Specifications for a Comprehensive Teacher Education Program, 1968). In a second teacher training application at the University of Pittsburgh, CIPP was applied in the evaluation of a graduate training program for educational research and development personnel. The evaluation study concluded that the CIPP model served a useful base for an overall design in a program which was relatively undefined (Woodwar & Yaeger, 1972).

A program evaluation conducted in support of the Pennsylvania Adult Basic Education Improvement Program, by the Continuing Education Division of Pennsylvania State University illustrated the role of evaluation in organizational development. The evaluation report concluded that the CIPP evaluation paradigm provided an approach to evaluation consistent with and supportive of the organizational development framework (Barnette, 1977).

CIPP's adaptability is substantiated by its application to evaluations in related educational agencies. A format for planning and evaluating statewide library services was formulated with the CIPP mode. Under a grant from the Illinois State Library, a manual as prepared in 1973 as a reference guide for 18 library systems to use in their five-year planning and evaluation programs. The Illinois libraries used this guide in various ways in

accordance with their specific needs. Thus, each library system was permitted maximum autonomy in the development of their programs and evaluation plan (Michael, 1976).

Development of New Models

Numerous CIPP applications have helped create new evaluation models. The first of these, a practical application, was the New Rochelle Evaluation Model, an eclectic framework developed for a district-wide program evaluation. The second and third applications were theoretical, being developed in dissertations, the second adapting CIPP for use in adult education, and the third modifying CIPP to assess school staff and school community involvement in the evaluation of local educational programs.

The New Rochelle Evaluation Model was district-wide. The model was used to evaluate district-wide reading programs and programs for children with special needs. The practicum committee concluded that CIPP provided the following features to meet the district's evaluation needs. It analyzed the district's characteristics, and provided a format to assess the data needs for decision making (Gess, 1974).

The first theoretical adaptation was designed to serve in the planning and implementation of adult education programs, by providing a conceptual framework and feedback networks for maintaining communication at all levels of evaluation (Shiplett, 1974).

A second theoretical adaptation was created to enable public and professional involvement in the planning and evaluation of educational programs. A new condensed evaluation model increased the extent of involvement and improved both the quality of decisions made and the climate of community

relationships.

The successful application of the CIPP model to theoretical evaluation situations, to actual program evaluations, and to the creation of new evaluation models, establishes CIPP's potential for use in varied educational programs across numerous settings. The CIPP applications cited show that the decision-making process is central to administrative procedures. CIPP provides a recognized systematic manner for making these sound decisions. The CIPP model, if properly applied, improves the quality of decisions and also the quality of programs, as well as identifying alternatives and making valid and reliable information accessible.

A Rationale for Examining Educators Practice
of Educational Program Evaluation Through
the CIPP Evaluation Model

That there is a continuing need to study the practice of program evaluation is evident. It now remains to explain why this present study employs the CIPP model to study the practice of and participation in program evaluation in order to adopt a new textbook in reading.

CIPP is uniquely suited to the investigative purposes of this study because it is balanced between the theoretical and practical. While CIPP is a comprehensive evaluation model with a strong philosophical basis, it is also a model which provides for direct application to program evaluation at the local district level. CIPP also facilitates both formative and summative evaluation to promote change and enhance accountability.

The actual means for assessment of program evaluations are inherent in CIPP's design. CIPP is best implemented by using its evaluation types and

tasks to analyze in-place program evaluation. Educational literature makes a strong case for the use of CIPP to study educators' practice of program evaluation in the elementary, secondary, post-secondary, and related educational settings.

This study will focus on educational practitioners' participation in program evaluation for the adoption of the new program. The role of local school personnel in the adoption of educational materials has varied in recent educational history. While the period of the fifties was characterized by the planning and evaluation of curriculum projects on a national level, the current emphasis is the involvement of school personnel at the local district level. The case for including practitioners in program evaluation at the local level has been well-substantiated and clearly defined. The basic assumption of the involvement approach to evaluation is that the most satisfactory, the most useful, and the most valid evaluations are those in which school people have had major input. Research has shown that the practice of involving program and administrative staff in evaluation studies eliminates many problems in disseminating results and increases overall study credibility. Further, staff involvement is likely to be the most effective staff development activity possible to improve awareness of evaluation's benefits. Input into the development of the evaluation design, the collection of data, and the interpretation of results are means by which teachers and administrators can be significantly involved in an evaluation effort (Novak, 1970).

Numerous studies have reported beneficial results from school personnel's participation in evaluation. Cohen reported in 1976 on a program

in which educators participated in a "collegial evaluation system" to create and maintain a team approach to evaluation. This cooperative system responded to the growing complexity of instructional materials available in elementary school. Nolin, in his 1976 dissertation, applied the CIPP model to study the involvement of various groups in program evaluation in four school districts in Los Angeles. Here the CIPP model was adapted to meet public, legislative, and professional desires for involvement of the school staff and school community in program evaluation (Nolin, 1976).

The textbook selection process is a natural subject for a study of this type. The practitioner in the field faces a wider choice of instructional materials every year. This multitude of available materials is a mixed blessing. Harriet Talmage has observed that the selection of educational materials has become the "Russian roulette of education" (Talmage et al., 1977). Faced with numerous competing products offered by enthusiastic salespeople, educators often make decisions without adequate guidelines for analyzing, judging and selecting materials. Under these circumstances, selection often rests on personal whim, persuasiveness of the company representative, and comparisons with other familiar materials, none of which is a rational or scientific approach to instructional improvement through the use of better quality educational materials (Eash, 1969).

Textbook selection has not always been the prerogative of the school or the teacher. Historically, the textbook selected by the teacher has played a critical role in determining curricula. In fact, the McGuffey Reader of a century ago had three roles: it was instructional material, it dictated instructional strategies and approaches to grouping, and it was the curriculum.

During the 1920's and 1930's, a mounting concern among groups of teachers about the purpose of education emphasized the philology of education as the necessary criterion for textbook selection. In the late 1930's and 1940's, the distinction between curriculum and instruction emerged. Textbooks were selected during this period to assure attainment of a school system's objectives. Here for the first time the hierarchical relationship between the curriculum and the textbook was reversed: curriculum decisions were no longer exclusively in the hands of teachers. As the trend toward curriculum development accelerated in the 1940's, teachers became overwhelmed by administrators, supervisors, and specialists in the selection process (Talmage et al., 1977.) However, this comprehensive involvement was reversed with the launching of Sputnik. A new philosophy of education based on the nation's international commitments resulted in the formation of curriculum committees on the national level to produce completed products for classroom use. The involvement of school personnel in curriculum planning was bypassed and once again the textbook and related materials became the curriculum. The school and community thereby lost their responsibility of defining the school's philosophy of education, leaving the philosophy to the instructional materials, specifically the textbooks. Currently the textbook and instructional materials have assumed one of two roles. One trend emphasizes the dominance of the textbook and the instructional package. A second trend emphasizes the role of school personnel and other concerned parties in making curricular and instructional decisions, which are then implemented through textbooks and instructional materials.

Talmage complains that the practice of textbook selection is plagued by

misleading publishers' blurbs, invalid and unreliable rating scales, the biases of outspoken teachers, perfunctory acts of ad hoc selection committees, influential partisal parent groups, and unethical arrangements between publishers and superintendents (Talmage, et al., 1977).

Financial considerations multiply the effects of these factors.

According to Talmage's analysis: "With instructional budgets cut to the bone, there is little room for error in the selection of materials. For a large school district, a major series for adoption could easily run from \$60,000 to \$300,000 or more, excluding implementation costs (Talmage, et al., 1977, p. 1). Thus the importance of the textbook adoption process and the existence of varied and unworkable adoption procedures are evident.

Several practical considerations further substantiate the choice of textbook adoption as the focus of this study, rather than the evaluation of a specific instructional program. Any comparative study of existing programs across school districts would be nearly impossible. Evaluations of existing programs differ in orientation, philosophy, and scope. Moreover, they are conducted at numerous times during the program period to serve a variety of purposes, and employ assorted evaluation methods. Program evaluations designed for textbook adoptions, however, have the same orientation and scope, are conducted during roughly the same period, and have a similar goal, namely: the selection of a new basal program. Therefore, textbook selection is an appropriate means for comparing a number of program evaluations.

Textbook adoption has been chosen as the focus of this study for a second practical reason. Existing programs are often subjected to ad hoc evaluations which differ in quality and quantity from program to program.

However, the long-range consequences of the adoption of a new basal program require implementing a formal evaluation procedure, including an evaluation committee and a final evaluation report. The process of textbook adoption, then, is a procedure which is appropriate to the nature of this study.

Finally, districts define textbook adoption as an evaluative procedure and recognize its importance in planning the program of the schools. Therefore, the commonality of the textbook adoption process itself provides a means for studying program evaluation in educational settings which differ in size, available resources, socio-economic make-up and philosophy of education.

The subject area of reading is also a natural choice for an evaluation study of this type. Reading is the content area most often chosen for study in school districts, and therefore reading program evaluations are readily available for study. Secondly, the study is confined to reading in order to keep the content area constant across districts. Finally, because reading itself is central to the learning process and at the same time transferable to all other content areas, the process of selecting a basal reading program is likely to be a widely influential task.

CHAPTER III

METHODOLOGY

This study analyzed the practice of educational program evaluation, as reported by district evaluation committees, through questionnaires, interviews and final evaluation reports. The instrument used was designed according to the four evaluation types (context, input, process and product), and the evaluation tasks appropriate to each evaluation type, as outlined in the CIPP evaluation model. The study was conducted in Chicago area elementary school districts.

Participants

Limiting the study to certain participants was necessary to give it both purpose and direction. The following criteria were used to identify potential districts which might serve as subjects for the study:

1. The school district should be within a 30 mile radius of downtown Chicago. The Chicago Public School System was excluded because it is atypical of most systems because of its size and the manner in which it evaluates educational programs.

2. The school district should employ a minimum of 200 certified elementary teachers. A district of this size is necessary to insure an adequate number of curriculum supervisors for study.

3. The school district should have completed a formal program evaluation in reading within the last three years to adopt a new basal reading program. For the purpose of this study a formal program evaluation in reading is defined as a process by which a committee specifically charged with the

task of studying alternative basal reading programs makes a recommendation to select a basal program according to a predetermined set of criteria.

4. The program evaluation committee in reading should be predominantly composed of classroom teachers.

The geographic boundary was imposed to isolate a group of municipalities which are suburbs of Chicago. The larger districts were chosen for study because only districts of this size would have an adequate number of specialists to conduct a comprehensive program evaluation large enough to be useful to the study.

The Directory for Illinois Schools, 1979, was consulted to identify those elementary school districts within a 30 mile radius of downtown Chicago, and which employ a certified staff of at least 200. The use of these criteria yielded a list of 20 possible districts, out of 60 districts which might be candidates for the study.

A second list of criteria was compiled to aid in selecting appropriate districts for the study. Telephone interviews were conducted with the assistant superintendent for instruction in each eligible district. The following list of questions was used for the interview:

1. Has the school district completed a program evaluation in reading within the last three years to adopt a new basal reading program?
2. Did the school district engage in a formal process of evaluation - that is, did a committee consider other basals according to predetermined objectives?
3. Did the school district's evaluation committee recommend a basal program for adoption in a written report of its decision?

4. Was the school district's evaluation committee predominantly comprised of elementary classroom teachers?

This telephone survey yielded the following results. Four suburban school districts responded positively to the queries, and agreed to participate in the study.

Data Sources

Ordinarily many groups are involved in adopting a new basal reading program, namely school personnel, including administrators, supervisors, and teachers; the school board; and the various groups that the school serves, pupils, parents and the general public. The school group most actively involved in the process of program evaluation is the evaluation committee. Therefore, the evaluation committee itself and its final evaluation report were the two primary sources of data for the study. The evaluation committee was the first source because it involved administrators, curriculum supervisors, resource specialists, and classroom teachers, who either volunteered or were appointed to serve on the evaluation committee. The work of the committee was analyzed over a period of time from the initial formation of the committee through implementation of the new program. Their efforts as part of this evaluation team gives them expertise in program evaluation in general and familiarity with the specific operation of their own program evaluation. Therefore, they were the logical group to report on their district's evaluation program. The final evaluation report is the written document produced by the committee and the second important source of data. It is the summary of the committee's work, including their report and recommendations to the board of education.

The Research Instrument

Format of the Instrument

The research instrument is based on the four types of evaluation contained in the CIPP Evaluation Model. Each type of evaluation is defined by its purpose as follows:

Context Evaluation. Purpose: To provide a rationale for determining program objectives.

Input Evaluation. Purpose: To provide information to determine how to use resources to meet program goals.

Process Evaluation. Purpose: To provide periodic feedback to persons responsible for implementing the new program.

Product Evaluation. Purpose: To measure and interpret results during the implementation and duration of the program. (The tasks identified for each type of evaluation are drawn from Stufflebeam, Educational Evaluation and Decision Making, 1971.)

The research instrument was used in three distinctive procedures to gather data for the study: as a questionnaire for the members of the evaluation committee, as a format for interviewing several committee members, and as a checklist to analyze the final evaluation report.

Validation Procedure

The research instrument was subjected to a careful three-step validation procedure, designed to provide a natural, logical plan for testing the instrument. The validation groups included teachers and administrators with varied experience in program evaluation.

The first draft of the research instrument was reviewed by a seminar

group of doctoral students in curriculum at Loyola University, who are presently working on doctoral dissertations. This group of teachers and supervisors was included, because they have become expert in curriculum development and evaluation as part of their professional experience. The following changes were made in the second draft of the instrument as a result of this group analysis: Educational jargon was removed from the task questions; the order of the task questions was altered to provide a more logical sequence; and the language in the instrument was clarified. (See appendix A for the first draft and the second draft of the research instrument and a background summary of the curriculum and instruction doctoral student group.)

The second draft was submitted to an evaluation specialist to establish content validity. This expert clarified the directions for the questionnaire, proposed to include an explanation for each type of evaluation, and created additional task questions by separating existing task questions.

The third draft of the instrument was then field tested in a suburban elementary school by a group of teachers and administrators who had evaluated programs in their district. These educators were able to complete the instrument unaided in a short period of time. They suggested that a cover letter accompany each questionnaire to instruct respondents to answer the task questions based on the entire committee's work, rather than based on their individual efforts. (See Appendix B for the third draft of the research instrument and the final research instrument.)

This validation process was most beneficial, resulting in an instrument with a more structured format, which was easier to use. Language and general readability were improved and the questions progressed more logically. As a

result of the validation, the evaluation instrument finally sent to the participants was easily usable.

Data Collection

A preliminary meeting was held with the assistant superintendent for curriculum in each district, in order to provide him/her with an overview of the study, to review data collection logistics, and to secure permission to conduct the study. In each meeting, specific steps were followed:

1. The researcher explained the reasons for each district's selection, outlining the selection criteria described above.

2. The researcher presented each assistant superintendent with an explanatory document which included a description of the proposed research study and a summary of staff involvement for the participating school district. (See Appendix C for the research proposal presented to participating school districts.)

3. The researcher then requested approval from the district to conduct the study.

After the researcher received approval to conduct the study in each district, she used the following procedures to collect the data:

Phase 1

The list of Evaluation Committee members was secured from each assistant superintendent. They were returned to the researcher in self-addressed envelopes.

Phase 2

Several evaluation committee members were chosen at random from the

list of committee members. These members were contacted by telephone and asked: Would you participate in a group interview with other members of the evaluation committee to discuss how the reading program evaluation was conducted in your district? Those present at the interviews included evaluation committee members selected for the interview, the researcher, and a fellow graduate student who is the author of the companion study identified earlier. In the interview, the researcher asked how the tasks were completed. To increase reliability in this phase of the data collection process, minutes were taken of each district interview by the author of the companion study.

Phase 3

Each district's evaluation report was analyzed by the author to ascertain which of the 40 CIPP evaluation tasks were specifically documented in the evaluation report. These tasks were tallied on a copy of the research instrument.

Coding of the Research Instrument

The research instruments which were completed by the evaluation committee members were coded with a three digit identification number. The first digit identified the school district. Districts were coded as 1,2,3,4 respectively. The second and third digits represent a sequential numeration of research instruments as they were sent to evaluation committee members. The coding appeared in the upper left hand corner of each research instrument as follows:

(District
1,2,3,4)

(Sequential Numeration
01,02,03...)

Hypotheses to be Tested and Research

Questions to be Addressed

The following null hypotheses were tested:

Questionnaire

1. Ho: There is no significant difference in the frequency of the tasks reported for each evaluation type (CIPP) within each district.

Analytical Techniques: One-way analysis of variance (ANOVA), Scheffe's Test on the difference of sample means.

2. Ho: There is no significant difference in the frequency of tasks reported for each evaluation type (CIPP) across all districts.

Analytical Techniques: Two-way analysis of variance, Scheffe's Test on the difference of sample means.

3. Ho: There is no significant difference in the ranking of the evaluation types in the group of districts.

Analytical Technique: Kendall's coefficient of concordance.

Questionnaire, Interview and Report

4. Ho: There is no significant difference in the coefficient of completion for each evaluation type across districts, as reported in the questionnaire, interview and report.

Analytical Techniques: One-way analysis of variance, Scheffe's test on the difference of sample means.

The level of significance set for each of the above analysis is .05 and .01.

The following research question was answered:

Interview

1. What evaluation tasks are most frequently reported by all four districts?
2. What methods did each district use in completing the evaluation tasks?

Report

3. What tasks are most frequently reported in the evaluation across districts?

Questionnaire, Interview and Report

4. What unique themes emerged in each district's completion of the program evaluation?

CHAPTER IV

RESULTS OF THE STUDY

Introduction: Data Presentation and Description of Sample Districts

The format for reporting the results of the study follows the data-gathering sequence used in the study. To analyze data within each district and across districts, the data from the questionnaire (Q), the interview (I), and the report (R), are reported in quantitative and qualitative terms. In addition, summary information from the three investigative procedures (Q,I,R) is analyzed and synthesized to create a composite picture of the practice of program evaluation as reported in all four districts. This method of organization meets the following three goals for presenting the results of the study:

1. To present individually the results from the three data sources;
2. To discuss both the statistical and qualitative results; and
3. To analyze and compare the results from each of the four districts.

These results will show the degree to which all districts practice program evaluation according to the CIPP framework and will also describe the unique practices of program evaluation used in each district.

District Descriptions

Four districts in the Chicago area met the specified criteria for the study. This sample group is made up of large school districts within a thirty mile radius of downtown Chicago which have, in the past three years, completed a comprehensive program evaluation in reading in order to adopt a new

basal program. Not only was the similarity among the districts important to the study, but also the distinctive nature of each district enhanced the interest and significance of the results.

The four districts studied represent the gamut of suburban communities surrounding Chicago. The districts and communities they serve differ in the following substantive ways: geographic location, socio-economic make-up, special community features and problems, and acknowledged purpose for initiating a program evaluation. A detailed description of each district follows:

District 1

District 1, a large suburban school district located to the far northwest of Chicago, serves a middle and upper-middle class bedroom community. Because it is located in a growth area, expansion and change typify the district's pattern as it continues to open new schools and expand educational programming. A special feature of the district's schools is the number of highly developed resource centers which serve as the central element in school program planning. To meet the needs of students transferring into and out of the schools as a result of parental job changes, the district conducted a reading program evaluation in order to adopt a more comprehensive reading program with a broader appeal.

District 2

District 2 is a large "urban" suburban school district directly bordering on the city of Chicago. Its school population offers the greatest socio-economic diversity of the four districts, with an integrated student population which includes numerous ethnic subgroups. The district attempts

to meet the needs of this diverse group of students through a program emphasizing human dignity and cultural pluralism. However, this special program is being threatened by the problems of declining enrollment and decreasing revenues. Before the program evaluation was implemented, the district was using a wide variety of reading series. But a need existed for a unified reading program which would include series for marginal readers, remedial readers and average readers. After adopting two separate series for the groups of special learners, the district formed an evaluation committee to select a series for the average readers that would complement the two previously adopted series. Thus, this district adopted not one but three complementary reading programs.

District 3

District 3, a relatively small school district in an older, established south suburban community, serves an integrated population and offers a large number of federally funded Title I programs. The district attempts to meet student needs by using a wide variety of special teachers and resource specialists. Because of district-wide declining standardized test scores, there was a need to conduct a comprehensive reading program evaluation. The evaluation committee, created to function as a study committee, was given the charge to prepare a district reading philosophy and build a unified district reading program, which would meet the needs of the district's wide variety of learners by emphasizing skill development.

District 4

District 4, located in a small, established western suburb, services a middle and upper-middle class school population. Administrators and teachers

report that a special strength of the school district is widespread community involvement in and support of school programming. An added feature in the district's operation is its close affiliation with a small private liberal arts college located in the community. The decision to undertake a program evaluation in reading was part of a larger district plan to reorganize the two key areas of math and language arts. Because reading programs varied from school to school within the district, the administration chose as the first step of their reorganization plan the adoption of a new basal reading series.

Analysis of Questionnaire Results

Introduction

The project director and members of the evaluation committee in each district completed the research instrument, the evaluation questionnaire. In answering "yes" or "no" to the series of forty task questions drawn from the CIPP framework, they provided information to answer two kinds of evaluation questions: (1) What evaluation tasks did the committee complete in the process of conducting their program evaluation? and, (2) To what extent did the committee complete each of the four evaluation types (CIPP)? Quantitative data provided by the questionnaire are analyzed and reported within each district and across all four districts.

Comparisons Within Each District - Null Hypothesis I

The questionnaire provides information to assess the practice of context, input, process, and product evaluation in each district. The following null hypothesis is a means for analyzing the frequency of tasks completed for each evaluation type:

There is no significant difference in the frequency of the tasks reported for each evaluation type (CIPP) within each district.

Statistical Analysis 1. Initially, the researcher prepared a table of means to show the percentage of tasks which were completed for each evaluation type in all districts to serve as a basis for comparing the means and as a source of information for further data analysis.

Results of 1:

Table 1
Means and Grand Means for CIPP in all Districts

| Evaluation Type | District | | | | Row \bar{x} |
|------------------|-----------|-----------|-----------|-----------|---------------|
| | 1 N=10 | 2 N=12 | 3 N=16 | 4 N=11 | |
| Context | 89.0% | 93.0% | 76.9% | 78.8% | 84.6% |
| Input | 87.9% | 80.0% | 77.6% | 81.4% | 81.75% |
| Process | 50.0% | 53.3% | 44.3% | 67.0% | 53.65% |
| Column \bar{x} | 77.0% | 78.0% | 62.7% | 73.3% | |

Discussion of 1. District 1 reported completion of 77% of all evaluation tasks and showed the greatest completion tasks reported for context evaluation ($\bar{x} = 89\%$) and the fewest completed tasks for product evaluation ($\bar{x} = 50\%$). District 2 reported completion of 77.8% of all evaluation tasks, showing the greatest completion of tasks reported for context evaluation ($\bar{x} = 93\%$) and the fewest completed of tasks for product evaluation ($\bar{x} = 53.3\%$). In District 3, the committee reported the completion of 62.7%

of all evaluation tasks, with input evaluation showing the greatest completion of tasks reported, ($\bar{x} = 77.6\%$) and product again showing the fewest completed tasks ($\bar{x} = 44.3\%$). Finally, District 4 completed 73.3% of all evaluation tasks with input evaluation ($\bar{x} = 81.4\%$) having the most completed tasks and process evaluation the fewest completed tasks ($\bar{x} = 66.1\%$).

Comparing the means of all four districts reveals similar frequency patterns from district to district. First, the most frequently reported evaluation types in all four districts are context and input evaluation. Secondly, the least frequently reported evaluation type in three districts is product evaluation. In addition to the agreement the districts show in reporting individual evaluation types, they display similar results in their column means for all tasks reported in their districts ranging from 62.7% to 78.0%.

Statistical Analysis 2. The ranges among the means of completed evaluation types within each district are broad: District 1 39.0% (89.0% - 50.0%); District 2, 39.7% (93.0% - 53.3%); District 3, 33.3% (77.6% - 44.3%); and District 4, showing the smallest range 15.3% (81.4% - 66.1%). The question is whether these differences are significant within each district. Therefore, a second statistical procedure is necessary to analyze these differences.

A one way analysis of variance using ANOVA for unequal N's is an appropriate way to analyze the within district means in Table 1. This statistic was chosen to test whether a significant difference exists among the means of the four evaluation types within each district (Ferguson, 1977, pp.215-217).

Results of 2.

Table 2

One-Way Analysis of Variance for Questionnaire Data

| District | Source of Variation | Sum of Squares | Degrees of Freedom | Variance Estimate |
|---|---------------------|----------------|--------------------|-------------------|
| 1 | Between | 7985.39 | 3 | 2661.8 |
| | Within | 12214.61 | 36 | 339.29 |
| Total | | | | F = 7.85 |
| F .05 (3.36) = 2.88 | | | | |
| F .01 (3.36) = 4.41 | | | | |
| F _{obt} = 7.85 is significant at .05 and .01 levels. | | | | |
| 2 | Between | 7087 | 3 | 2362.3 |
| | Within | 2167.4 | 36 | 338 |
| Total | | | | F = 6.99 |
| F .05 (3.36) = 2.88 | | | | |
| F .01 (3.36) = 4.41 | | | | |
| F _{obt} = 6.99 is significant at .05 and .01 levels. | | | | |

Table 2 (Continued)
 One-Way Analysis of Variance for Questionnaire Data

| District | Source of Variation | Sum of Squares | Degrees of Freedom | Variance Estimate |
|----------|---------------------|----------------|--------------------|---|
| 3 | Between | 8144.6 | 3 | 2716.5 |
| | Within | 20299.5 | 36 | 563.9 |
| | Total | | | F = 4.817 |
| | | | | F .05 (3.36) = 2.88 |
| | | | | F .01 (3.36) = 4.41 |
| | | | | $F_{obt} = 4.817$ is significant at .05 and .01 levels |
| 4 | Between | 1880.5 | 3 | 626.8 |
| | Within | 10501.9 | 36 | 291.7 |
| | Total | | | F = 2.15 |
| | | | | F .05 (3.36) = 2.88 |
| | | | | F .01 (3.36) = 4.41 |
| | | | | $F_{obt} = 2.15$ is not significant at .05 or .01 levels. |

Discussion of 2. The use of a one way ANOVA to analyze the questionnaire data from each of the four districts yielded the following conclusions:

1. There is a significant difference among the context, input, process and product means in districts 1, 2 and 3.

2. There is no significant difference among the context, input, process, and product means in district 4.

Statistical Analysis 3. Because the ANOVA results (Statistical Analysis 2) indicate a significant difference among the evaluation type means within District 1, 2 and 3, the researcher employed the Scheffe Method of Multiple Comparisons to compare sets of means to identify where the significant differences occur. For a number of reasons, Scheffe's method is an appropriate statistic for analysis of this data. While it is a rigorous criterion, its simplicity and versatility are useful over a wide variety of situations. Further, it is applicable to this study because it is appropriate for analysis of data drawn from unequal sample sizes (Hayes, p. 606). Formula for Scheffe's method:

$$F = \frac{(\bar{x}_1 - \bar{x}_2)^2}{S_w^2/n_1 + S_w^2/n_2}$$

Results of 3. The use of Scheffe's method yielded significant results in the following table of means (Table 3):

Table 3
Comparisons of Means of Evaluation Types
Reported on Questionnaire

| District | Comparison |
|----------|--|
| 1 | Context \bar{x} = 89.0% vs. Product \bar{x} = 50.0% * ** Input \bar{x} = 87.9% vs. Product \bar{x} = 50.0% * ** Process \bar{x} = 81.8% vs. Product \bar{x} = 50.0% * ** |
| 2 | Context \bar{x} = 93.7% vs. Product \bar{x} = 53.3% * ** Input \bar{x} = 80.0% vs. Product \bar{x} = 53.3% ** Process \bar{x} = 84.9% vs. Product \bar{x} = 53.3% ** |
| 3 | Input \bar{x} = 77.6% vs. Product \bar{x} = 44.3% ** |
| 4 | None |

Note: * refers to level .01
** refers to level .05

Discussion of 3. In District 1, product evaluation is significantly lower than all other evaluation types.

Data gathered through the use of the questionnaire indicated that the evaluation committee practiced context (\bar{x} = 89%) input (\bar{x} = 87%) and process evaluation (\bar{x} = 81.8%) to a greater degree than product evaluation (\bar{x} = 53.3%).

An analysis of the data from District 2 yielded similar results. Again the evaluation committee gave greater attention to context (\bar{x} = 93.7%) input (\bar{x} = 80%) and process evaluation (\bar{x} = 84.9%) than to product evaluation (\bar{x} = 53.3%). The comparison of means in District 3 yielded one significant

difference - input evaluation ($\bar{x} = 77.6\%$) was completed to a greater extent than product evaluation ($\bar{x} = 44.3\%$). There was no significant differences in the means for District 4.

Summary Discussion of Null Hypothesis I. Based on the foregoing results, null hypothesis I is rejected. Results obtained from the use of two statistical analyses: the one way ANOVA, and Scheffe's method for comparison of means indicate that there is significant difference in the frequency of tasks reported for each evaluation type (CIPP) within each district. Both the descriptive means table and these two analytic statistics indicate that context, input and process evaluation were given greater importance than product evaluation in all districts.

Comparisons Within the Group of Districts - Null Hypothesis II

Statistical analyses have identified the significant differences which exist between evaluation type means within each district. Now, the question remains: Is there a significant difference among the evaluation types across districts? The second null hypothesis relative to the questionnaire addresses this issue:

There is no significant difference in the frequency of tasks reported for each evaluation type (CIPP) across all districts.

Statistical Analysis 1. Initially, the researcher prepared a tally to indicate the number of districts which reported completing each evaluation task on the questionnaire. The tasks were then given 1 point for each time they were reported by a district. These transformations are reported in Table 4:

Table 4
 Number of Evaluation Types Completed by all
 Districts as Reported on Questionnaire

| Evaluation Type | Number of Times Reported | Per cent |
|-----------------|-----------------------------|----------|
| Context | 32 | 80.0% |
| Input | 38 | 68.0% |
| Process | 26 | 72.0% |
| Product | 7 | 25.0% |

Discussion of 1. When the evaluation tasks reported in the questionnaire are tabulated in all districts, it becomes apparent that context, input and process tasks were reported frequently: 80% of the context tasks were reported, as being completed, 72% of the process tasks, and 68% of the input tasks. However, the four districts only reported completing 25% of the product tasks.

Statistical Analysis 2. Analysis 1 indicates that there is a difference in the degree to which each evaluation type was reported by all districts. A second statistical analysis was used to determine whether this difference is significant. The researcher used a two-way analysis of variance (using ANOVA for unequal N's) (Winer, 1962, p.242).

The two-way ANOVA makes possible examination of the individual effects of the separate variables, evaluation type and district, as well as the interaction effect (between evaluation type and district).

Three questions are of interest:

1. Are these systematic effects due to the district itself?

(District 1,2,3,4) (Called variable A in Table 5)

2. Are these systematic effects due to the evaluation types (CIPP)?

(Called variable B in Table 5)

3. Are these systematic effects due neither to evaluation type alone, nor to district alone, but attributable only to the combination of a particular evaluation type with a particular district? (Called variable AB in Table 5).

Results of 2:

Table 5

Two-Way Analysis of Variance for Questionnaire

Data from all Four Districts

| Source of Variation | Sum of Squares | Degrees of Freedom | Variance Estimate |
|---------------------|----------------|--------------------|-------------------|
| A (District) | 5586.3 | 3 | 1862.1 |
| B (Evaluation Type) | 2420.5 | 3 | 7473.1 |
| AB | 625.8 | 9 | 69.5 |
| Within Cell | 50183.4 | 144 | 348.5 |

Obtained Values of F: $F_a = 5.34$, $F_b = 21.44$, $F_{ab} = .20$

Critical Values of F: F for A is F at (3.144) 2.60 3.78
 F for B is F at (3.144) 2.60 3.78
 F for AB is F at (9.144) 1.88 2.41

F_{obt} for A is F = 5.34 - Significant at .05 and .01 levels

F_{obt} for B is F = 2.44 - Significant at .05 and .01 levels

F_{obt} for AB is F = 2.40 -Not Significant

Discussion of 2. The results of the two-way ANOVA provide answers to the questions concerning effects of the two variables, evaluation type and district:

1. Since variable A - District - is significant ($F_{obt} = 5.34$), there are systematic effects due to the district.

2. Since variable B - evaluation types - is significant ($F_{obt} = 21.44$) there are systematic effects due to the evaluation type.

3. Because the interaction effect ($F_{obt} = .20$) is not significant, the systematic effects are not due to the combination of a particular evaluation type with a particular district.

Statistical Analysis 3. The results of the two-way ANOVA indicate significant effects due to both evaluation type and district. It is necessary to determine which evaluation type or types and which district or districts account for these significant effects. Therefore the researcher used Scheffe's method to make comparisons of means of evaluation type and district variables.

Table 6

Means and Grand Means for CIPP in all Districts

| Evaluation Type | District | | | | Row \bar{x} (Evaluation Type) |
|---------------------------------|----------|-------|-------|-------|------------------------------------|
| | 1 | 2 | 3 | 4 | |
| Context | 89.0% | 93.7% | 76.9% | 78.8% | 84.6% |
| Input | 87.9% | 80.1% | 77.6% | 81.4% | 81.0% |
| Process | 81.0% | 84.9% | 52.0% | 66.1% | 71.0% |
| Product | 50.0% | 53.3% | 44.3% | 67.0% | 53.0% |
| Column \bar{x} (Districts) | 77.0% | 78.0% | 62.7% | 73.3% | |

Results of 3. Result of comparison of means for evaluation types: Product evaluation (53.7%) is significantly lower than context evaluation (84.6%) in all four districts at $\alpha = .05$.

Result of comparison of means for districts: District 3 (62.7%) is significantly lower than District 2 (78%) and District 1 (77%) across all evaluation types (CIPP) at $\alpha = .05$ and $\alpha = .01$.

Discussion of 3. Scheffe's method has pinpointed the specific means of both evaluation types and districts which account for the overall systematic effects. In the evaluation type category, context evaluation has the greatest influence on the district results and product evaluation has the least influence. In the district category, both District 1 and District 2 reported completion of a significantly larger percentage of CIPP evaluation tasks than has District 3.

Summary Discussion of Null Hypothesis II. The findings from the previous statistical procedures result in the rejection of null hypothesis number 2. Both descriptive and analytic means indicate that there is significant difference in the frequency of tasks reported for each evaluation type (CIPP) in the group of districts. The descriptive frequency table indicates that context, input, and process evaluation were practiced to a greater extent than was product evaluation. While the two-way ANOVA showed that this difference was indeed significant, Scheffe's method pinpointed where the significance existed. Comparison of evaluation type means showed context evaluation significantly ahead of product evaluation. Comparison of district results showed that District 1 and District 2 completed a greater

percent of the evaluation types than District 3. Therefore both evaluation type and district had a systematic effect on the practice of program evaluation.

Ranking the Evaluation Types - Null Hypothesis III

A difference in the frequency of tasks and evaluation types reported within and across districts has been demonstrated. Now, it is instructive to use this information to rank the evaluation types within each district and then compare the rankings in the group of districts. The third null hypothesis relative to the questionnaire follows:

There is no significant difference in the ranking of the evaluation types in the group of districts.

Statistical Analysis. The researcher ranked the evaluation types for the CIPP means within district, and then used Kendall's Coefficient of Concordance to statistically compare the rankings. Kendall's statistic determines the extent to which members of a set of m distinct rank orderings of N things tend to be similar. The coefficient W shows how much rank orders tend to agree or show concordance (Guilford, 1964). Formula of Kendall's Coefficient of Concordance:

$$W = \frac{d^2}{x_t^2}$$

Results:

Table 7
 Ranking of Evaluation Types for Each District

| District 1 | District 2 | District 3 | District 4 |
|------------|------------|------------|------------|
| Context | Context | Input | Input |
| Input | Process | Context | Context |
| Process | Input | Process | Product |
| Product | Product | Product | Process |

Coefficient of concordance $W = .675$

Discussion. Evaluation types are ranked according to their type means; therefore, the ranking displays the relative importance each district placed on the four evaluation types. A visual analysis of each ranking reveals a unique rank ordering for each district. Several trends emerge. Context evaluation occupies either the first or second position in three of the four districts. Product evaluation was least used in three of the districts.

Summary Discussion of Null Hypothesis III. The coefficient of concordance adds a quantitative component to the descriptive analysis of the rankings. When perfect agreement exists between the rankings, $W = 1$, and when maximum disagreement exists, $W = 0$ (Ferguson, p. 312). The coefficient of concordance ($W = .675$) indicates that there is a 67.5% agreement in the ranking of evaluation types in the group of districts. This result indicates again, moderate, agreement in the practice of the four types of program evaluation in the four districts. They tend to view the relative merit of CIPP somewhat similarly.

Summary of Questionnaire Results

This statistical analysis of the data from the questionnaire presents a picture of each district's report and a composite picture of all four districts' reports on the practice of evaluation. In two of the four individual districts, product evaluation tasks were completed less often than any of the other evaluation types. In one of the four districts, product evaluation tasks were significantly behind only input evaluation tasks.

In the ANOVA of the group of districts, "evaluation type" and "district" were both significant variables in determining the reporting of evaluation practice. When "evaluation type" was analyzed as an individual effect, product evaluation again was reported to a lesser degree than context evaluation. Further, when "district" was analyzed as an individual effect, District 3 completed a significantly smaller percentage of the evaluation types than did either District 1 or District 2. Further, the fact that the districts did vary in the degree to which they completed the evaluation tasks and types is reinforced by a coefficient of concordance (.675) which indicates only moderate agreement in the four districts on the relative merit of CIPP.

Analysis and Summary of Interview Results

The researcher conducted a group interview with the project director and several members of the evaluation committee using the questionnaire as an interview format. She posed each task question to the group and when the answer was affirmative, she asked the committee members to name the methods they used in completing each task. The interview results, therefore,

provide quantitative results verifying task completion and qualitative results relative to the methods used for task completion. Taken together, these two indexes provide a second descriptive method for assessing CIPP's utilization in the four districts. Two major "research questions" were developed to summarize the interview data in order to coordinate that data with the questionnaire results. The first research question deals with quantitative information gathered in the interview.

Research Question 1. What evaluation tasks are most frequently reported by all four districts?

Table 8
Number of Evaluation Types Completed by all
Districts as Reported in the Interview

| Evaluation Type | Number of Times Reported | Percent |
|-----------------|--------------------------|---------|
| Context | 37 | 93% |
| Input | 45 | 80% |
| Process | 27 | 75% |
| Product | 22 | 79% |

Discussion

The evaluation types show similar percentages of completed tasks.

This quantitative information on the number of tasks completed for each evaluation type is complemented and amplified by the addition of the committee's descriptive response to the interview questions.

Research Question 2. What methods did each district use in completing the individual evaluation tasks? The following chart is a report of the district interviews listing the interview questions with a summary answer for each, including the pertinent information from all the districts.

CHART I

Summary of Interview Results

| <u>Context Evaluation</u> | <u>Question</u> | <u>Response</u> |
|---------------------------|--|---|
| | 1. How did you identify the learning outcomes which existed as a result of your current program? | 1. All four districts made primary use of standardized tests and supplemented these with teacher-written evaluations of programs already in use in the district. |
| | 2. How did you identify the learning outcomes you hope to achieve with your new program? | 2. Committee members and reading specialists established objectives, basing them on local curriculum guides or input from evaluation specialists and classroom teachers. |
| | 3. How did you identify the needs that were not being served by your current program? | 3. The committee considered standardized test results, teacher-made tests results, and records of requests for supplementary materials to determine the effectiveness of the current program. |
| | 4. How did you identify the potential human resources, such as faculty, staff, and volunteers? | 4. The committees used resource guidelines established by book company consultants to informally assess available human resources. |
| | 5. How did you identify the potential material resources such as classroom space, audio-visual materials, supplementary materials or library services? | 5. Committee members decided to buy a total comprehensive program, which made use of available classroom space, existing materials in district resource centers, and district library services. |
| | 6. How did you gather information from sources outside your district, such as research findings or outside consultants? | 6. The committees used three information-gathering methods: book company consultants gave presentations on the series under consideration; committees conducted district visitation programs to observe series in use; district reading coordinators shared relevant research findings with the committees. |
| | 7. How did you explore other available programs in terms of the impact of change on students, faculty, parents and community? | 7. After hearing presentations from reading coordinators and book company consultants, the committees analyzed the various series in terms of the committee's objectives and reported their conclusions through the use of questionnaires and voting sheets. |
| | 8. How did you assess the community values, attitudes, and priorities concerning a new program? | 8. Community representatives gave presentations at board meetings, served on PTA curriculum committees, and functioned as members of the evaluation committees. The community as a whole had access to the display copies of the series, and were invited to complete evaluation forms on these. |

CHART I (Continued)

| <u>Context Evaluation (Continued)</u> | <u>Question</u> | <u>Response</u> |
|---------------------------------------|---|---|
| 9. | How did you consult various data bases such as standardized test results, or parent surveys? | 9. The committees used data sources, which were external and internal to the district: statistics from the publishers, district standardized test scores, and the results from community needs assessments. |
| 10. | How did you identify the discrepancies that exist between your present program and the objectives you outlined for any new program? | 10. The committees considered existing data: standardized test results, and use-records for supplementary materials. In addition, they solicited new information from teachers through questionnaires to evaluate existing series and pilot series in the district. |
| <u>Input Evaluation</u> | | |
| 1. | How did you determine what you were already doing to meet your new set of objectives? | 1. The committees solicited teacher verbal input on a formal basis during building meetings and district institute days, and on an informal basis through personal conversations. |
| 2. | How did you determine what new strategies you could employ to meet the objectives? | 2. The committees chose to adopt a consistent, all-inclusive program. Book company consultants suggested appropriate learning strategies and the reading coordinator and committee subsequently adapted them to meet the district's needs. |
| 3. | How did you identify the costs of these new strategies? | 3. Financial cost was not a factor, but districts did identify a series of related costs, considering such factors as time requirements and space requirements. |
| 4. | How did you identify the benefits of these new strategies? | 4. Committee members and teachers completed evaluation forms on the series already in use in the districts, and the series under consideration (including pilot series and series observed during district visitation). |
| 5. | How did you identify staff training requirements to implement your new plans? | 5. District inservice plans to train teachers and administrators were designed through the cooperative efforts of the following groups of professionals: book company consultants, central office administrators, principals, committee members, and teachers. |
| 6. | How did you design procedures to implement a given program? | 6. The committees and the superintendents, with input from book company consultants and reading coordinators, chose blanket adoption as the procedure for implementing the new series. |
| 7. | How did you determine how existing staff could be used to implement the new program? | 7. The committee, principals, and teachers designed training programs to strengthen and support existing staff during implementation. |

CHART I (Continued)

| Input Evaluation (Continued) | <u>Question</u> | <u>Response</u> |
|---|-----------------|--|
| 8. How did you ascertain how present facilities and resources could be used to implement the new program? | | 8. Committee members and resource specialists assessed the districts' existing resources in terms of the resource guidelines established by the book company consultants. |
| 9. How did you set a schedule of events and activities to guide the implementation of the new program? | | 9. The school board or the superintendent set the date of adoption, and the committee established the schedule of events on the advice of book company consultants. |
| 10. How did you identify possible side effects which might result from implementing the new program? | | 10. While most districts did not identify side effects, one committee piloted a series in their own classrooms and warned of the possible need for study time and preparation time for teachers. |
| 11. How did you assess the attitudes of students, parents, teachers, etc. toward the new program? | | 11. While teachers' completed questionnaires on district visitations and pilot series, parents gave their oral and written input through parent-teacher conferences, PTA meetings and reactions to the display series. |
| 12. How did you assess students', parents', and teachers' knowledge of the new program? | | 12. The committees maintained communication with teachers through building meetings and inservice institutes, and with parents through PTA meetings, press releases, and building newsletters. |
| 13. How did your design involve evaluation during the implementation phase? | | 13. The committees assumed several active evaluative roles during the implementation year. They chaired building meetings to assess the program; they served as trouble-shooters in one or more buildings; they served as facilitators in teaching mock-lessons and aiding teachers in implementing the new program. |
| 14. How did you design involve evaluation of the outcomes of the program? | | 14. Program outcomes were assessed formally through the use of standardized tests and the evaluation component of the new program, and informally through verbal feedback from building reading specialists, teachers, and parents. |
| <u>Process Evaluation</u> | | |
| 1. How did you develop an implementation plan for your new program? | | 1. The committees designed the implementation plan with input from the book company consultants, superintendents, and reading coordinators. |
| 2. How did you determine the adequacy of the resources? | | 2. The committees received verbal feedback from resource specialists and teachers. |

CHART I (Continued)

| Process Evaluation (Continued) | <u>Question</u> | | <u>Response</u> |
|--------------------------------|--|--|---|
| 3. | How did you determine what kind of feedback is needed during piloting and implementation? | | 3. Written and verbal feedback within the district was addressed to committee members, building principals, reading coordinators, and the assistant superintendents for curriculum. In addition, book company consultants met regularly with teacher groups during the implementation year. |
| 4. | How did your feedback include monitoring the students', faculties', and parents' understanding of and agreement with your program? | | 4. One committee distributed questionnaires through the PTA. |
| 5. | How did you assess the physical facilities employed in program implementation? | | 5. The committees polled a number of groups, including reading specialists, principals, and teachers. |
| 6. | How did you account for staff effectiveness in making the program operational? | | 6. A variety of professionals became involved in the assessment of staff effectiveness. One committee sought input from principals, reading specialists, and teachers through the use of questionnaires and mid-year instruments. District administrative staff members assumed a variety of active roles in this regard. The assistant superintendent for curriculum in one district hosted a brunch at each building to get input from teachers and principals. Other members of the central office consultant staff conducted theme inservice meeting for teachers, and the reading coordinators often served as trouble-shooters. |
| 7. | How did you determine the adequacy of time schedules in facilitating program operation? | | 7. The reading coordinators set reading schedules for each building with the building principals and teachers. |
| 8. | How did you monitor informal interpersonal relationships among staff and students? | | 8. Informal feedback from teachers went to committee members, building principals, reading specialists and the reading coordinators. In addition, reading coordinators conducted regular visits to buildings and classes to get formal feedback from staff members. |
| 9. | How did you design formal lines of communication? | | 9. Formal lines of communication varied from district to district with teachers reporting to committee members or reading specialists in some districts, principals or coordinators in other districts. However, all districts used written questionnaires and other assessment forms to evaluate the series. |

CHART I (Continued)

Question

Response

Product Evaluation

- | | |
|---|--|
| <p>1. How did you determine whether the program is meeting its objectives?</p> <p>2. How did you assess the gain or loss in pupil achievement?</p> <p>3. How did you identify the unanticipated outcomes and their effects on the students?</p> <p>4. How did you determine the effect of the project on staff?</p> <p>5. How did you assess the attitudes of students, staff, parents, and community regarding the outcomes of the program?</p> <p>6. How did you determine whether or not the program's results justify the finances and efforts needed to maintain it?</p> | <p>1. The committee combined the use of standardized tests, and assessment components of the new series with verbal and written feedback from teachers, parents and reading specialists.</p> <p>2. Again, the assessment component of the series and standardized tests were the tools used to assess pupil achievement.</p> <p>3. Committee members handled complaints that came directly to them or the principals.</p> <p>4. The committee solicited informal verbal feedback from all teachers and administrators, who were directly involved with implementing the new program.</p> <p>5. The committee used staff questionnaires on institute days and elicited verbal feedback from parents and teachers throughout the implementation period.</p> <p>6. Cost effectiveness was not a factor of importance in the evaluation studies.</p> |
|---|--|

Analysis of Report Results

Each district evaluation committee presented a final evaluation report to their board of education to summarize their findings and recommend a basal series for adoption. Like the interview transcript, this written report is quantitative in that it delineates the tasks which were completed and qualitative in that it describes the methods each district used to complete the evaluation tasks.

The researcher, using the research instrument as a check-list in analyzing the final evaluation reports, prepared a tally indicating which evaluation tasks were reported. In addition, the researcher studied the various elements of the report, the chronological account of steps taken, the research instruments and reports of results, and written communications with various publics. The researcher used this information and the related information gathered from the interview to devise a list of the themes (see following section) which emerged in the completion of the program evaluation in each district.

Statistical Analysis and Results

The report yielded data on evaluation tasks and types completed in all districts. Again "research questions" are the basis for compiling the district responses.

Research Question 3. What tasks are most frequently reported in the written report across districts?

Initially the researcher prepared a tally to indicate the number of districts which reported completing each evaluation task in the written report. The tasks were given one point for each time they were reported by

a district. The transformations are reported in Table 9.

Table 9
Number of Evaluation Tasks Completed by all Districts as
Reported in the Written Evaluation Report

| Evaluation Type | Number of Times Reported | Percent |
|-----------------|--------------------------|---------|
| Context | 33 | 83.0% |
| Input | 44 | 79.0% |
| Process | 13 | 36.0% |
| Product | 4 | 14.0% |

Discussion and Summary of Report Results

A visual inspection of the percentage of evaluation types reported by all districts in their written evaluation report reveals several trends. The reports indicated that while similar attention was given to context (83%) and input (79%), the committee completed a smaller percentage of process (36%) and product (14%) evaluation tasks.

Summary Analysis of Combined Questionnaire, Interview, and Report Results (QIR)

The researcher conducted a quantitative and qualitative summary analysis of the results obtained for the completion of each evaluation type in the questionnaire, the interview, and the report combined in all four districts. For the quantitative analysis she calculated a coefficient of completion for each task and evaluation type to indicate to what extent all four districts reported completing each evaluation type in the questionnaire,

interview and report. Tables 10-12 below present the quantitative results of the combined questionnaire, interview and written report. For the qualitative analysis, the researcher used data from the questionnaire, interview and report to compile a list of the themes, including methods and characteristics unique to each district's practice of program evaluation.

Quantitative Analysis of Questionnaire, Interview, and Report Results -

Null Hypothesis IV

Statistical Analysis 1. The data gathering process addresses Null Hypothesis IV:

There is no significant difference in the coefficient of completion for each evaluation type across districts as reported by the questionnaire, interview, and report.

Results of 1. The coefficient of completion shows the percentage of evaluation types completed by all districts and reported in the questionnaire, interview and written report.

Table 10

Coefficients of Completion for CIPP in all Districts

| Evaluation Type | Coefficient of Completion |
|-----------------|---------------------------|
| Content | .825 |
| Input | .773 |
| Process | .631 |
| Product | .369 |

Discussion of 1. Analysis of the coefficients of completion yields the following rank order: Context is first in degree of completion (.825), then input (.773), process (.631), and finally product (.369). The districts have given considerable importance to the first three of the evaluation types. There is a large discrepancy, however, between the coefficient of completion for context, input and process evaluation, and the coefficient of completion of product evaluation (.369). In the questionnaire, interview, and report, evaluation committees indicated that they completed two-thirds of the tasks for context, input and process evaluation. However, they completed just over one-third of the tasks included in the product evaluation.

Statistical Analysis 2. Because the coefficients of completion range from a high of .825 for context evaluation to a low of .369 for product evaluation, it is necessary to determine whether the difference between the coefficient of completion for all four evaluation types is significant. The researcher used a one-way analysis of variance to determine whether or not there were significantly different means (Ferguson, pp. 215-217).

Results of 2. The one-way ANOVA yielded the following results:

Table 11

One-Way Analysis of Variance for Coefficient of Completion as
Reported in Questionnaire Interview and Written Report

| Source of Variation | Sum of Squares | Degrees of Freedom | Variance Estimate |
|---------------------|----------------|--------------------|-------------------|
| Between | 10309.6 | 3 | 3436.5 |
| Within | 17577.2 | 36 | 488.3 |
| Total | | | F = 7.04 |

From the table = 13.36

F .05 (3.36) = 2.88

F .01 (3.36) = 4.41

$F_{obt} = 7.04$ is significant at .05 and .01 levels.

Discussion of 2. The use of a one-way ANOVA to analyze the coefficient of completion for each of the four evaluation types indicates a significant difference between context, input, and process and product.

Statistical Analysis 3. Because the ANOVA indicated significant difference among the means, the Scheffe Method of Multiple Comparisons was employed to compare sets of means to determine where significant differences occur.

Formula:

$$F = \frac{(\bar{x}_1 - \bar{x}_2)^2}{S_w^2/n_1 + S_w^2/n_2}$$

Results of 3: The use of Scheffe's method yielded significant results in the following comparisons of means:

Table 12

Comparisons of Means of Evaluation Types as Reported in Questionnaire,
Interview and Written Report

| Comparison | Significance |
|---|---------------------------|
| Content \bar{x} = .825 vs. Product \bar{x} = .369 | * ** |
| Input \bar{x} = .773 vs. Product \bar{x} = .369 | * ** |
| +++++ | |
| Content \bar{x} = .825 vs. Input \bar{x} = .773 | No Significant Difference |
| Content \bar{x} = .825 vs. Process \bar{x} = .631 | No Significant Difference |
| Input \bar{x} = .773 vs. Process \bar{x} = .631 | No Significant Difference |
| Process \bar{x} = .631 vs. Product \bar{x} = .369 | No Significant Difference |

Note: * refers to level .01
** refers to level .05

Discussion of 3. When the questionnaire, interview and report results were analyzed for all four districts, the use of Scheffe's method revealed that the coefficient of completion of product evaluation (.369) is significantly lower than the coefficient of completion of both context evaluation (.825) and input evaluation (.773). All four districts reported completing more tasks for both context and input evaluation than for product evaluation.

Summary Discussion of Null Hypothesis IV. The findings from the foregoing statistical procedures suggest rejection of null hypothesis number four. The use of a one-way ANOVA and Scheffe's method to analyze the coefficient of completion for each evaluation type as measured in all districts by the questionnaire, Interview, and Report indicate that product evaluation

(.369) is significantly lower than context evaluation (.825) and input evaluation (.773).

Qualitative Analysis of Questionnaire, Interview and Report Results

As in the analysis of interview results, the researcher developed a "research question" to summarize the questionnaire, interview and report data in order to coordinate the data with the coefficient of completion.

Research Question 4. What unique themes emerged in each district's completion of the program evaluation?

Although all districts studied engaged in a comprehensive process of program evaluation for the singular purpose of basal reading series adoption, each district committee used several methods to fashion a program evaluation to meet the peculiar needs of the district. The researcher used information gathered with the questionnaire, interview, and report to compile a list of unique evaluation themes which emerged in each district program evaluation.

District 1. The special nature and purpose of the District 1 program evaluation dictated the committee's choice of its specific evaluation methods. Because of the district's large size, and its commitment to adopt a new reading philosophy, District 1 used representatives from many diverse groups in an investigative process to reach its goal. The following themes characterize the program evaluation of District 1.

1. District 1 used three groups of resource people in the program evaluation: the language arts standing committee, an ad hoc study committee (the evaluation committee), and a cadre of teachers who were specially trained in the new program. This method was used in a large school district in an effort to insure sufficient representation by all groups of teachers.

2. In District 1, a unique supportive role was played by resource specialists and learning centers: a new resource program had been the focus of district efforts and subsequently became a major feature of the programming in individual schools. This source of assistance, including both facilities and personnel, was used in both planning and implementing the new basal reading program.

3. District 1 used controlled sample testing for longitudinal evaluation: Prior to the program evaluation, the district identified a single group of learners for longitudinal assessment. They tested a sample from this group each year to make possible the comparison of scores from old and new reading programs.

4. District 1 conducted a comprehensive visitation program: committee members identified this method as the major focus of their data-gathering process to precede textbook adoption. The committee made on-site visitations to districts using the numerous series under consideration, using both formal and informal assessment tools. This program provided a vehicle for professional evaluation of the series in question, because both committee members and teachers who were using the series, provided verbal and written evaluations.

District 2. The evaluation committee in District 2 fashioned its program evaluation to provide ethnically and socially relevant curricula and at the same time address problems of declining enrollment and decreasing revenues. The unique feature of their program evaluation was its attempts to assess the needs of the diverse groups the district services, and still use the abundant broad-based resources the community offers the school district.

1. District 2 based its program evaluation on the curriculum guide, The Teaching of Reading - Phase I. In 1977 the district prepared the guide for the teaching of reading - kindergarten to grade 6, including the district reading philosophy, goals, and scope and sequence of requisite reading skills. Designed to meet the specific needs of the schools' changing student population, this document served to guide the committee in preparing all components of their program evaluation.

2. District 2 piloted seven basal series in all district schools: all teaching personnel participated in the pilot program by using the series, or by observing a series in use and submitting a formal, written evaluation.

3. District 2 designed a sequential program of teacher involvement in evaluation: at the beginning of the evaluation, the committee prepared a time-line which specified that classroom teachers be involved at every stage of decision-making. The teacher involvement components included a needs assessment check-list, a reading inservice survey, and a checklist of necessary characteristics for a basal series. Finally, teacher vote determined the final selection of a basal series for the district.

District 3. The committee in District 3 faced the challenging problems of designing a program evaluation to address the needs of large groups of remedial learners and of adopting a reading program that would improve declining scores on standardized tests. For these reasons the work of the committee was analytic in purpose and educative in direction, turning the program evaluation into a learning experience for the committee and district teachers.

1. The District 3 committee was a study committee: the committee

spent an initial three months studying the area of reading, analyzing district needs, and defining the district philosophy and goals of reading. Teaching and learning continued to characterize the committee work as they performed the following functions during the implementation year: troubleshooting for the district, educating teaching personnel, serving as pilot teachers for the series, demonstrating mock lessons at district institutes.

2. The final report of District 3 is a study report: the report includes an in-depth analysis showing the relationship of the adopted series to district needs, goals, and objectives. Further, the committee provided the following tools to aid program implementation; an implementation plan, an in-service format, a district-wide reading schedule, and a set of specific guidelines for teaching the new series.

3. In District 3 the book company consultants provided the major force during the implementation phase: after the committee selected the new basal reading program they turned to the book company consultant for the implementation plan and the inservice program for general preparation and continued guidance in introducing the new series.

District 4. In District 4, the community had identified and expressed support for improved programming in all areas of the language arts. The district acted upon this mandate by initiating a five year program evaluation cycle to evaluate all reading and language programs. Both the community and district established as the purpose of this first evaluation the replacing of the district's many reading series with a single unified reading program. Because education is highly valued in this district of middle and upper-middle class professionals, the community gave the district support and

resources for conducting the evaluation.

1. District 4 conducted an evaluation course for the committee taught by a college professor: an evaluation specialist from a local university specifically designed an evaluation course to meet their needs. The committee used the course as a forum for establishing a philosophy of reading, reading goals, and criteria for evaluating a basal reading program. These three components formed the basis of their work.

2. District 4 used a systematic, cyclic plan for program evaluation based on a five year schedule:

1st year - Preparation - - In this phase, the committee, established by the administration, initiates the study by formulating the philosophy, goals, and objectives of the program evaluation.

2nd year - Study - - During the second year, the committee secures materials, evaluates them according to the predetermined criteria, adopts a new program, and initiates staff orientation.

3rd year - Implementation - - In addition to the major tasks of staff development, the committee conducts formative evaluation of goals, objectives, materials, instruction, and learning in order to make recommendations for program modification.

4th year - Second year of modification with improvements - - The whole program in general is evaluated and particular suggestions made for program improvement.

5th year - Program continuation - - No change is made during this final year of the program schedule. Then the cycle repeats.

3. In District 4, this first 5-year cycle of program evaluation served as an "agent of change": as a result of the program evaluation the district: hired a second district reading coordinator and reading aides for each school, brought reading specialists into the district for in-service training of teachers, set an agenda to identify reading needs in the next two years, and formulated a district-wide testing program to supplement the new reading program.

Because each district represented a unique socio-economic and educational setting, and because each district committee undertook its program evaluation in keeping with the district's particular philosophy and needs, four distinct program evaluations resulted from their efforts. The statistical evidence indicates that all four districts completed the recognized tasks under each evaluation type. However, they devised unique methods to fashion their program evaluations to accommodate the human and material resources of their settings.

Summary of Results

Several significant relationships emerged from the data analyses of the individual district's use of evaluation.

Districts 1 and 2 were noticeably similar: both completed more context, input and process tasks than product tasks. Both also completed significantly more of all evaluation tasks than did District 3. These results are significant in view of the large size of both Districts 1 and 2 and their large central office staffs. They had the personnel to complete more tasks and seemed to use that personnel effectively. District 3, on the other hand, reported significantly more input tasks than product tasks and completed less

of the evaluation tasks overall than Districts 1 and 2, but also had less staff than Districts 1 and 2. District 4 is noticeable because it did not stand out as distinct from the other three districts in its practice of the four evaluation types.

The analyses of the entire group of districts also produced some notable relationships. First, all districts completed significantly more context and input tasks than product tasks. Further, all districts reported similar patterns of practice in using the evaluation types. Each method of measuring - the questionnaire, interview and written report - yielded similar results. All districts reported completing more context tasks than any others, while their completion of product tasks was uniformly low.

Finally, although all districts completed similar tasks, the way they completed these tasks was often different from district to district.

CHAPTER V

CONCLUSIONS

This study began as an assessment of the state of the art in program evaluation. But it yielded much more than just a description of that art. The problem, as stated in Chapter I, was that a gap exists between theory and practice in actual program evaluation. This study showed that the gap is much narrower than we at first imagined.

We used a classic, theoretical model in our assessment and found, in fact, that school districts are doing the very evaluation tasks that the model proposes. A new gap emerged, however, because theorists have been saying that program evaluation is largely product-oriented. But such is not the case. Rather, in this study, context emerged as the primary focus and major task of program evaluation, with product tasks seeming least important because they were least completed. The theorists were not completely wrong in their assumptions, for we showed, as Novak (1977) asserted, that staff/teacher involvement in program evaluation is the most effective means of staff development. In District 3, for instance, the "study" orientation of the program evaluation involved both teachers and central staff and functioned as a kind of in-service training. In District 4, the process of program evaluation promoted significant changes in staff and thus promoted development within the district.

Another of our initial contentions was that to investigate formal program evaluation by analyzing real world evaluations is a meaningful way to assess the state of the art in program evaluation. In the study we focused

on a single comprehensive model, looking for quantitative data only. But, at the completion of the study we were able to identify multi-faceted models, adapted to meet particular district needs, and, in addition to significant quantitative data, enough qualitative data to add perspective and breadth to the findings.

A further contention to begin this study was that the analysis of systems of evaluation must be based on an accepted theory and model. The CIPP model provided an ideal framework for assessing individual districts, making cross-district comparisons, and constructing a total picture of program evaluation practice.

Not only did the study prove or disprove earlier theories and contentions, but it also fills the need for data to describe and delineate the actual practice of program evaluation, adding to the state of the art, not merely describing it.

The study also proved the usefulness of the CIPP model as an effective way to study the practice of program evaluation. It functioned as a true practitioners' model, for the staff of the four districts were able to report their practice of program evaluation in terms of the CIPP evaluation tasks and still feel free to report their own unique procedures.

Not only did the study challenge some of the theories about program evaluation, but it also challenged our own definition of comprehensive program evaluation. Our understanding of a comprehensive program evaluation consisted of four assumptions:

1. An evaluation committee should be formed.
2. The committee should establish objectives and select several basal

reading series for consideration according to the objectives.

3. The committee should recommend the series for adoption in a final written report.

4. The committee should be comprised predominantly of elementary classroom teachers.

The four districts' definition of comprehensive program evaluation was so much broader and more elaborate than the understanding of evaluation set forth in this study that the criteria could only be regarded as minimal, rather than as a standard by which to judge program evaluation practice.

The demographic characteristics of the four districts influenced the practice of program evaluation. Larger districts, (1 and 2) with extensive central office staffs, tended to conduct a more complete program evaluation, reporting more tasks completed than the smaller districts (3 and 4). Thus districts used staff, facilities and money wisely and effectively in program evaluation (as defined by formal reports and series adoption). District 3, on the lowest socio-economic level, used the program evaluation model as an opportunity for in-service training of teachers, thus also making efficient use of this evaluation opportunity. In the most homogeneous and stable district, District 4, the program evaluation functioned as an "agent-of-change" to foster staff development and provide an opportunity to examine and change long-established practices.

The two larger districts completed more context, input and process tasks than product tasks. Context and input tasks, focusing as they do on meeting needs, were ideally suited to the two larger districts which have diverse and mobile populations with multiple needs. Product tasks would be

less important in such a continuously changing school population. District 3, with its focus on program evaluation as "study" and "learning experience" could be expected to concentrate on input over product as it did, because input is essential to learning and absorption of new information. District 4 completed all evaluation tasks nearly equally, due to its stability and homogeneity, needing to address no particular special needs. However, its program evaluation process produced the most significant and far-reaching changes. The effectiveness of the program evaluation model and practice is unmistakable.

References

- Alkin, M.C. Evaluation development. Evaluation Comment. 1969, 2, 2-7.
- Barnette, J. The role of evaluation in organizational development.
New York: Paper presented at the Annual Meeting of the American Educational Research Association, 1977. (ERIC Document Reproduction Service No. ED 152824).
- Bleakley, L. Rhetorical stance and the teaching of literature. Purdue, Indiana: Purdue University, 1974. (ERIC Document Reproduction Service No. ED 101349).
- Bloom, B. Taxonomy of educational objectives. New York: David McKay, 1956.
- Bloom, B. Human characteristics and school learning. New York: McGraw Hill, 1978.
- Bloom B., Hastings, J., & Madus, G. Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.
- Borich, G. (Ed.). Evaluating educational programs and products. Englewood Cliffs, New Jersey: Educational Technology Publications, 1974.
- Browder, L. Developing an educationally accountable program. Berkeley, California: McCutchan Publishing, 1973.
- Carter, W. A taxonomy of evaluation models: use of evaluation models in program evaluation. Dallas, Texas: Dallas Independent School District, 1975. (ERIC Document Reproduction Service No. ED 109244).
- Cohen, E. Problems and prospects of teaming: research and development memorandum no. 143. Stanford, California: Stanford University, 1976. (ERIC Document Reproduction Service No. ED 126081).
- Cooley, W., Lohnes, P. Evaluation research in education. New York: Irvington Publishers, Halsted Press, 1976.
- Cronbach, L.J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.
- Cronbach, L.J. Evaluation in a context of accomodation. In E. Bossone (Ed.), The second annual conference on testing: major issues, proceedings. New York: 1978 (ERIC Document Reproduction Service No. ED 166245).
- Directory of Illinois Schools. Illinois: Office of the Superintendent of Public Instruction, 319, 1978-1979.

- DuBois, P. A history of psychological testing. Boston: Allyn & Bacon, 1970.
- Eash, M.J. Assessing curriculum materials: a preliminary instrument. EPIE Educational products report, 1969, 2, 18-24.
- Ebel, R. Measuring educational achievement. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.
- Eisner, E. Mind as cultural achievement. The John Dewey Lecture, the John Dewey Society for the Study of Education and Culture, Gainesville, Florida, 1979.
- Ewy, R. & Chase, C. Program and product evaluation: an information packet. Denver, Colorado: Colorado State Department of Education, 1977. (ERIC Document Reproduction Service No. ED 109822).
- Ferguson, G. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1971.
- Gess, D. The implementation of a district-wide evaluation model. Ft. Lauderdale, Florida: Doctoral Dissertation, Nova University, 1974. (Eric Document Reproduction Service No. ED 109822).
- Glass, G. The growth of evaluation methodology. Research paper no. 27. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, 1969, (mimeo).
- Good, T., Biddle, B., & Brophy, J. Teachers make a difference. New York: Holt, Rinehart and Winston, 1975.
- Greenbaum, W. Measuring educational progress. New York: McGraw-Hill, 1977.
- Guba, E. Evaluation and the process of change. In Notes and working papers concerning the administration of programs. Title III ESEA. Washington, D.C.: U.S. Senate, Committee on Labor and Public Welfare, Subcommittee on Education, 1967.
- Guba, E., & Stufflebeam, D. Evaluation: the process of stimulating aiding and abetting insightful action. Bloomington, Indiana: Indiana University, 1970. (ERIC Document Reproduction Service No. ED 055733).
- Hammond, R. Evaluation at the local level. In B. Worthen & J. Sanders (Eds.), Educational evaluation: theory and practice. Worthington, Ohio: Charles A. Jones, 1973.
- Hayes, W. Statistics for the social sciences. New York: Holt, Rinehart & Winston, 1973.

- Hecht, A. Utility of the CIPP Model for evaluating an established career program in a community college. San Francisco, California: Paper presented at the annual meeting of the American Educational Research Association, 1976. (ERIC Document Reproduction Service No. ED 120203).
- Hinkle, D. The conceptualization of the Stufflebeam CIPP Evaluation Model in a multivariate context (Doctoral dissertation, University of Toledo, 1971). Dissertation Abstracts International; 1971, 32, 3682A. (University Microfilms No. 72-3268).
- Kunder, L. Procedures for textbook and instructional materials selection: ERS report. Arlington Virginia: Educational Research Service, 1976. (ERIC Document Reproduction Service No. ED 132677).
- Krathwohl, D. Taxonomy of educational objectives handbook II: affective domain. New York: McKay, 1964.
- McLure, G. Sex role stereotyping and evaluation: a systems approach. Chicago, Illinois: Paper presented at the 78th Annual Meeting of the North Central Association, 1973. (ERIC Document Reproduction Service No. ED 078326).
- Metfessel, N., & Michael, W. A paradigm involving multiple criterion measures for the education of the effectiveness of school programs. Educational and Psychological Measurement, 1967, 27, 931-943.
- Michael, M. Planning and evaluating library system services in Illinois using the CIPP Evaluation Model. Urbana, Illinois: University of Illinois, 1976. (ERIC Document Reproduction Service No. ED 118158).
- Nevo, D. & Stufflebeam, D. Evaluation priorities of students, teachers and principals. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., 1975. (ERIC Document Reproduction Service No. ED 104935).
- Nolin, S. School staff and community involvement in the planning and evaluation of educational programs through the use of the CIPP Evaluation Model in selected school districts in Los Angeles County (Doctoral dissertation, University of Southern California, 1976). Dissertation Abstracts International, 1976, 37, 6856A-6857A.
- Novak, C. An involvement approach to the evaluation of local district programs. New York: Paper presented at the Annual Meeting of the American Educational Research Association, 1977. (ERIC Document Reproduction Service No. ED 152819).
- Ornstein, A. Teaching in a new era. Champaign, Illinois: Steps Publishing, 1976.

- Ornstein, A. An introduction to the foundations of education. Chicago, Illinois: Rand McNally College Publishing, 1977.
- Pasch, M. Examining the use of paraprofessionals at Parma Schaaf Jr. High: an evaluation study. Parma, Ohio: Parma Schaaf Jr. High School, 1976. (ERIC Document Reproduction Service No. ED 163643).
- Pipho, C. (Ed.). Minimum competency testing. Phi Delta Kappan, 1978, 59, 585-625.
- Popham, W.J. Objectives '72. Phi Delta Kappan, 1972, 53, 435.
- Popham, W.J. (Ed.) Evaluation in education: current applications. Berkeley, California: McCutchan Publishing, 1974.
- Popham, W.J. Educational evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.
- Provus, M. Discrepancy evaluation. Berkeley, California: McCutchan, 1971.
- Reeves, J., & Michael, W. The application of the Stufflebeam educational decision-making model to the evaluation of a dental team training program. Ellenville, New York: Paper presented at the Annual Meeting of the Northeastern Educational Research Association, 1973. (ERIC Document Reproduction Service No. ED 094004).
- Reinhard, D. Methodology development for advocate team use for input evaluation. New Orleans, Louisiana; Paper presented at the Annual Meeting of the American Educational Research Association, 1973. (ERIC Document Reproduction Service No. ED 074137).
- Rice, E. An exploratory study to investigate the aspects of the construction and use of instruments encompassing the evaluative strategies of context, input, process and product of inservice programs (Doctoral dissertation, University of Iowa, 1968). Dissertation Abstracts International, 1968, 29, 3922A-3923A. (Univeristy Microfilms No. 69-8796).
- Rose, C., & Nyre, G. The practice of evaluation, ERIC/TM Report 65. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1977. (ERIC Document Reproduction Service No. ED 152859).
- Scriven, M. Evaluation perspectives and procedures. In W.J. Popham (Ed.), Evaluation in education: current applications. Berkeley, California: McCutchan, 1974.
- Scriven, M. The methodology of evaluation. In A. Bellack & H. Kliebard (Eds.), Curriculum and evaluation. Berkley, California: McCutchan, 1977.

- Shiplett, G. An adaptation of the CIPP Model of evaluation for use in adult education (Doctoral dissertation, Indiana University, 1974). Dissertation Abstracts International, 1974, 35, 5763A. (University Microfilms No. 75-5573).
- Stake, R. School accountability laws. The Journal of Educational Evaluation: Evaluation Comment, 1973, 4.
- Stake, R. The validity of district-wide test means. In E. Bossone (Ed.), The second annual conference on testing: major issues, proceedings. New York, 1978. (ERIC Document Reproduction Service No. ED 166245).
- Stead, F. A third party evaluation of the Appalachian Maryland experience-based career education project: 1976-1977. Morgantown, West Virginia: West Virginia University, 1977.
- Stufflebeam, D. The relevance of the CIPP Evaluation Model for educational accountability. Journal of Research and Development in Education, 1971, 5, 19-25.
- Stufflebeam, D., Foley, W., Gephart, W., Guba, E., Hammond, R., Merriman, H., & Provus, M. Educational evaluation and decision-making in education. Itasca, Illinois: Peacock, 1971.
- Summary of the educational specifications for a comprehensive elementary teacher education program, summary of the final report. Toledo, Ohio: Toledo University, 1968. (ERIC Document Reproduction Service No. ED 035595).
- Taba, H. Curriculum development: theory and practice. New York: Harcourt, Brace & Javanovich, 1962.
- Talmage, H. The textbook as the arbiter of curriculum and instruction. The Elementary School Journal, 1971, 10, 20-25.
- Talmage, H., Walberg, H., & Nicholas, S. An evaluation research model for system-wide textbook selection. Paper presented at the American Educational Research Association Conference, Division B, New York, 1977.
- Taylor, J. An administrator's perspective of evaluation. Paper No. 2. Kalamazoo, Michigan: Western Michigan University, 1974. (ERIC Document Reproduction No. ED 163059).
- Thorndike, R., & Hagen, E. Measurement and evaluation in psychology. New York: John Wiley and Sons, 1969 & 1977.
- Tyler, R. General statement on evaluation. Journal of Educational Research, 1942, 35, 492-501.

- Tyler, R. Basic principles of curriculum and instruction. Chicago: University of Chicago Press, 1950.
- Tyler, R. The evaluation of teaching. In R. Cooper (Ed.), The two ends of the log. Minneapolis, Minnesota: University of Minnesota Press, 1958.
- Tyler, R. Some persistent questions on the defining of objectives. In C. Lindvall (Ed.), Defining educational objectives. Pittsburgh: University of Pittsburgh Press, 1964.
- Wallace, R., & Shavelson, R. A systems analytic approach to evaluation: a heuristic model and its application. Syracuse, New York: Eastern Regional Institute for Education, 1970. (ERIC Document Reproduction Service No. ED 058292).
- Winer, B. Statistical principles in experimental design. New York: McGraw-Hill, 1962.
- Womer, F. What is national assessment? Ann Arbor, Michigan: National Assessment of Educational Progress, 1970.
- Woodward, J., & Yaeger, J. Evaluation of programs to train educational R & D personnel. Pittsburgh, Pennsylvania: Pittsburgh University, 1972. (ERIC Document Reproduction Service No. ED 064376).
- Worthen, B., & Sanders, J. Educational evaluation: theory and practice. Worthington, Ohio: Charles A. Jones, 1973.

APPENDIX A

- FIRST DRAFT -

DRAFT OF RESEARCH INSTRUMENT: STUDY II

The task of educational program evaluation has been conceptualized into a model by the Phi Delta Kappa National Study Committee on Evaluation. This model, known as the CIPP evaluation model, divides the process of educational program evaluation into four stages: context, input; process; product.

The following is a list of evaluation tasks that can be done in the process of doing a program evaluation. They are divided into four phases or aspects of evaluation. Check the tasks which your evaluation committee completed in doing your program evaluation.

Formulating Objectives

- | | | |
|-----|----|--|
| YES | NO | |
|-----|----|--|
1. Did you identify the learning outcomes which existed as a result of your current programs?
 2. Did you identify the learning outcomes you hope to achieve with your new program?
 3. Did you identify the needs that were not being served by your current program?
 4. Did you identify potential human and material resources?
 5. Did you gather information from sources outside your district such as research findings or outside consultants?
 6. Did you explore alternative programs in terms of the impact of change on concerned publics (ie., students, faculty, parents, etc.)?
 7. Did you assess community values, attitudes, and priorities concerning a new program?
 8. Did you consult various data bases (ie standardized test results, parent surveys, etc.)?
 9. Did you identify the discrepancies that exist between your existing program and the objectives you outlined for a new program?

*This is the Cipp Evaluation Framework used for summative evaluation of the project.

FIRST DRAFT

Examining Present Resources and Developing an Implementation Plan

- | | | |
|-----|----|--|
| YES | NO | <ol style="list-style-type: none"> 1. Did you determine what you were already doing to meet your new set of objectives? 2. Did you determine what new strategies you could employ to meet the objectives? 3. Did you identify both the costs and benefits of these new strategies? 4. Did you identify staff training requirements to implement your new plans? 5. Did you design procedures to implement a given program? 6. Did you determine how existing staff and facilities and resources could be used to implement the new program? 7. Did you set a schedule of events and activities to guide the implementation of the new program? 8. Did you identify possible side effects which might result from implementing the new program? 9. Did you assess the attitudes of students, parents, teachers, etc., toward the new program? Did you assess their knowledge of the new program? 10. Did you determine how this program should be administered, evaluated, and reviewed at various levels? 11. Does your design involve evaluation during the implementation process? 12. Does your design involve evaluation of the outcomes of the program? |
|-----|----|--|

Providing Feedback to Project Director

1. Did you develop an implementation plan for your new program?
2. Did you determine the adequacy of the resources, physical facilities, staff and time schedules?

Providing Feedback to Project Director

- | | | |
|-----|----|--|
| YES | NO | |
|-----|----|--|
3. Did you determine what kind of feedback is needed during piloting and implementation?
 4. Did your feedback include monitoring the various publics' understanding of and agreement with your program?
 5. Did you monitor interpersonal relationships among staff and students?
 6. Did you design and assess communication channels?

Assessing the Program's Effectiveness

1. Did you determine whether the program is achieving its objectives?
2. Did you assess whether or not students' needs are being met by the program?
3. Did you assess the gain or loss in pupil achievement?
4. Did you identify unanticipated outcomes and their effects on the students?
5. Did you determine the effects of the project on staff as well as students?
6. Did you assess the attitude of students, staff, parents, and community regarding the outcomes of the program?
7. Did you determine whether or not the program results justify finances and efforts needed to maintain it?

- SECOND DRAFT -

The task of educational program evaluation has been conceptualized into a model by the Phi Delta Kappa National Study Committee on Evaluation. This model, known as the CIPP evaluation model, divides the process of educational program evaluation into four stages: context; input; process; product.

The following is a list of evaluation tasks that can be done in the process of doing a program evaluation. They are divided into four phases or aspects of evaluation. Check the tasks which your evaluation committee completed in doing your program evaluation.

Formulating Objectives

- | | | |
|-----|----|--|
| YES | NO | |
|-----|----|--|
1. Did you identify the learning outcomes which existed as a result of your current program?
 2. Did you identify the learning outcomes you hope to achieve with your new program?
 3. Did you identify the needs that were not being served by your current program?
 4. Did you identify the potential human resources, such as faculty, staff or volunteers?
 5. Did you identify the potential material resources, such as classroom space, audio visual materials, supplementary materials or library services?
 6. Did you gather information from sources outside your district such as research findings or outside consultants?
 7. Did you explore other available programs in terms of the impact of change on students, faculty, parents and community?
 8. Did you assess community values, attitudes, and priorities concerning a new program?
 9. Did you consult various data bases such as standardized test results, or parent surveys?
 10. Did you identify the discrepancies that exist between your present program and the objectives you outlined for any new program?

- SECOND DRAFT -

Examining Present Resources and Developing an Implementation Plan

YES NO

1. Did you determine what you were already doing to meet your new set of objectives?
2. Did you determine what new strategies you could employ to meet the objectives?
3. Did you identify both the costs and benefits of these new strategies?
4. Did you identify staff training requirements to implement your new plans?
5. Did you design procedures to implement a given program?
6. Did you determine how existing staff and facilities and resources could be used to implement the new program?
7. Did you set a schedule of events and activities to guide the implementation of the new program?
8. Did you identify possible side effects which might result from implementing the new program?
9. Did you assess the attitudes of students, parents, teachers, etc., toward the new program? Did you assess their knowledge of the new program?
10. Did you determine how this program should be administered, evaluated, and reviewed at various levels?
11. Did your design involve evaluation during the implementation process?
12. Did your design involve evaluation of the outcomes of the program?

Providing Feedback to Project Director

1. Did you develop an implementation plan for your new program?
2. Did you determine the adequacy of the resources, physical facilities, staff and time schedules?

- SECOND DRAFT -

Providing Feedback to Project Director

YES NO

3. Did you determine what kind of feedback is needed during piloting and implementation?
4. Did your feedback include monitoring the students', faculties and parents' understanding of and agreement with your program?
5. Did you monitor informal interpersonal relationships among staff and students?
6. Did you design and assess formal lines of communication?

Assessing the Program's Effectiveness

1. Did you determine whether the program is achieving its objectives?
2. Did you make provision to assess whether or not students' needs are being met by the program?
3. Did you assess the gain or loss in pupil achievement?
4. Did you identify unanticipated outcomes and their effects on the students?
5. Did you determine the effects of the project on staff?
6. Did you assess the attitudes of students, staff, parents, and community regarding the outcomes of the program?
7. Did you determine whether or not the program's results justify the finances and efforts needed to maintain it?

- SECOND DRAFT -

Examining Present Resources and Developing an Implementation Plan

YES NO

1. Did you determine what you were already doing to meet your new set of objectives?
2. Did you determine what new strategies you could employ to meet the objectives?
3. Did you identify both the costs and benefits of these new strategies?
4. Did you identify staff training requirements to implement your new plans?
5. Did you design procedures to implement a given program?
6. Did you determine how existing staff and facilities and resources could be used to implement the new program?
7. Did you set a schedule of events and activities to guide the implementation of the new program?
8. Did you identify possible side effects which might result from implementing the new program?
9. Did you assess the attitudes of students, parents, teachers, etc., toward the new program? Did you assess their knowledge of the new program?
10. Did you determine how this program should be administered, evaluated, and reviewed at various levels?
11. Did your design involve evaluation during the implementation process?
12. Did your design involve evaluation of the outcomes of the program?

- SECOND DRAFT -

Providing Feedback to Project Director

YES NO

1. Did you develop an implementation plan for your new program?
2. Did you determine the adequacy of the resources, physical facilities, staff and time schedules?
3. Did you determine what kind of feedback is needed during piloting and implementation?
4. Did your feedback include monitoring the students', faculties' and parents' understanding of and agreement with your program?
5. Did you monitor informal interpersonal relationships among staff and students?
6. Did you design and assess formal lines of communication?

Instrument Validation: Background summary of curriculum and
instruction doctoral seminar group

Position

- | | |
|--|---------------------------------------|
| 1. High School History Teacher | Chicago Board of Education |
| 2. Teacher/Director Alternative High School | Chicago Board of Education |
| 3. Director Elementary School Reading Laboratory | Chicago Board of Education |
| 4. School Facilities Analyst | State Board of Education |
| 5. Chairperson, Department of Education | Private Illinois College |
| 6. Director of Early Childhood Education | Private Illinois College |
| 7. Curriculum Coordinator, Department of Occupational Therapy | Public University Medical Center |
| 8. English Department Chairperson | Chicago Suburban High School |
| 9. Teacher/District Teacher Inservice Coordinator | Chicago Suburban Elementary School |
| 10. Superintendent of Training | Public Utility Company |

APPENDIX B

- THIRD DRAFT -

Research Instrument: Phase I

The task of educational program evaluation has been conceptualized into a model by the Phi Delta Kappa National Study Committee on Evaluation. This model, known as the CIPP Evaluation Model, divides the process of educational program evaluation into four stages: Context; input; process; product.

The following is a list of evaluation tasks that can be done in the process of doing a program evaluation. They are divided into four phases or aspects of evaluation. Check the tasks which your evaluation committee completed in doing your program evaluation. A number of examples are given under some tasks. If your committee completed at least one of the suggested examples, please answer yes for that item.

Context Evaluation

Purpose: To provide a rationale for determining program objectives.

YES NO

1. Did you identify the learning outcomes which existed as a result of your current program?
2. Did you identify the learning outcomes you hope to achieve with your new program?
3. Did you identify the needs that were not being served by your current program?
4. Did you identify the potential human resources, such as faculty, staff or volunteers?
5. Did you identify the potential material resources, such as classroom space, audio visual materials, supplementary materials or library services?
6. Did you gather information from sources outside your district, such as research findings or outside consultants?
7. Did you explore other available programs in terms of the impact of change on students, faculty, parents and community?
8. Did you assess community values, attitudes, and priorities concerning a new program?

- THIRD DRAFT -

Context Evaluation

Purpose: To provide a rationale for determining program objectives.

YES NO

9. Did you consult various data bases such as standardized test results, or parent surveys?
10. Did you identify the discrepancies that exist between your present program and the objectives you outlined for any new program?

Input Evaluation

Purpose: To provide information for determining how to utilize resources to meet program goals.

1. Did you determine what you were already doing to meet your new set of objectives?
2. Did you determine what new strategies you could employ to meet the objectives?
3. Did you identify the costs of these new strategies?
4. Did you identify the benefits of these new strategies?
5. Did you identify staff training requirements to implement your new plans?
6. Did you design procedures to implement a given program?
7. Did you determine how existing staff could be used to implement the new program?
8. Did you ascertain how present facilities and resources could be used to implement the new program?
9. Did you set a schedule of events and activities to guide the implementation of the new program?
10. Did you identify possible side effects which might result from implementing the new program?
11. Did you assess the attitudes of students, parents, teachers, etc. toward the new program?

- THIRD DRAFT -

Input Evaluation

Purpose: To provide information for determining how to utilize resources to meet program goals.

YES NO

12. Did you assess students', parents', and teachers' knowledge of the new program?
13. Did your design involve evaluation during the implementation process?
14. Did your design involve evaluation of the outcomes of the program?

Process Evaluation

Purpose: To provide periodic feedback to persons responsible for implementing the new program.

1. Did you develop an implementation plan for your new program?
2. Did you determine the adequacy of the resources?
3. Did you determine what kind of feedback is needed during piloting and implementation?
4. Did your feedback include monitoring the students', faculties', and parents' understanding of and agreement with your program?
5. Did you assess the physical facilities employed in program implementation?
6. Did you account for staff effectiveness in making the program operational?
7. Did you determine the adequacy of time schedules in facilitating program operation?
8. Did you monitor informal interpersonal relationships among staff and students?
9. Did you design and assess formal lines of communication?

- THIRD DRAFT -

Product Evaluation

Purpose: To measure and interpret attainments during the implementation and duration of the program.

YES NO

1. Did you determine whether the program is achieving its objectives?
2. Did you make provision to assess whether or not students' needs are being met by the program?
3. Did you assess the gain or loss in pupil achievement?
4. Did you identify unanticipated outcomes and their effects on the students?
5. Did you determine the effects of the project on staff?
6. Did you assess the attitudes of students, staff, parents, and community regarding the outcomes of the program?
7. Did you determine whether or not the program's results justify the finances and efforts needed to maintain it?

Dear Colleague,

We are completing a graduate program in curriculum and instruction at Loyola University and would greatly appreciate your participation in a research study which we are conducting. Your participation involves completing a questionnaire which requires approximately fifteen minutes.

This study is being conducted in several elementary school districts in which classroom teachers have been involved in the process of evaluating educational programs. A random sample of fifty classroom teachers and all curriculum supervisors in your school district are being asked to participate in the study.

We have received permission to conduct this study from your school district's administrative office. Anonymity to you and the school district is guaranteed in all phases and reports of this study. The results of the study will be available in each participating school district.

Although your participation is voluntary, we are asking you to please take a few minutes to participate in this research endeavor. We thank you in advance for your participation.

Sincerely,

Michael Palmisano
Kay Smith

Please return your questionnaire by U.S. Mail in the stamped addressed envelope by _____ . Thank you.

- FINAL RESEARCH INSTRUMENT -

Research Instrument: Phase II

The following is a list of evaluation tasks that can be done in the process of doing a program evaluation. They are divided into four phases or aspects of evaluation. Check the tasks which your evaluation committee completed in doing your program evaluation. A number of examples are given under some tasks. If your committee completed at least one of the suggested examples, please answer yes for that item.

Context Evaluation

Purpose: To provide a rationale for determining program objectives.

YES NO

1. Did you identify the learning outcomes which existed as a result of your current program?
2. Did you identify the learning outcomes you hope to achieve with your new program?
3. Did you identify the needs that were not being served by your current program?
4. Did you identify the potential human resources, such as faculty, staff or volunteers?
5. Did you identify the potential material resources, such as classroom space, audio visual materials, supplementary materials or library services?
6. Did you gather information from sources outside your district, such as research findings or outside consultants?
7. Did you explore other available programs in terms of the impact of change on students, faculty, parents and community?
8. Did you assess community values, attitudes, and priorities concerning a new program?
9. Did you consult various data bases such as standardized test results, or parent surveys?
10. Did you identify the discrepancies that exist between your present program and the objectives you outlined for any new program?

- FINAL RESEARCH INSTRUMENT -

Input Evaluation

Purpose: To provide information for determining how to utilize resources to meet program goals.

YES NO

1. Did you determine what you were already doing to meet your new set of objectives?
2. Did you determine what new strategies you could employ to meet the objectives?
3. Did you identify the costs of these new strategies?
4. Did you identify the benefits of these new strategies?
5. Did you identify staff training requirements to implement your new plans?
6. Did you design procedures to implement a given program?
7. Did you determine how existing staff could be used to implement the new program?
8. Did you ascertain how present facilities and resources could be used to implement the new program?
9. Did you set a schedule of events and activities to guide the implementation of the new program?
10. Did you identify possible side effects which might result from implementing the new program?
11. Did you assess the attitudes of students, parents, teachers, etc. toward the new program?
12. Did you assess students', parents', and teachers' knowledge of the new program?
13. Did your design involve evaluation during the implementation process?
14. Did your design involve evaluation of the outcomes of the program?

- FINAL RESEARCH INSTRUMENT -

Process Evaluation

Purpose: To provide periodic feedback to persons responsible for implementing the new program?

YES NO

1. Did you develop an implementation plan for your new program?
2. Did you determine the adequacy of the resources?
3. Did you determine what kind of feedback is needed during piloting and implementation?
4. Did your feedback include monitoring the students', faculties', and parents' understanding of and agreement with your program?
5. Did you assess the physical facilities employed in program implementation?
6. Did you account for staff effectiveness in making the program operational?
7. Did you determine the adequacy of time schedules in facilitating program operation?
8. Did you monitor informal interpersonal relationships among staff and students?
9. Did you design and assess formal lines of communication?

Product Evaluation

Purpose: To measure and interpret attainments during the implementation and duration of the program.

1. Did you determine whether the program is achieving its objectives?
2. Did you make provision to assess whether or not students' needs are being met by the program?
3. Did you assess the gain or loss in pupil achievement?
4. Did you identify unanticipated outcomes and their effects on the students?
5. Did you determine the effects of the project on staff?

- FINAL RESEARCH INSTRUMENT -

Product Evaluation

Purpose: To measure and interpret attainments during the implementation and duration of the program.

YES NO

6. Did you assess the attitudes of students, staff, parents, and community, regarding the outcomes of the program?
7. Did you determine whether or not the program's results justify the finances and efforts needed to maintain it?

APPENDIX C

A RESEARCH STUDY OF PROGRAM EVALUATION

Conducted by:

Michael J. Palmisano

and

Kay M. Smith

Loyola University

Chicago, Illinois

DESCRIPTION OF PROPOSED RESEARCH STUDY

The purpose of this study is to provide information to help understand substantively the discrepancy that exists between educational evaluation theory and the evaluation practices of local school districts. These evaluation practices are criticized in the literature for their emphasis on measurement and educational outcomes, their overall lack of comprehensiveness, and their lack of systematic efforts to obtain accurate and relevant information. This situation coexists with the availability of a body of educational evaluation theory and methodology.

Contrary to the current trend of the evaluation literature, several school districts have been identified which conducted systematic comprehensive curriculum evaluations for the purpose of selecting a new basal reading series. These districts offer a unique opportunity for studying comprehensive program evaluation at the local level.

The study describes two aspects of program evaluation in each of these districts:

1. Classroom teachers' and curriculum supervisors' perceptions of educational program evaluation; and
2. The tasks pursued in the process of an educational program evaluation.

Two research instruments have been devised to describe the perceptions and practices of educational practitioners in terms of the CIPP Evaluation Model developed by the Phi Delta Kappa National Study Committee on Evaluation. In

the first phase of the study, classroom teachers and curriculum supervisors are asked to assess the relative importance of the four types of program evaluation and a series of evaluation tasks representative of each type of evaluation delineated in the CIPP Evaluation Model. The study also attempts to determine whether perceptions of program evaluation are related to such educator variables as: position, experience, level of education, major area of graduate study, or experience on an inservice curriculum evaluation committee. In the second phase of the study, the members of the evaluation committee in each district are asked to identify the tasks their committee pursued in the process of selecting a new basal reading program. Several members of this committee are asked to participate in an interview for the purpose of identifying how the tasks identified above were completed. Finally, the written evaluation report will be examined to determine which tasks were reported in this document.

This study will not interfere with a participating district's instructional program, nor will it involve students in any manner. Anonymity to the school district is guaranteed and participation by individuals will be on a voluntary basis. The school district will not incur any expense or risk by participating in the study.

Summary of Staff Involvement for Participating School Districts

| | Participants | Role in the Study | Time Required |
|---------|---|--|---------------|
| Phase 1 | A random sample of 60 K-6 classroom teachers to include all members of the evaluation committee. | Complete Research Instrument: Phase 1 | 20 minutes |
| | All district curriculum supervisors. This includes educators whose job descriptions include supervision or assistance to K-6 classroom teachers in their implementation of curriculum such as: principals curriculum coordinators, directors, consultants etc.; and, assistant superintendent for curriculum. | Complete Research Instrument: Phase 1 | 20 minutes |
| Phase 2 | All members of the evaluation committee. | Complete Research Instrument: Phase 2 | 20 minutes |
| | 5-7 members of the evaluation committee. | Participate in an interview for the purpose of explaining how the tasks identified in Research Instrument: Phase 2 were completed. | 45 minutes |

APPROVAL SHEET

The dissertation submitted by Kay Monroe Smith has been read and approved by the following Committee:

Dr. Barney M. Berlin, Chairman
Associate Professor, Department of Curriculum and Instruction
Loyola

Dr. Robert C. Cienkus
Associate Professor, Department of Curriculum and Instruction
Loyola

Dr. Todd J. Hoover
Assistant Professor, Department of Curriculum and Instruction
Loyola

Dr. John S. Shea
Associate Professor
Chairman, Department of English
Loyola

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dec 1, 1980
Date

B. M. Berlin
Director's Signature