1981

An Analysis of Classroom Teachers' and Curriculum Supervisors' Perceptions of the Education Program Evaluation Process in Terms of the CIPP Evaluation Model

Michael J. Palmisano
Loyola University Chicago

AN ANALYSIS OF CLASSROOM TEACHERS' AND CURRICULUM SUPERVISORS' PERCEPTIONS OF THE EDUCATIONAL PROGRAM EVALUATION PROCESS IN TERMS OF THE CIPP EVALUATION MODEL

BY

Michael J. Palmisano

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

Doctor of Education

January

1981

ABSTRACT

for

Michael J. Palmisano

Loyola University of Chicago

AN ANALYSIS OF CLASSROOM TEACHERS' AND CURRICULUM SUPERVISORS' PERCEPTIONS OF THE EDUCATIONAL PROGRAM EVALUATION PROCESS IN TERMS OF THE CIPP EVALUATION MODEL

ABSTRACT

The study describes classroom teachers' and curriculum supervisors' perceptions of the educational program evaluation process. Classroom teachers, elementary school principals, and certificated curriculum personnel were selected from four Chicago metropolitan area school districts which recently underwent a formal program evaluation in reading for the purpose of selecting a new basal reading program. These practitioners were asked to assess the relative importance of the four types of program evaluation and five evaluation tasks representative of each type of evaluation which were delineated by the CIPP Evaluation Model.

The study also investigated the relationship of these educators' perceptions of program evaluation to the variables of: professional position; years of experience; highest level of education; major area of graduate study; and the experience of having served on an inservice program evaluation committee.

Classroom teachers' and curriculum supervisors' perceptions of the importance of context and product evaluation were more often affected by these variables than were their perceptions of the importance of input and process evaluation. Perceptions of context and product evaluation were found to be related to professional position, years of experience and highest level of education. Perceptions of input and product evaluation were found to be related to professional position and years

of experience.

The experience of having served on an inservice evaluation committee was found to affect educators' perceptions of program evaluation more frequently than did any of the other variables. Highest level of education was found to effect perceptions of context evaluation only. No evidence was found indicating that a particular major area of graduate study influenced perceptions of program evaluation.

This study demonstrated that classroom teachers and curriculum supervisors recognized the importance of the four types of evaluation and tasks representative of each evaluation type which are identified by the CIPP Evaluation Model. The findings suggest that educational practitioners have the perceptual base, or readiness, to pursue comprehensive program evaluations. The discrepancy noted between the availability of evaluation theory and methodology and the program evaluation practices of school systems is apparently not due to the inability of practitioners to recognize the need for a comprehensive approach to program evaluation. What appears to be lacking, is supervisory personnel trained in the science of evaluation who can direct the evaluation efforts of school personnel to more comprehensive program evaluations.

ACKNOWLEDGEMENTS

In addition to all of the participants in this study, this author is grateful to the school personnel in each of the school districts who assisted in the data collection. Without such clerical and administrative support, studies of this kind are not possible.

The completion of this study would not have been possible without the guidance and expertise of the members of my committee. Dr. Barney Berlin, chairman, and Dr. Robert Cienkus consistently offered encouragement and critical questions which assisted throughout the completion of this work. The author is particularly indebted to Dr. Todd Hoover who, as methodology director, assisted the author in formulating research questions and helped with the voluminous computer data analyses.

I am greatly indebted to my colleague and friend, Kay Smith. Our collaboration in this research endeavor was vital to delineating the problem and collecting the data. I am also deeply appreciative to Dr. Frank Tavano who, as my friend and superior, provided encouragement and many opportunities to take the time necessary to complete this study. Also, I am grateful to my friend and former secretary, Nancy James, who typed numerous correspondences and drafts of the research instrument in addition to providing encouragement.

Finally, I would like to thank my family for their understanding, patience and encouragement. Although she is too young to appreciate its significance, I am greatly appreciative of my daughter Linea's patience and cooperation. Most importantly, however, I am thankful for the strength, encouragement and sacrifice of my wife Lynnette, who has truly been a partner in this study.

VITA

MICHAEL J. PALMISANO

PERSONAL DATA	Birthdate: 7/24/47 Married, one child	578 Tarpon Court Schaumburg, Illinois 60193 (312) 893-3759
PROFESSIONAL EXPERIENCE	1969-1972	Classroom teacher Elementary and Junior High School, Lincoln School, School District No. 81, Schiller Park, Illinois
	1972-1976	Assistant Principal, Intermediate team leader, John Muir School, School District No. 54, Schaumburg, Illinois
	1976-1980	Learning Resource Center Director, Frederick Nerge School, School District No. 54, Schaumburg, Illinois
	1980-	Curriculum Director, School District No. 102, Prairie View, Illinois
EDUCATION	Chicago Public Schools B.S. (Psychology) Loyola University, June 1969 M.Ed. (Guidance and Counseling) Loyola University, February, 1972	
PROFESSIONAL QUALIFICATIONS	Secondary Teaching Certificate, 1970 Elementary Teaching Certificate, 1971 Administrative Certificate, 1972 School Services Certificate (Guidance), 1972	
PROFESSIONAL AFFILIATIONS	Phi Delta Kappa Association for Supervision and Curriculum Development Illinois Association for Supervision and Curriculum Development Lake County Curriculum Resource Council	

TABLE OF CONTENTS

LIST OF TABLES

CONTENTS FOR APPENDIXES

CHAPTER I

OVERVIEW OF THE PROBLEM

Introduction

Twenty three years have elapsed since the Russian launching of Sputnik I in September, 1957. This event challenged American educators and lawmakers to rethink educational priorities and practices. During this period, American educational services and opportunities flourished in an unprecedented era of governmental concern for all levels of education. The National Defense Education Act of 1965, coupled with the increase of state and local funding during this period, fostered the expansion and development of educational programs from early childhood through graduate education. Never before were so many Americans involved in education.

This proliferation of educational services continued into the 1970's. During that decade, however, forces emerged which compelled educators to demonstrate the effectiveness of their programs and threatened the development of new programs. On the popular level, individuals and community groups became increasingly vocal in expressing their educational concerns. Local schools and boards of education were compelled to form advisory groups to provide a medium for incorporating these concerns into school policy-making and administration. The

1970's also witnessed an unprecendented rejection of tax and bonding referenda to maintain and expand educational programs. On the legislative level, state and federal legislatures became more explicit in their requirements that funded programs be evaluated to demonstrate their effectiveness. By 1980, the majority of state legislatures had responded to a growing public disenchantment with the cost and outcomes of present educational programs by enacting minimum competency legislation requiring local school boards to establish minimum competencies for promotion and graduation. Thus, a popular and legislative mandate emerged for educators to be accountable for the effectiveness of their educational programs.

Statement of the Problem

A search of the literature of educational program evaluation has revealed that this mandate for educators to demonstrate the effectiveness of their programs has not been satisfied. In several sources, Daniel Stufflebeam and Egon Guba have described the status of educational program evaluation during the years immediately following the Elementary and Secondary Education Act (ESEA). They have identified several factors responsible for this condition.

First of all, those involved in educational evaluation did not adhere to a single definition of educational evaluation. Consequently, evaluation was equated with such practices as

administering tests, determining the congruence of performance to previously determined objectives, or exercising professional judgment on the basis of observation (Stufflebeam, Foley, Gephart, Hammond, Merriman, and Provus, 1971).

Secondly, although evaluations were conducted for the purpose of demonstrating the effectiveness of educational programs, evaluations were characteristically conducted as research studies. The assumptions and methodologies of research were mistakenly equated with those of evaluation. Evaluation consisted of comparing randomly assigned treatment and control groups based on quantitative measures and augmenting the often inconclusive results with masses of illustrative data (Guba & Stufflebeam, 1970).

Finally, due to this confusion over the nature and process of evaluation, evaluations of this period failed to provide decision-makers with the data necessary to judge the value of an educational program or to determine how an educational program might be reshaped to become more effective (Guba, 1967).

Guba and Stufflebeam (1970) identified several areas of need in the interest of ameliorating the above deficiencies in the practice of educational program evaluation. These needs were: "(a) A need for trained evaluators, (b) a need for appropriate evaluation instruments and procedures, and most crucially, (c) a need for evaluation theory" (p.9).

In the decade that followed this pronouncement, attempts were made to satisfy these needs. Regional centers for the study of educational evaluation were organized and maintained through ESEA funding. A body of evaluation theory and methodology also evolved from the contributions of numerous authors. In spite of these circumstances, however, the current evaluation practices of school districts have been criticized on several counts: Their essential product orientation (Rose & Nyre, 1977); their dependence upon quantitative measures and lack of qualitative assessment (Eisner, 1979; Stake, 1978); their confusion of evaluation with research (Cronbach, 1978); and their overall lack of comprehensiveness (Popham, 1975).

It is apparent, then, that a discrepancy exists in the theoretical sophistication available for planning and conducting formal educational program evaluation and the current evaluation practices of school systems. Since evaluation theory and methodology is available to educators, there is need for a study which facilitates the transition of contemporary evaluation theory and methodology into practice. This need has been expressed in the introduction to the ERIC TM Report: The Practice of Evaluation (Rose & Nyre, 1977).

> Less than five years ago, our collection of non-journal works on evaluation consisted of a few well-worn monographs and even fewer books. Today, our file drawers and shelves are filled. There are well over a dozen hard-cover books complete with artist-designed jackets; most written in the last two or three years. But, with all their instructional value, there is not one casebook among them that describes

> real-world evaluations in the context of recommended evaluation models and designs. After all the theory has been studied and the methodologies learned, only such a book can provide guidance to fledgling evaluators (or even seasoned ones) in the practice of program evaluation. (p.1)

This study attempts to make a contribution to contemporary evaluation literature by describing two groups of educational practitioners' perceptions of the process of program evaluation within the context of a major evaluation model. The CIPP Evaluation Model developed by the Phi Delta Kappa National Study Committee on Evaluation has been selected to serve as the referent through which classroom teachers' and curriculum supervisors' perceptions of the process of program evaluation are described.

Purpose of the Study

The purpose of this study is to provide information to help understand substantively the apparent discrepancy that exists between the sophistication of evaluation theory and methodology and the current program evaluation practices of school systems. The study describes two groups of educational practitioners' perceptions of the process of educational program evaluation as delineated in the CIPP Evaluation Model. Cronbach (1978) expressed the belief that the process of evaluation is misunderstood by educational practitioners conducting program evaluations:

> I will say, though, that one of our strongest motivations in pressing for a reform of evaluation is the sense that some enormous fraction of the studies--I might go as high

> as 90 percent--are not worth the effort. The talent of many keen young social scientists is going down the drain because the task of evaluation is misconceived. (p.22)

This study attempts to facilitate the transition of evaluation theory and methodology into practice by describing the nature of educational practitioners' perceptions of the practice of evaluation. With this understanding, attempts to improve the evaluation efforts of school systems could be focused on specific areas of need.

The need for studies of the systematic involvement of school personnel in evaluation research has been expressed by several authors (Good, Biddle and Brophy, 1975; Talmage, Walberg and Nicholas, 1977). Only several such studies, however, have been reported in the literature. The findings of these studies have demonstrated that educational practitioners--classroom teachers as well as administrators--were willing and able to contribute constructively to the program evaluation process (Cohen, 1975; Novak, 1977; Talmage et al., 1977). Two studies relating specifically to practitioners' perceptions of program evaluation were reported in the literature. These studies focused on educational practitioners' roles in the process of evaluation (Nolin, 1976) and the availability and preferences for specific types of evaluation information (Nevo & Stufflebeam, 1975). Studies describing practitioners' understanding of the process of program evaluation, however, are lacking.

This study describes two groups of educational practi-

tioners' perceptions of the process of program evaluation as it applies to the selection of a new curriculum program. Classroom teachers and curriculum supervisors were selected from several school districts which recently underwent a formal program evaluation in reading for the purpose of adopting a new basal reading program.

The textbook adoption process has been selected as the context from which, for several reasons, program evaluation is described. Textbook adoption has been one of the most crucial decision-making tasks a school district faces in the selection of appropriate instructional materials. This is particularly true in reading, mathematics and science where, in most districts, the textbook has been the arbiter of the curriculum (Talmage, 1972; Talmage et al., 1977). Secondly, the process of textbook adoption has been employed as a context for studying the involvement of practitioners in the process of a program evaluation (Kunder, 1976; Talmage et al., 1977). Also, many school districts have involved classroom teachers, as well as administrators, in the decision-making process of textbook adoption. Finally, textbook adoption in reading has been chosen because school districts have traditionally expended their most comprehensive evaluation efforts in this subject area.

The practitioners in this study, classroom teachers and curriculum supervisors, are asked to assess the relative importance of the four types of program evaluation and to assess the

relative importance of a series of evaluation tasks representative of each type of evaluation. The four types of evaluation and the representative tasks were delineated in the CIPP Evaluation Model. Further, the study 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 the experience of having served on an inservice curriculum evaluation committee.

A companion study describes the evaluation practices of these same school districts' program evaluations in reading for the purpose of adopting a new basal reading program (Smith, 1981). This companion study analyzes the tasks each district pursued within the four types of program evaluation which were delineated in the CIPP Evaluation Model.

Research Questions

The following research questions were presented in the study to describe classroom teachers' and curriculum supervisors' perceptions of program evaluation in terms of the CIPP Evaluation Model:

1. What are classroom teachers' and curriculum supervisors' perceptions of the program evaluation process?

2. Do classroom teachers and curriculum supervisors differ in their perceptions of the relative importance of the four types of evaluation and tasks representative of each evaluation type?

3. Do the variables of professional position, the experience of having served on an inservice program evaluation committee, years of experience, highest level of education or major area of graduate study affect classroom teachers' and curriculum supervisors' perceptions of program evaluation?

Significance of the Study

The popular and legislative mandates requiring educators to demonstrate the effectiveness of their programs have resulted in the collection of a plethora of data. Attempts to deduce meaningful conclusion from these data, however, have been criticized in the literature for their inconsistency, lack of systematic analysis and their emphasis on measurement techniques. Although a great deal of study has been pursued under the guise of evaluation, the results of many evaluation endeavors have been questionable.

These circumstances need not continue. The theoretical and methodological sophistication necessary to pursue comprehensive educational program evaluation is available in the literature. This discrepancy in evaluation theory and practice has not been studied substantively. Such study requires an examination of the evaluation practices of school districts, the available resources of evaluation theory and methodology, and the understanding of the evaluation process by those conducting evaluations. This study is significant in that, in addition to

examining the literature on educational program evaluation, it specifically focuses on the need for examining the understanding, or perceptions, of evaluation held by two groups of educational practitioners--classroom teachers and curriculum supervisors--who are directly responsible for implementing and evaluating curriculum.

A companion study to this dissertation addresses the need for examination of the program evaluation practices of school districts. It is significant in that it analyzes the decision-making process of educational evaluation. The combined analysis of this study and the companion study describes educational practitioners' understanding and practice of educational program evaluation. Thus these studies will help define the nature of the discrepancy between educational evaluation theory and practice.

Assumptions

1. The field of educational program evaluation is in need of data describing practitioners' perceptions of evaluation.

2. A theoretical body of literature exists which can be applied to describe educational practitioners' perceptions of program evaluation.

Limitations

1. The CIPP Evaluation Model has been chosen as a formal,

comprehensive evaluation model to serve as a referent in describing classroom teachers' and curriculum supervisors' perceptions of program evaluation.

2. Program evaluation is limited to the formal process by which a school district selects a new basal reading program. Such a limitation serves to clarify the study and to serve as a basis for comparisons across school districts.

3. The study is descriptive in nature. Information is reported to add to the discipline of evaluation.

4. The purpose of the study is to describe educational practitioners' perceptions of program evaluation rather than to study an application of the CIPP Evaluation Model per se.

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. A rationale for examining educators' perceptions of educational program evaluation through the CIPP Evaluation Model.

3. An analysis of the CIPP Evaluation Model.

4. Applications of the CIPP Evaluation Model.

The following resources were consulted in searching out current literature:

1. The computerized searches of ERIC; Dissertation Abstracts; and Psych 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

The practice of formal evaluation can be traced to the twenty-second century B.C. in China. Chinese emperors conducted examinations of their officials every third year to ascertain their fitness for continuing in office. Under the Han dynasty (202 B.C.-200 A.D.), elaborately structured written civil service examinations stressing classical scholarship were developed to identify the most highly qualified individuals for government positions.

In the western world, formal evaluation emerged centuries later in the early middle ages in Europe. Early university examinations were conducted orally by university faculty to determine a candidate's eligibility for a degree. The use of written examinations as the primary means of evaluation appeared during the twelfth century following the introduction of paper-making. The development and systematic use of written tests for placement of students and evaluation was pioneered by the Jesuits in the 1600's. This emphasis on written examination flourished and later spread to England and the United States. By the middle of the nineteenth century, the written examination was the recognized basis for such determinations as: the awarding of academic degrees; who should be permitted to practice

a profession; and who should serve in a government position.[1]

The contemporary development of formal evaluation was guided by several influences. The first evidence of formal program evaluation appeared in the United States just prior to this century. Early evaluations reflected the measurement tradition of evaluating individuals by means of written examinations. This tradition persisted into the 1930's when goal-centered or product evaluation was formulated. During this same period, evaluation by professional judgment also came into prominence. It was not until the mid 1960's with the advent of the ESEA that formal program evaluation was studied seriously. Much of the existent body of evaluation theory and methodology has evolved since that time.

The first recorded formal program evaluation in the United States was conducted in 1897-1898 by Joseph Rice, a pediatrician. Rice conducted a comparative study of the spelling performance of students in a large metropolitan school system. He demonstrated that student achievement in spelling was not related to the repeated drills which characterized spelling instruction at that time. Methodologically, the evaluation consisted of administering a spelling test developed by

---

[1]The reader is referred to Du Bois (1970) as an excellent source for the history of evaluation in the measurement tradition.

Rice to over thirty-thousand students. This landmark study served as a prototype for evaluation efforts for the next thirty years (Rose & Nyre, 1977).

During the early twentieth century, the measurement tradition was refined and it became firmly rooted in the practice of educational evaluation (DuBois, 1970). A measurement technology emerged in the form of increasingly sophisticated tools of human ability and achievement testing. The work of Edward Thorndike and other pioneers in the testing movement stressed the importance of measuring human change and educational evaluation became defined in terms of this emergent measurement technology. Educators typically evaluated programs and students through the administration of tests in different subjects.

The measurement tradition in evaluation has been evident in the writing of contemporary researchers and in the current evaluation practices of school districts. Thorndike and Hagan (1969, 1977) and Ebel (1965) have written treatises on educational measurement. In these works, the process of measurement has been the primary means for acquiring data for educational decision-making. This testing orientation to educational evaluation has been characteristic of the evaluation efforts of school systems as they have traditionally based the judgment of student achievement and program effectiveness pri-

marily on the results of testing information (Popham, 1975, pp. 1-17).

The recent growth of the minimum competency testing movement has been further evidence that this tradition has remained a central focus of educational decision-making. The definition of educational competency, or literacy, in terms of test performance, and the contention that tests can be accurate measures of literacy, have been central premises of the efficacy of the minimum competency testing movement (Pipho, 1978).

Several limitations to adopting a measurement orientation to evaluation have been identified in the literature. These limitations have included: the need for discriminating between the processes of measurement and evaluation; the narrow perspective of test data; and the invalid use of test scores. In the measurement tradition, evaluation has been defined as being roughly synonomous with measurement. This definition of evaluation was evident in the writings of the measurement specialsits Thorndike and Hagan (1969, 1977) and Ebel (1965). Worthen and Sanders (1973), however, described the process of evaluation as a decision-making process. Thus, they maintained that the failure of measurement specialists to discriminate between the processes of measurement and evaluation has limited

evaluation to the collection and interpretation of test data.[2]

The limitation of evaluation to measurement practices has been criticized by several authors. Cronbach (1963) pointed out that focusing evaluation on measurement isolated the process of developing paper and pencil tests. The principles pertinent to test construction thereby came to be regarded as the principles of evaluation. Furthermore, Stufflebeam et al. (1971, p. 11) extended this argument by stating that the "real limits" of evaluation then became the limits of instrumental sophistication. Stufflebeam also noted that such an instrumental focus obscured the necessity of value judgments in the evaluation process and ignored those variables for which measurement devices were not available. Such variables included sociological, cultural, economic, sociometric and philosophical influences. This position was endorsed by Cronbach (1978) again in a later source where he directly presented the caveat that test scores were but one of many sources of information necessary for evaluation.

In the same source, Stake (1978) discussed the common practice in which school districts judge the effectiveness and quality of their instructional programs on the results of

---

[2]Thorndike and Hagan (1977) have broadened their conception of testing and measurement presented in earlier editions of their work. Although testing and measurement are presented as an aid for decision-making, this process of decision-making is not defined as a process distinct from measurement.

standardized tests. He pointed out that test developers have yet to provide evidence, i.e. empirical studies, that demonstrated the accuracy of test means as effective indicators of school or district performance. Thus, Stake maintained that the value of test data depended upon "the professional experience and intuition" of the educator using them (p. 276).

A second theme which has influenced the development of formal evaluation emerged in the 1930's as a result of the Tyler and Smith Eight Year Study of the Progressive Education Association (Tyler, 1942). Tyler and Smith defined evaluation as the process of determining the degree to which the objectives of a program have been realized. Evaluation entailed the use of a variety of tests, scales, inventories, questionnaires, checklists, pupil logs and other measures to assess the achievement of high school students in terms of curricular objectives. The Eight Year Study and Tyler's subsequent work broadened the scope of educational evaluation to include the use of a variety of data, systematic processes of assessment, and the focus of evaluation on the acheivement of objectives. This goal-attainment model greatly influenced the character of evaluation for the next three decades (Worthen & Sanders, 1973).

This practice of focusing evaluation on the degree to which students attain instructional objectives has been termed product evaluation. Product evaluation was based on the tech-

nical model of curriculum development and it was characterized by Tyler's evaluation framework (Tyler, 1950). More recently, Tabe (1962) elaborated upon Tyler's curriculum development rationale and stressed this same orientation to evaluation. Hammond (n.d. Mineo, in Worthen & Sanders, 1973) developed an evaluation model in the product orientation to evaluation. Hammond's model was developed for the purpose of assessing the effectiveness of current and innovative programs at the local level by comparing behavioral data with objectives. Many school systems have adopted this product orientation to evaluation when evaluating programs (Guba, 1968; Rose & Nyre, 1977; Womer, 1970).

Several authors have specifically addressed the issue of product emphasis in their critiques of current evaluation practices of school systems. Their criticisms have focused on several issues: the limitations of product data; the need for assessment of educational goals; and the need for qualitative as well as quantitative assessment.

Scriven (1967) contended that product data are essential, but limited data. He criticized a strictly product oriented evaluation as being little more than an estimation of goal achievement. Bloom, Hastings and Madus (1971) maintained that evaluation should begin with an assessment of the goals of a program. They 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 (1975) maintained that a product orientation focuses attention on the quantitative aspect of curriculum evaluation and thereby limits essential qualitative analysis.

During the same period as goal-attainment emerged as the focus of educational evaluations, another means of evaluation came into prominence, evaluation by accreditation. Popham (1975) defined evaluation in the accreditation mode as the judgment of programs by intrinsic criteria. The accreditation movement began in the 1870's at the University of Michigan. High schools of that state were invited to seek university approval of their academic programs which thereby enabled their graduates to be admitted to the university without examination. Later, the power to approve high school educational programs was delegated to regional accrediting associations which set standards as a basis of membership. Finally, state departments of education constituted a third agency for accrediting schools. Presently, they constitute the primary means of elementary and secondary school program appraisal (Glass, 1969).

Stake (1967; 1973) and Glass (1969) have both criticized evaluation by accreditation. Both authors maintained that the merit of evaluation by accreditation depended upon the validity of the standards used by the evaluators, the kinds of data col-

lected for making decisions and the competencies of the evaluators. Stake (1967) further maintained that evaluation by accreditation can have a positive value as a catalyst in encouraging the refinement of a developing curriculum. Both authors, however, have also pointed out what Glass (1969) referred to as "the genetic flaw" in the accreditation model, "Evaluation will not enhance the value of an educational program if it demands conformity to standards which themselves cannot be demonstrated to lead to valued goals" (p. 27).

The modern history of educational evaluation began in the late 1960's. An immediate public response to the Russian launching of Sputnik I in 1957 was to seriously question the quality of American schools. What transpired was a mandate for reform. At the same time, the emerging power of civil rights groups also pressed for reform in fair treatment of minority children in schools. These forces prompted the federal government to provide a greater share of the schools' financial support. This support came in the ESEA of 1965.

Through its various titled programs, the ESEA provided for thousands of grants to educational agencies throughout the country. To hold the receiving agencies accountable to the federal government for its investment in local programs, monies granted under Titles I and III carried the proviso that programs be evaluated by the receiving agency in order to

continue receiving federal funds. Annual evaluations of the effectiveness of funded programs were to be filed with the federal government to insure that federal funds were accomplishing their intended purposes. Thus funding was provided for the specific purpose of evaluation of educational programs (Taylor, 1974).

What ensued was a massive demand for evaluators and evaluation technology. The demands, however, could not be satisfied as few trained evaluators were available and evaluation technology was not sufficiently developed to handle the diverse needs of the multitude of programs requiring evaluation. Consequently, the evaluations which were conducted on local, state and federal levels were stongly criticized on several counts (Rose & Nyre, 1977).

The premier criticism of the early ESEA mandated evaluations was that these evaluations were conducted as research endeavors. Guba and Stufflebeam (1970) specifically addressed this criticism of these early evaluations in a major position paper:

> The authors of this paper have taken a rather specific position in this controversy, rejecting the proposition that evaluation is equivalent to research, that is, that the same assumptions and methodologies hold for the two fields. The writers assert that professors of educational research are largely to blame for the confusion and ineptness which persist in the field of evaluation. The authors think many researchers make wrong assumptions about what an evaluation study should accomplish, and that, based on these erroneous assumptions, researchers foist bad

> advice upon unsuspecting and unsophisticated practitioners. As a consequence, evaluations are usually useless, and practitioners are largely justified in the jaundiced view they typically have taken about evaluation and its utility. (p. 7)

The result of these early ESEA evaluations having been conducted in a research model yielded several specific criticisms. Evaluations were heavily product oriented and essentially consisted of testing students and compiling masses of data to file in the evaluation reports. Methodologically, comparisons of randomly assigned treatment and control groups became the standard evaluation practice. Thus, the practice of testing and reporting quantitative data coupled with vast amounts of illustrative data became the sine qua non of program evaluation. The resulting evaluation reports were of little use to the federal government (Rose & Nyre, 1977).

Guba (1967) pointed out that the lack of trained evaluators was not the only reason these early evaluations were so strongly criticized. In disseminating ESEA legislation information to education agencies receiving federal funds, the United States Office of Education did not provide adequate guidelines as to the content or process required of the mandated evaluations. In the absence of such guidelines, evaluation designs and reports were drafted by inexperienced personnel at the receiving agencies. This situation led Guba to conclude that "The present guidelines are markedly inadequate, they do little more than

to encourage sloppily conceived product evaluations" (p. 313).

Large scale evaluations conducted by the federal government were also criticized. Guba and Stufflebeam (1970) reviewed many such evaluations and found the reports lacking in the information needed to support decision-making related to the programs being evaluated. They noted several shortcomings. Many reports contained only "impressionistic information" (p. 8) lacking the level of credibility required by decision-makers. Attempts to conduct rigorous research studies characteristically yielded "no significant differences" (p. 8) between experimental and control group results. Yet, those involved in the programs repeatedly reported that these programs were producing such significant differences that the programs could not be terminated. Finally, Title III staff members in the U.S. Office of Education repeatedly ranked the quality of the evaluation reports filed under Title III as "poor"--nearly the lowest ranking of the fifteen quality criteria of Title III projects.[3]

---

[3]Three evaluation reports discussed illustrating the shortcomings which Guba and Stufflebeam noted above included: The First Annual Reports: Title I and III of the Elementary and Secondary Education Act and the evaluation report for New York City's Higher Horizon's Program (Guba & Stufflebeam, 1970).

On the basis of the criticisms of evaluation reports filed with the federal government, Guba and Stufflebeam (1970) identified several fundamental impediments to the prospect of ameliorating the deficiencies present in educational program evaluations. These impediments included: "(a) The lack of trained evaluators and training programs; (b) the lack of appropriate evaluation instruments and procedures; and most crucially, (c) the lack of adequate evaluation theory" (p. 9).

The years following these early ESEA evaluations witnessed the development of more sophisticated evaluation theory and methodology in response to the deficiencies outlined above. What evolved were theories and models of evaluation and a considerable fund of writings about the task of evaluation itself. Popham (1975) suggested four orientations from which to describe the models which have been developed: goal-attainment models; judgmental models emphasizing intrinsic criteria; judgmental models emphasizing extrinsic criteria; and decision-facilitation models.

The goal-attainment models of educational evaluation approach evaluation as "The determination of the degree to which an instructional programs' goals were achieved" (Popham, 1975, p. 22). As discussed earlier, the first application of evaluation as goal-attainment was in Tyler and Smith's Eight Year Study in the 1930's. Tyler presented the goal attainment framework

to evaluation in several sources (1942, 1950, 1958, 1964). Essentially, Tyler recommended the formulation of educational goals according to an analysis of three sources--the student, society and the subject matter--and two screens--a psychology of learning and a philosophy of education. Goals were then to be written as measureable objectives. Evaluation consisted of measuring pupil progress against these measureable objectives.

More recently, Hammond (n.d. Mimeo, in Worthen & Sanders, 1973) developed an evaluation model which focused on determining the effectiveness of innovative programs in achieving expressed objectives. Hammond's model attempted to analyze in detail the nature of the institutional and instructional factors relevant to the realization of program objectives.

Another example of a goal-attainment model was developed by Metfessel and Michael (1967). These authors recommended the use of multiple criterion measures representing a comprehensive assessment of factors influencing the goal-attainment of an educational program.

Popham (1975) categorized judgmental models according to the criteria bases for professional judgment upon which they are based. In these approaches to evaluation, the evaluator exercized considerable influence on the nature and outcome of the evaluation. Popham described both instrinsic and extrinsic applications of judgmental models. He described intrinsic criteria

as process criteria focusing on the nature of a program and extrinsic criteria as product criteria focusing on the effects of a program. Popham also pointed out that judgmental approaches to evaluation emphasizing intrinsic criteria are quite common in education; but, with one major exception, are too haphazard to be classed as systematic evaluation of educational programs.[4] The exception was the accreditation model of educational evaluation discussed earlier in which professional colleagues attempt to identify strengths and deficiencies in educational programs in a prearranged systematic process.

An example of a judgmental approach emphasizing extrinsic criteria was the Countenance Model developed by Stake (1967). Stake differentiated the descriptive and judgmental aspects of evaluation according to phases of program development, implementation and outcome.

The final orientation of evaluation models described by Popham (1975) were the decision-facilitation models. These models were distinct from those discussed above in that the

---

[4]Such judgmental approaches criticized for their narrow conceptual base and emphasis on budgetary and accounting procedures included the approaches developed from the Planning-Programming-Budgeting System (PPBS). (See CASEA Progress Report, 1971 and Jenkins and Lehmen, 1972) A second model, School Planning and Evaluation Communication System (SPECS) was likewise criticized for its narrow focus and especially for the lack of goal validation upon which it is based. (See CASEA Progress Report, 1972).

role of the evaluator in these models was to collect and present information to decision-makers, not to assess the worth or merit of a program. Several models of evaluation have been representative of this orientation.

The Discrepancy Model devised by Provus (1971) specifically focused on the discrepancies between program goals and outcomes.[5] The model consisted of five evaluation stages in the process of defining program standards, determining discrepancies between program performance and program goals and providing this discrepancy information to decision-makers to determine whether the program should be terminated, unaltered or to require a change in performance or standards.

The CIPP Evaluation Model developed by the Phi Delta Kappa Committee on Evaluation (Stufflebeam, et al., 1971) approached

---

[5]That authors have exercised license in the categorization of evaluation models is illustrated in the case of the Discrepancy Model. Worthen and Sanders (1973) have identified three frameworks from which to describe evaluation theory and models: (a) judgmental; (b) decision-management; and (c) decision-objective. While these frameworks correspond to Popham's (1975) orientations--judgmental, decision-facilitation and goal-attainment respectively--Worthen and Sanders classified the Discrepancy Model as decision-objective focusing on its product orientation to the delineation of information for decision-makers.

evaluation as a collaborative process between evaluators and decision-makers. Evaluation entailed the delineating, obtaining and providing of information to decision-makers wherein the delineation and providing functions were collaborative and the obtaining function was the technical role of the evaluator. The CIPP Evaluation Model distinguished between four decision-settings, four types of evaluation and the decision-making and accountability roles of evaluation.[6]

A third decision-facilitation model was developed by the UCLA Center for the Study of Evaluation (CSE) and was described by Alkin (1969). Although similar to the CIPP Evaluation Model in many respects, the process evaluation stage was reconceptualized to attend to enroute products as well as process of the program being evaluated. The model delineated five stages of evaluation, each with a specific decision focus: needs assessment; program planning; implementation evaluation; progress evaluation; and outcome evaluation.

---

[6]The CIPP Evaluation Model is discussed in detail in a subsequent section of this chapter.

In addition to the theories and models discussed above, other authors have contributed significantly to the growing fund of evaluation literature. Although their contributions have not been in the form of a model, their work has been influential in the development of contemporary evaluation theory and methodology.

Scriven (1967, 1974) did not delineate a formal evaluation model, but his extensive writings have contributed importantly to the emerging comprehensive perspective of contemporary evaluation theory. These contributions included the first distinction between the formative and summative roles of evaluation, attention to the quality of goals, the description of evaluation focusing on extrinsic criteria as "Payoff Evaluation," an emphasis on comparative evaluation, the need for a "goal-free" component of evaluation, and the "Modus Operandi Method" useful to educational evaluators in situations where experimental or quasi-experimental approaches have not been feasible.

Popham (1975) also has contributed to the literature of educational evaluation. He described evaluation as a holistic, systematic and adaptive process. He maintained that it is necessary to approach educational program evaluation with a "gestalt" of the evaluation process. Such an approach focuses on the total program in the context in which it operates, is

an on-going cyclic process, and adapts the evaluation itself to meet the specific needs of the program under consideration.

Cronbach (1978) expressed a similar view regarding the adaptive role of evaluation. He described the need for an eclectic approach to the task of devising an evaluation strategy to meet the needs of a given program. Cronbach claimed that the best approach to evaluative inquiry was a "mix of studies. The evaluation planners task is to find the right mix among styles (within as well as between studies) to enlighten persons concerned with the program" (p. 22).

Bloom and his associates (1971) extended the distinction between the formative and summative roles of evaluation originally described by Scriven (1967). They utilized this distinction extensively in their treatment of evaluation as it applied to the development, implementation and refinement of instructional programs. They also contributed significantly to another aspect of evaluation theory--the need for evaluating the goals and objectives of an instructional program. This theme was also expressed by Popham (1972) and by Grobman (1968, 1970).

Eisner (1979) advanced the view that conventional modes of evaluation examined only a "slender slice of educational life" (p. 20). He advocated the development and use of alternative methodology to examine those competencies necessary in "conceptualizing, expressing and recovering meaning" (p. 13).

Eisner's work stressed the need for a qualitative focus to evaluation in contrast to the essentially quantitative approaches which have typically been pursued.

Guba and Stufflebeam (1970), in addition to their work on the CIPP Evaluation Model, championed the case for clearly differentiating the purpose and function of evaluation and research. These authors and others (Cronbach, 1978; Worthen & Sanders, 1973) have contributed significantly to the emergence of evaluation as a process qualitatively distinct from research and they have pointed out the shortcomings of approaching evaluation in the traditional social science research tradition of comparing the performance of a control group and experimental group on the basis of null hypotheses.

Finally, a number of authors have contributed to the development of an evaluation activity designed to assess the merits of proposed and completed evaluation efforts. This activity, meta-evaluation, was defined by Scriven (1969) as a procedure for describing an evaluation activity and comparing it against a set of ideas concerning what constitutes good evaluation. Several authors have described procedures for conducting meta-evaluations (Gowin, 1978; Scriven, 1969; Stufflebeam, 1967, 1977). Stufflebeam (1977) stated two purposes of meta-evaluation. On the formative level, meta-evaluation is a means for assuring that evaluations will produce results which are valid, useful and

cost-effective. On the summative level, meta-evaluation provides a system of accountability for the evaluator.

In conclusion, these authors' contributions represent the major influences which have guided the development of contemporary evaluation theory and methodology. These numerous and somewhat diverse viewpoints make up a comprehensive fund of evaluation literature which educational theoreticians and practitioners can draw upon in their writings and practices of evaluation.

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

Cronbach (1978) summarized the criticisms of contemporary program evaluation practices in his contention that "The task of evaluation is misconceived" (p. 20). Attempts to study the involvement of educational practitioners in evaluation research and the ability of practitioners to contribute constructively to evaluation research efforts are well documented in the literature. What is lacking, however, are attempts to specifically describe educational practitioners' perceptions, or understanding, of the program evaluation process.

The need for the systematic involvement of educational practitioners in evaluation research has been expressed by numerous authors. In the early 1950's, Corey (1953) described an action research design which involved field investigation by practitioners (teachers, supervisors and administrators) to improve educational practice. The methodology associated with action research was strongly criticized and by the late 1950's few research studies made reference to action research (Clifford, 1973; Hogkinson, 1957).

Renewed interest in action research has been expressed by several authors. Talmage et al. (1977) developed an evaluation research model which involved an action research design carried out in a naturalistic setting using process and attitude

measures as well as achievement measures. The authors acknowledged the earlier criticisms of action research and expressed the view that closer collaboration between evaluation consultants and school personnel, recent advances in statistical analysis and greater sophistication in research methodology may overcome the former methodological weaknesses in action research. This view has also been expressed by Good et al. (1975) who contended that there is both need for and value to the systematic involvement of school personnel in evaluation research.

Several researchers have demonstrated that educational practitioners can make important contributions to evaluation research. In the study described earlier, Talmage et al. (1977) examined the involvement of classroom teachers in the process of adopting a new basal reading program from several alternative basal reading series. The results of this study suggested several important considerations for school districts in their textbook selection practices.

In another study, Novak (1977) integrated the findings of research in evaluation and applied these to the involvement of the potential users of this evaluation information. Novak provided an opportunity for teachers and administrators to contribute to an evaluation design and implementation plan for a reading disabilities program. From his findings, Novak identi-

fied guidelines for effective, meaningful involvement of teachers and administrators in program evaluation.

In a study of the effectiveness of team-teaching programs, Cohen (1975) identified several characteristics present in successful team-teaching situations. These characteristics included the involvement of teachers and administrators in the process of curriculum and program evaluation on a short and long term basis.

In another study, Nolin (1976) employed an adaptation of the CIPP Evaluation Model to meet legislative and professional aspirations for the involvement of school staff and community in planning and evaluating educational programs. Nolin concluded that school administrators perceived a need for both the involvement of teachers and community in the educational program decision-making process, and, that a vehicle for incorporating this involvement was needed.

Finally, in a study by Nevo and Stufflebeam (1975) the authors examined the evaluation priorities of teachers, principals and students. These school personnel were asked to assign priorities to alternative information items derived from the CIPP Evaluation Model which were useful in educational evaluation. The authors found that school personnel perceived context and product information as the most available evaluative information in schools and that they perceived a need for other kinds of evaluative information.

The research studies described above demonstrate that the study of educational practitioners' involvement in evaluation research is both necessary and fruitful. The findings of these studies, however, are lacking in information regarding how practitioners perceive, or understand, the program evaluation process. If the process of evaluation is misunderstood by those conducting evaluations, as Cronbach (1978) maintains that it is, then information describing practitioners' perceptions of the evaluation process is needed. With such information, efforts to improve the evaluation efforts of practitioners could be focused on specific areas of misunderstanding and misinformation.

The choice of the CIPP Evaluation Model as the referent from which to examine educational practitioners' perceptions of the program evaluation process was based on several factors. These factors included: the need to examine evaluation practices from a well-founded theoretical basis; the underlying principles of the CIPP Evaluation Model; and the demonstrated effectiveness of the model in describing evaluation practices in a variety of applications.

The ERIC/TM Report (Rose & Nyre, 1977) highlighted the discrepancy that exists in the theoretical sophistication available for planning and conducting educational programs evaluation and the current evaluation practices of school systems.

The need for study of evaluation within the context of well-founded evaluation models and designs was suggested as a means for facilitating the transition of contemporary evaluation theory and methodology into practice. The CIPP Evaluation Model is based on a comprehensive and practical theory of evaluation, and, thus satisfies the need for a well-founded theoretical referent from which to conduct evaluation research.

The underlying principles of the CIPP Evaluation Model were derived in response to several specific criticisms of evaluation practices. These principles focused on providing information for decision-making, adapting evaluative strategies to meet the demands of an existing program, conceptualizing types of decisions and evaluation designs generalizable to varied educational decision settings, and satisfying criteria of practical utility as well as scientific adequacy (Stufflebeam & Guba, 1970). The CIPP Model is, therefore, an action oriented practitioners' model useful in describing practitioners' understanding of the program evaluation process.

The effectiveness of the CIPP Evaluation Model has been demonstrated in several theoretical applications. Hinkles (1971) applied the model to assess a hypothetical and complex educational change activity; Reinhard (1973) employed the model to develop alternative evaluation strategies in innovative programs; and McLure (1973) utilized the model for the purpose of

identifying areas of need for social change. Finally, in the study by Nevo and Stufflebeam (1975) the model was employed to assess evaluation priorities of teachers, administrators and students. In this application, the CIPP Evaluation Model was successful in describing educational practitioners' attitudes and needs in regard to evaluation information.

In conclusion, the involvement of educational practitioners in evaluation research has been studied in a variety of contexts. Although practitioners' perceptions of the educational program evaluation process have not been studied per se, their involvement in several studies has demonstrated that practitioners can contribute importantly to evaluation research. What is still needed, however, are studies which focus on practitioners' perceptions, or understanding of the program evaluation process within the context of well-founded evaluation theory. The CIPP Evaluation Model was selected as the context within which to describe practitioners' perceptions of the program evaluation process because numerous and varied applications of the model have resulted in valid and reliable information for decision-making.

An Analysis of the CIPP Evaluation Model

Conceptual Background of CIPP

One of the provisions of the ESEA was that funds were allocated to support the organization and maintenance of regional centers for the study of educational evaluation. One such center was established at Ohio State University under the direction of Daniel Stufflebeam. In 1968, the Phi Delta Kappa Professional Education Fraternity, in conjunction with the Ohio State University Evaluation Center, organized an advisory committee under the chairmanship of Daniel Stufflebeam. This advisory committee was charged with the task of identifying the problems facing educational evaluation. The committee identified three essential problems:

> 1. A lack of understanding of decision processes and information requirements in current programs of educational change;
> 2. The lack of a definition of educational evaluation pertinent to emergent requirements for educational evaluation; and
> 3. A lack of appropriate evaluation designs. (Carter, 1975, p. 2)

The Phi Delta Kappa Advisory Committee recommended the establishment of a National Study Committee on Evaluation to devise a definition of evaluation and to provide a conceptual and methodological description of the process of evaluation based on this definition. Members of the committee were selected from universities and other regional centers for the study of

evaluation. The CIPP Evaluation Model was the result of this committee's work (Gess et al., 1974).

Upon embarking on the study, the committee delineated questions to which program evaluation should address itself:

> What are the needs, problems, and opportunities?
> What decisions need to be made to respond to them?
> What are some possible alternative ways to respond to them?
> What are reasonable bases for choosing among the available alternatives?
> When a choice has been made, how can one know whether the selected response mode works?
> And finally, how can one make the selected response work even better? (Stufflebeam et al., 1971, p.3)

It was the committee's contention that "Evaluation, in its present state, has failed to provide even a minimally acceptable way of responding to these requirements" (Stufflebeam, 1971, p. 4).

The committee's first task in the development of an evaluation theory was to define the task of evaluation. Evaluation was defined as "The process of delineating, obtaining, and applying descriptive and judgmental information; concerning some object's merit; as revealed by its goals, design, implementation, and results; for purposes of decision-making and accountability" (Ewy & Chase, 1977, p. 2). The definition was stated more simply as the process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam, et al., 1971, p. 40).

There were three essential elements in this definition of evaluation. Evaluation was described as an ongoing process. Secondly, evaluation included three main steps: delineating the questions to be answered; obtaining relevent information so the questions may be answered; and providing this information to decision makers. The delineating and obtaining steps were interface activities requiring collaboration between evaluator and decision makers, the obtaining step was a technical activity which was performed by the evaluators. Finally, the purpose of evaluation was to provide relevant information to decision makers.

The Phi Delta Kappa Committee's definition of evaluation provided for both decision-making and accountability. Evaluation was to be performed in the service of decision-making in that it was designed to provide information to decision-makers either for drawing conclusions or projecting future action (Wallace & Shavelson, 1970). Accountability was to be served in acquiring evaluation information which met the scientific criteria of validity and reliability, and, the utility criteria of relevance, significance, scope, credibility, timeliness, pervasiveness and efficiency. Also, by maintaining a record of past decisions and a record of the information that was available to support them, evaluation was to aid decision-makers in being accountable for their past decisions and actions (Ewy & Chase, 1977).

After defining the task of evaluation, the committee's next step was to specify principles from which a theory and model of evaluation could be derived. Five such principles underlied the CIPP Evaluation Model:

> 1. The purpose of evaluation is to provide information for decision making: To evaluate, therefore, it is necessary to know what decisions are to be served.
> 2. Different evaluation strategies are required depending upon the nature of different decision making settings to be served.
> 3. A generalizable evaluation model should be based on a conceptualization of the types of decisions and evaluation designs which are generalizable to all educational decision settings.
> 4. While the content of different evaluation designs varies, a single set of generalizable steps can be followed in the design of any sound evaluation.
> 5. Evaluation studies should answer questions posed by decision makers. Therefore, designs for such studies should satisfy criteria both of scientific adequacy and practical utility. (Guba & Stufflebeam, 1970, p. 15)

Description of CIPP

The CIPP Evaluation Model is a comprehensive, complex evaluation model. In delineating the model, the Phi Delta Kappa Committe identified four types of evaluation, four types of decisions and four decision settings. These characteristics serve as the framework from which the model is described below.[7]

---

[7]This description of the CIPP Evaluation Model was drawn from several sources: Guba and Stufflebeam (1970) a major position paper outlining the CIPP Evaluation Model; Stufflebeam et al., (1971), the report of the Phi Delta Kappa National Study Committee on Evaluation; Stufflebeam (1971), a major article on the relevance of CIPP for accountability; and the ERIC/TM Report: The Practice of Evaluation (Rose & Nyre, 1977), a detailed report on contemporary theories of evaluation.

The four types of evaluation identified by the CIPP Evaluation Model were: context; input; process and product. The purpose of context evaluation was to provide a rationale for the justification of a particular type of program. Providing this rationale required specifying the population to be served, determining the needs of that population, and formulating objectives to satisfy these needs. This process entailed the description of the environment in which the change was to occur; identification of unmet needs and necessary and available resources; identification of sources of problems in meeting these needs; and prediction of future deficiencies by considering the desired, expected, possible and probable outcomes. Context evaluation was a systematic and macroanalytic process.

Four questions were addressed in the delineation of goals and objectives of a program:

> 1. What unmet needs exist in the context served by a particular institution?
> 2. What objectives should be pursued in order to meet these needs?
> 3. What objectives will receive support from the community?
> 4. Which set of objectives is most feasible to achieve? (Rose & Nyre, 1977, p. 18)

The methodology of context evaluation involved a contingency and congruency mode of data collection. In the contingency mode baseline data was collected by searching for opportunities and pressures outside of the system to foster improvement within. In the congruency mode, actual and intended system performances were compared and discrepancy information was reported concerning the agency's statement of

goals, policies, laws and general policies governing education.

The purpose of input evaluation, the second type of evaluation identified in the CIPP Evaluation Model, was to determine how to use resources to meet the objectives established for the program. Input evaluation required the identification and assessment of relevant capabilities of the individuals, agencies or strategies responsible for achieving program goals, and designs for implementing a selected strategy. The goal of input evaluation was to provide an analysis of alternative procedural designs in terms of their potential costs and benefits. Input evaluation performed a "diagnostic" function in the detection of resource problems in implementing a selected strategy. Its function was also "theraputic in problem solving within the overall system" (Stufflebeam, 1971, p. 17)

Four questions were addressed in the determination of how to use resources to meet the objectives of a program:

1. Does a given project strategy provide a logical response to a set of specified objectives?
2. What strategies already exist with potential relevance for meeting previously established objectives?
3. What procedures and time schedules will be needed to implement a given strategy?
4. What are the operating characteristics and effects of alternate strategies under pilot conditions? (Rose & Nyre, 1977, p. 19)

The third type of evaluation identified in the CIPP Evaluation Model was process evaluation. The purpose of process evaluation was to provide feedback to project directors concerning project progress in its implementation phase. The objectives of process evaluation were detection of defects in

the program design or its implementation, and monitoring the various elements of the project to identify and remedy potential problems or sources of failure. These included: interpersonal relationships among staff and students, communication channels, logistics, adequacy of the resources, physical facilities, staff and time schedules.

Four questions were addressed in providing feedback to project directors during implementation:

1. Is the project on schedule?
2. Should the staff be reoriented or retrained prior to completion of the present project cycle?
3. Are the facilities and materials being used adequately and appropriately?
4. What procedural barriers need to be overcome during the present cycle? (Rose & Nyre, 1977, p. 19)

Process evaluation required the delineation, obtaining and reporting of information by project personnel as often as it was required. Thus, in addition to providing continuous feedback for ongoing program improvement, process evaluation also provided a record of the project cycle which may be useful in future reference.

The final type of evaluation identified in the CIPP Evaluation Model was product evaluation. The purpose of product evaluation was to provide information concerning the degree to which overall goals and objectives of a program have been realized. Product evaluation entailed measuring and interpreting attainments as often as necessary during the project term and at

the completion of the project cycle.

Four tasks were specified in product evaluation:

> 1. Identifying congruencies and descrepancies between the intended objective and the actual achievements;
> 2. Identifying unintended outcomes;
> 3. Providing for objectives that have not been met by recycling the program; and
> 4. Providing appropriate information to decision makers regarding the continuation, modification or termination of the program. (Rose & Nyre, 1977, p. 20)

The four types of evaluation formed the basis for the CIPP Evaluation Model. As a comprehensive interrelated network of evaluation processes, they provided for both systematic and ad hoc analyses of a program. Their interrelationship provided continuity in the process of evaluation.

Four decision settings were also delineated in the CIPP Evaluation Model: homeostatic; incremental; neomobilistic and metamorphic. These decision settings arose directly from the authors' definition of evaluation as delineating, obtaining and providing information for decision making. The extensiveness of an evaluation was determined by the impact of the decision that was to be serviced and the availability of information. These two factors formed two intersecting lines which, when combined, yielded four classes of decision settings. The classes were labled "small versus large change" and "high versus low understanding" (Stufflebeam et al., 1971, p. 80). The factor distinguishing small from large change was the degree of contro-

versy associated with the change. Large changes involved major restructuring in an education program; small change dealt with less far reaching decisions.

The four decision settings each referred to the extent of the intended change. Homeostatic decisions were designed for maintaining the status quo. Incremental decisions referred to developmental activities. Neomobilistic decisions denoted major innovative activities, and metamorphic decisions involved drastic changes in school programs.

Each of these four decision settings was crossed with the available knowledge grasp on the part of those to be affected by a program. This class varied from a low degree of knowledge to a high degree of knowledge concerning an educational change.

Within each of these four decision settings, there were four categories of decision types: planning decisions to determine objectives; structuring decisions to design the procedures to be used to achieve the objectives; implementing decisions to monitor and refine the procedures; and recycling decisions to judge the outcomes or attainments of the objectives.

Planning decisions were made to determine the major changes that were needed in a program. They were concerned with such questions as: What are the conditions which are preventing the objectives from being achieved? What priorities

should the program serve? and What new objectives would best service the philosophy and general goals of the program?

The second decision type was structuring decisions which determine the means to be used in a program to attain the program objectives. Structuring decisions specified action to operationalize a program. The prescribed program variables included content, organization, personnel, schedules, human and material resources.

The third decision type was implementing decisions which were concerned with the operational procedures of using resources to make an educational program work. These decisions dealt with such questions as: Should the schedule be modified? Is effective use being made of human and material resources? Are additional personnel and/or resources needed?

The fourth decision type was recycling decisions which were made to determine the congruence between the original objectives and the quality of the project attainments. Recycling decisions indicated whether a program should be continued, modified or terminated. Recycling decisions asked 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?

Finally, the four types of evaluation, decision settings and decision types provided an accountability dimension to

evaluation in the CIPP Evaluation Model. Stufflebeam defined educational accountability as "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. 19).

Context, input, process and product evaluation each provided for accountability. Context evaluation provided the record which identified the objectives which were chosen, the reasons why they were chosen, and the goal-related reasons for the choice of procedural designs. Input evaluation indicated whether stated objectives were the ones pursued and which procedural design was selected. Process evaluation confirmed whether stated objectives were pursued and whether procedural specifications were implemented. Finally, product evaluation provided the report on the degree to which objectives were achieved during and at the end of the project term.

Evaluation conducted for the purpose of accountability was retroactive in nature, and therefore assumed a summative role. Evaluation conducted for the purpose of decision-making was proactive, and assumed a formative role. All four evaluation types--context, input, process and product--served a formative role when they provided information for program improvement and served a summative role when they provided information for decisions regarding the future of a program.

Applications of the CIPP Evaluation Model

The Phi Delta Kappa Committee's intention that the CIPP Model serve the needs of educational practitioners in their efforts to conduct evaluations has been realized in a variety of settings. The model has been applied to existing educational programs in schools and other agencies of education and training. The model has also been tested in several contexts to ascertain its usefulness in providing useful evaluation information for decision-making. Also, the model has been adapted by educators to meet the needs of specific program evaluations.

One context in which the CIPP Evaluation Model has been applied is the evaluation of programs in elementary and secondary schools. The model has been employed to evaluate a variety of educational programs. These applications have included evaluation of: innovative programs; curriculum content areas; learning strategies; and system and state-wide educational programs.

On the elementary school level, the CIPP Evaluation Model was employed to assess the merits of an educational innovation which had been implemented at the school-wide level (Pasch, 1976). The model was used as a framework for examining the use of paraprofessionals at the Schaff Junior High School in Parma, Ohio. The evaluation resulted in a series of recommendations to improve the training system that assisted instructional aides and their faculty supervisors. In this instance, the CIPP

Evaluation Model was employed to describe an innovative program in analytical and generalizable terms.

Another application of the CIPP Evaluation Model to study the effectiveness of a school-wide program was conducted in the area of career education (Stead, 1977). The CIPP Evaluation Model was used in a third party evaluation of the Appalachian Maryland Experienced Based Career Education Project (ECBE). The evaluation demonstrated that ECBE students showed academic and attitudinal progress as a result of the project. Here the CIPP Model was used to determine the interrelationship of procedural and intended elements of a program with program results.

The CIPP Evaluation Model has also been employed to evaluate the effectiveness of a particular mode of instruction (Bleakley, 1973). The model was employed to examine the relationship of rhetorical stance to the teaching of literature to twelfth grade students in a midwestern high school. In this instance, the CIPP Model was instrumental in teachers reaching the conclusion that the use of a particular teaching technique, rhetorical stance, had a positive effect on students cognitive recall of factual knowledge in a literature course.

In a study of the nature and types of impact evaluation available to vocational administrators, Grasso (1979) employed the CIPP Evaluation Model to delineate vocational educators' needs for evaluation information relating to federal legislation,

planning and accountability. This application of the CIPP Model demonstrated the differences between impact evaluation and research and provided vocational educators with a means for assessing the effectiveness of their current evaluation practices.

The CIPP Evaluation Model has been employed at the system-wide level to restructure the program evaluation plan of a city school system (Taylor, 1974). In Michigan, the Saginaw Public Schools' Evaluation Department was reorganized to facilitate conducting evaluations in the CIPP framework. In their experiences, administrators reported that the CIPP Evaluation Model was useful in providing information for answering 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? The usefulness of the CIPP Model in this application was that the reorganization resulting from adopting the CIPP framework facilitated providing information to administrators on the general quality of the educational programming of a school system.

A state-wide application of the CIPP Evaluation Model was employed in Colorado to assist in complying with Colorado's Educational Accountability Act (Ewy & Chase, 1977). The CIPP Model was used to develop information materials for implementing local educational improvement plans. These materials were used by teachers, administrators, and school board members. In this large scale application of the CIPP Model, the model provided

a framework for approaching evaluation as well as the basis for the numerous resources designed for the actual evaluations. This application of the CIPP Model also demonstrated its practical application by practitioners who were not formal evaluators.

In a study of eight state departments of education administration of ESEA Title I Programs, Forgione (1979) employed the CIPP Evaluation Model to describe the current status of program evaluations for Title I funded programs. This application of the CIPP Model delineated means employed in assessing effectiveness of funded programs, the relationship of program evaluation to other administrative practices, objectives against which to evaluate Title I programs and instructional processes within Title I programs.

The CIPP Evaluation Model has also been employed in several post-secondary school program evaluations. These applications have included program evaluation conducted on community college university and professional school programs, as well as an evaluation of a state-wide college program.

The usefulness of the CIPP Evaluation Model was appraised in the evaluation of an established career education program in a community college (Hecht, 1975). The focus of the evaluation study was to establish procedures for maximizing the impact of evaluation results on institutional decision-making. The results of the study demonstrated that the CIPP Evaluation Model could

be modified for evaluating a single program in a small community college. The authors cited the following characteristics of CIPP as particular advantages of using the model: comprehensiveness; flexibility; integration; and decision orientation.

The CIPP Evaluation Model was applied at the university level to evaluate the effectiveness of a comprehensive undergraduate inservice model for the preparation of elementary school teachers (Summary of the Educational Specifications for a Comprehensive Elementary Teacher Education Program, 1968). This inservice model was developed to accomodate the forces of societal and educational change. The evaluation component consisted of the four CIPP Evaluation types. In this instance, the CIPP Model was useful in evaluating a program designed for both individual and group study.

Other applications of the CIPP Evaluation Model at the university level concerned determining the effectiveness of innovative programs. The conclusions of an evaluation concerning a graduate program to train educational research and development personnel (Woodward & Yaeger, 1972), and a second study to evaluate a dental training program involving the use of paraprofessionals (Reeves & Michael, 1973) were similar. The CIPP Model was found to be a viable and useful evaluation methodology for providing information necessary for educational decision-making.

A state-wide post secondary school application of the CIPP Evaluation Model was conducted by the Continuing Education Division of Pennsylvania State University (Barnette, 1977). An evaluation conducted of the Pennsylvania Adult Basic Educational Improvement Program demonstrated that the CIPP Model was consistent with and supportive of the Organizational Development framework upon which the Pennsylvania Adult Basic Educational Improvement Program was based.

The CIPP Evaluation Model was also employed in the evaluation of programs in educational related institutions. The model was used to formulate a plan for evaluating the library services for the State of Illinois (Michael, 1976). Under a grant from the Illinois State Library Association, a manual explaining the CIPP Model was prepared as a reference guide for eighteen library systems to use in their five year planning and evaluation of programs. Several library systems reported that the CIPP Model was especially helpful in the analysis of their programs.

Merkel (1979) field-tested the CIPP Evaluation Model with the Community Education Program in Lakewood, New Jersey. This application of the CIPP Model described the effectiveness of programs which concern the well-being of all citizens in a community. The model provided a framework for assessing long-term programs with a particular emphasis on services for the elderly and minorities.

In addition to the applications of the CIPP Evaluation Model discussed above, there have been several theoretical applications of the model designed to assess the model's effectiveness. These applications have involved an application of the entire model as well as applications of selected pertinent elements of the model.

In a dissertation entitled "The Conceptualization of the Stufflebeam CIPP Evaluation Model in a Multivariate Context" (Hinkles, 1971), Hinkles applied the model to a hypothetical and complex educational change activity. Hinkles found that the inherent thoroughness of the model required a painstaking and time consuming process. He concluded, however, that utilization of the model as intended for designing and implementing an evaluation strategy would result in information that was valid, reliable, timely, pervasive, and credible.

The usefulness of the CIPP Evaluation Model in the evaluation of innovative programs was analyzed in a study which applied input evaluation to develop alternative evaluation strategies (Reinhard, 1973). The author found that the CIPP Model was a useful framework for extending the theory to the particular task of evaluating innovative programs.

Another theoretical application of the CIPP Evaluation Model explored its relevance in evaluating programs in the social realm (McLure, 1973). In this instance, the CIPP Eval-

uation Model was used to identify areas where institutionalized sex-role stereotyping would occur. This study demonstrated the model's usefulness in the identification of need for social change.

Finally, in a study conducted to assess the evaluation priorities of students, teachers, and principals, Nevo and Stufflebeam (1975) found that while information on outcomes is the most available evaluation information in schools, school people showed a great concern for other kinds of evaluation information. In this instance, the CIPP Evaluation Model demonstrated a particular sensitivity to evaluation attitudes and needs.

The final context in which the CIPP Evaluation Model has been applied has been as the basis for the formulation of a new evaluation model. In these instances, the CIPP Model was adapted to meet the unique needs of a particular program evaluation. These applications have included: an adaptation of the CIPP Model for use in adult education program evaluation; modifying the CIPP Model to assess school staff and community involvement in the evaluation of local educational programs; and the formulation of an eclectic model developed for a district-wide program evaluation.

In a dissertation entitled "An Adaptation of the CIPP Model of Evaluation For Use In Adult Education," Shiplett (1974) devised an evaluation model for practitioners in adult

education. The new model was designed to serve in the planning and implementation of adult education programs. Shiplett cited the CIPP Model for providing the new model's conceptualization system and the feedback network for maintaining communication at all levels of evaluation.

Nolin (1976) developed a condensed evaluation model to meet public and professional aspirations for involvement in the decision-making process in the planning and evaluation of educational programs. The condensed model increased the extent of these groups' involvement in the planning and decision-making process with the intent of improving the quality of decision-making. The new model retained the CIPP Model's definition and rationale for decision-making and the guidelines CIPP offered for involving various groups in the process of decision-making.

The city school district of New Rochelle, New York devised an evaluation model for collecting, storing and displaying pertinent data for use in planning educational programs at the district and school level (Gess, 1974). The model was used to evaluate district-wide reading programs and programs for exceptional children. The CIPP Model was an integral aspect of the new model in several respects: providing a rationale for analyzing program goals; developing a plan for gathering relevant data; and integrating the evaluation process into the study

of district curriculum.

The numerous and varied applications of the CIPP Evaluation Model have demonstrated its flexibility, comprehensiveness and practical orientation. These applications have revealed that the CIPP Model is a logical structure for an evaluation design to examine programs in a variety of educational contexts. The general consensus of authors who have reported on their use of the model has been that its application has resulted in obtaining valid and reliable information for decision-making (Gess, 1974).

CHAPTER III

RESEARCH DESIGN

Introduction

This study compared the responses of two groups of educational practitioners, classroom teachers and curriculum supervisors, to an instrument designed to describe their perceptions of the four types of evaluation identified by the CIPP Evaluation Model: context; input; process; and product. The study was conducted in several Chicago metropolitan area elementary school districts.

Participants

Two groups of educational practitioners were identified for this study, classroom teachers and curriculum supervisors. Classroom teachers were defined as teachers in grades kindergarten through sixth grade who are assigned to full-time classroom instruction. Curriculum supervisors were defined as educators whose job descriptions included supervision or assistance to classroom teachers in their implementation of curriculum. Such positions included elementary school principals, assistant superintendents for curriculum and supervisors, consultants, coordinators and directors of curriculum.

Classroom teachers and curriculum supervisors from sev-

eral Chicago metropolitan area elementary school districts were selected to participate in this study. These school districts were identified through consideration of the following criteria:

1. The school district was within a thirty mile radius of the downtown business area of Chicago. Excluding, however, the Chicago Public School System which is atypical of most school systems due to its size and the manner in which it evaluates educational programs.

2. The school district employed a minimum of two-hundred certified elementary teachers. A district of this size was necessary to insure an adequate number of curriculum supervisors for study.

3. The school district completed a formal program evaluation in reading within the last three years for the purpose of adopting a new basal reading program. For the purpose of this study, a formal program evaluation in reading was defined as a process by which a committee specifically charged with the task of studying different basal reading programs made a recommendation for adoption of a basal reader based upon its merits in terms of a predetermined set of criteria.

4. The program evaluation committee in reading was predominantly comprised of classroom teachers who would implement the new basal reading program.

The Directory of Illinois Schools (1979) was consulted

in the identification of elementary school districts. Sixty elementary school districts were included within the thirty mile radius of downtown Chicago, and twenty of these school districts employed a certified teaching staff of at least two-hundred.

The assistant superintendent for curriculum in each district was contacted by telephone and was asked the following questions:

1. Has your school district completed a program evaluation in reading within the last three years for the purpose of adopting a new basal reading program?

2. Did your school district engage in a formal process of evaluation--that is, did a committee consider alternative basals in terms of predetermined objectives?

3. Did your school district's evaluation committee recommend a basal reading program for adoption in a written report of its decision?

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

Four school districts responded positively to these questions. Curriculum supervisors from each of the four school districts agreed to participate in this study.

In conducting this study, the assistant superintendent for curriculum in each district served as the primary resource

person for his district. Through this resource person, the following participants were identified:

1. All curriculum supervisors within the school district who comprised the curriculum supervisor group. This included elementary school principals, assistant superintendents for curriculum, and curriculum supervisors, consultants, coordinators and directors.

2. All members of the program evaluation committee in reading were included in the classroom teacher group.

3. A sample of all kindergarten through sixth grade elementary school teachers within the school district was also included in the classroom teacher group.

Data Collection

A preliminary meeting was held with the assistant superintendent for curriculum in each district. The purpose of these meetings was to provide each assistant superintendent for curriculum with an overview of the study, review data collection logistics, and to secure permission to conduct the study in each district. The following were discussed: (a) purpose of the study, (b) how each school district was selected to participate in the study, and (c) an explanation of each participating school district's role in the study. A research proposal outlining the above was presented to each assistant superintendent

for curriculum prior to this preliminary metting. (See Appendix A for the research proposal presented to participating school districts.)

The assistant superintendent for curriculum was asked to provide the following information: (a) the names and base schools of all kindergarten through sixth grade teachers; (b) the names and base schools of all members of the reading evaluation committee; and (c) the names of personnel satisfying the criteria for the curriculum supervisor group. It was explained that the classroom teacher research sample for the study would include all members of the reading evaluation committee who were classroom teachers and a randomly selected group of forty classroom teachers. The curriculum supervisor group would include all supervisors identified by the assistant superintendent for curriculum.

The research instruments were distributed via each school district's inter-office mail system. Subjects were instructed to return the research instruments to the author via U.S. Mail in the self-addressed stamped envelope attached to the research instruments. These instruments were coded with a four digit number which identified the participants' school districts and groups.

Classroom Teacher Participants

Classroom teacher participants from each of the four school districts were selected as follows. This group included all members of the reading evaluation committee who were classroom teachers. In addition, a random sample of forty other kindergarten through sixth grade classroom teachers were selected for this group. This was accomplished by dividing the total number of kindergarten through sixth grade classroom teachers by forty and determining an identifier "X." Then, by counting down an alphabetical listing of kindergarten through sixth grade faculty arranged by the school, every "Xth" teacher was selected to participate in the study.

Curriculum Supervisor Participants

Curriculum supervisory personnel were identified by the assistent superintendent for curriculum in each of the four school districts. Personnel who satisfied the criteria for the curriculum supervisor group were included in the study.

The Research Instrument

Classroom teachers and curriculum supervisors were asked to assess the relative importance of the four types of evaluation identified by the CIPP Evalution Model and to assess the relative importance of a series of evaluation tasks contained in each

type of evaluation. Data were collected by means of a research instrument which participants completed individually.

Background of the Research Instument

The CIPP Evaluation Model was developed by the Phi Delta Kappa Committee on Evaluation. As described earlier, the CIPP Evaluation Model was selected as the referent from which to describe the perceptions of educational practitioners concerning the process of program evaluation for several reasons. The model was derived from a comprehensive theory of evaluation; it describes evaluation as a formal systematic process; it is a model developed by educational practitioners, for practitioners; and, its administrative orientation is adaptable to the current program evaluation practices of school systems.

This study sought to describe classroom teachers' and curriculum supervisors' perceptions of the process of program evaluation. These educators were asked to identify the relative importance of the four types of evaluation delineated by the CIPP Evaluation Model--context, input, process and product. According to the CIPP Evaluation Model, these four types of evaluation comprise the larger task of program evaluation. In addition, these educators were asked to rank order five evaluation tasks representative of each of these four types of evaluation in terms of their relative importance. (See Appendix

A pp. 144-148 for this portion of the research instrument.)

In addition to describing classroom teachers' and curriculum supervisors' perceptions of the program evaluation process, the research instrument also sought to describe these educators in terms of several common variables. The work of Good et al. (1975) and Talmage (1977) identified the need for systematic involvement of practitioners in evaluation research. If practitioners were to participate in evaluation research, a need existed to provide descriptive information concerning these participants. In previous studies which focused on practitioners' perceptions of the program evaluation process (Nevo & Stufflebeam, 1975; Nolin, 1976), and other studies which demonstrated the value of practitioners' participation in evaluation research (Cohen, 1975; Novak, 1977; Talmage, 1977), such descriptive information was lacking. This study described the practitioners who participated in the study in terms of several common variables which could be expected to influence their perceptions of the program evaluation process. These variables included: position, years of experience, highest level of education, major area of graduate study, and the experience of having served on an inservice program evaluation committee.

Format of the Instrument

The research instrument identified the four types of eval-

uation delineated in the CIPP Evaluation Model. Each type of evaluation was defined by its purpose as follows:

Context Evaluation: to provide a rationale for determining program objectives.

Input Evaluation: to provide information for determining how to utilize resources to meet program goals.

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

Product Evaluation: to measure and interpret attainments during the implementation and duration of the program.

Both groups were asked to:

> Assume that the process of educational program evaluation can be represented by 100 points. Assuming that each point represents an equal measure of value, or importance, divide the 100 points among the four types of evaluation according to your perception of the relative importance of each type.

The five tasks identified for each type of evaluation were as follows:

Context Evaluation

1. Identify learning outcomes of current program and identify learning outcomes as desired from a new program.

2. Identify needs not being served by the current program.

3. Identify potential human and material resources available to implement a new program.

4. Gather information from sources outside the school district such as research findings or outside consultants.

5. Explore other available programs in terms of the impact of change on students, faculty, parents and community.

Input Evaluation

1. Determine what is already being done to meet a new set of objectives.

2. Identify potential costs and benefits of other available programs.

3. Determine how existing staff, facilities and resources can be utilized to implement the new program.

4. Determine a specific schedule of events and activities to guide the new program's implementation.

5. Determine how the new program should be administered, evaluated and reviewed at various levels.

Process Evaluation

1. Develop an implementation plan for the new program.

2. Determine the adequacy of resources, facilities, staff and time schedules during implementation of the new program.

3. Determine the kinds of feedback needed during piloting and implementation.

4. Monitor the various publics' (teachers, students, administrators, parents) understanding of and agreement with the new program.

5. Design and assess communication channels between teachers, consultants, and administrators.

Product Evaluation

1. Determine whether or not the program is achieving its objectives.

2. Assess gains (or losses) in pupil achievement.

3. Assess the attitudes of students, staff, parents and community regarding the outcomes of the program.

4. Identify unanticipated outcomes and their effects on students and faculty.

5. Determine whether or not the program results justify the finances and efforts needed to maintain it.

Both groups were asked to:

> Within each type of evaluation, rank the five tasks according to your perception of their relative importance. Rank the most important task as "1" and the remaining tasks as "2" through "5" accordingly.

The descriptive variables included: position, years of experience, highest level of education, major area of graduate study, and the experience of participating in an inservice program evaluation. The variables identified for classroom teachers were:

1. Years of full time teaching experience.

2. Highest level of education.

3. Major area of graduate study if participant holds a graduate degree in education.

4. Participation in an inservice program evaluation committee.

The variables identified for curriculum supervisors were:

1. Present position.
2. Years of experience in present position.
3. Highest level of education.
4. Major area of graduate study.

(See Appendix A pp. 145-146 for this portion of the research instrument.)

Scoring the Research Instrument

In order to determine the difference in participants' ranking of the five representative evaluation tasks between types of evaluation, the following technique was employed. Participants' ranking of tasks from one through five were reversed and multiplied by the number of value points assigned to the type of evaluation each task represents. The total possible value points based on the original devision of one-hundred points across the four types of evaluation became one-thousand-five-hundred.

Instrument Validity

Construct validity of the research instrument was insured in that the four types of evaluation, their definition of purpose and the tasks included under each type of evaluation, were derived from the text Educational Evaluation and Decision-Making (Stufflebeam et al., 1971). This text served as the report of

the Phi Delta Kappa Committee on Evaluation which presents the CIPP Evaluation Model. Several Loyola University faculty members with expertise in the field of evaluation reviewed the instrument for accuracy in its interpretation of the CIPP Evaluation Model.

Instrument validity was also established by several other means. First, the instrument was reviewed by a group of graduate students who were writing doctoral dissertations in curriculum and instruction at Loyola University. Second, the instrument was administered to a Loyola University graduate class in elementary school curriculum. Finally, the revised research instrument was then field tested by administering it to an elementary school faculty and several curriculum supervisors. (See Appendix C p. 159 for the results of this field testing.)

The initial review of the research instrument by the curriculum and instruction doctoral students was conducted for the purpose of allowing a group of individuals with expertise in curriculum evaluation to critique the content and design of the research instrument. On the basis of this group's suggestions, and the suggestions of the faculty members who reviewed the research instrument, several modifications were made. These modifications included revising several categories within the educator variables so as to more precisely describe the two groups of educational practitioners and rewording several of

the evaluation tasks so as to achieve clarity of task description. (See Appendix B p.149 for initial and revised research instrument and a background summary of the curriculum and instruction doctoral student group.)

The revised research instrument was then administered to the graduate class in elementary education. This class included elementary school teachers and curriculum supervisors who were asked to respond to the instrument to ascertain the amount of time necessary to complete the instrument and the clarity of the instructions. This group experienced some difficulty in assigning value points to the individual tasks comprising each of the four stages of evaluation. They also did not perceive task 2 under product evaluation to be distinct from task 1 under the same stage. The task of completing the research instrument required approximately one-half hour.

This group's experience resulted in several more modifications to the research instrument. First of all, the directions were revised so that a simple ranking of the tasks in terms of importance was substituted for the assigning of value points. Task 2 under product evaluation was deleted and a different product evaluation task was included in the instrument. The order of tasks within process evaluation was also altered. Finally, in order to clarify that the tasks identified in the research instrument were representative of the four types of

evaluation and not intended to represent a sequence of evaluation tasks, context, input, process and product evaluation were hereafter presented as "types of evaluation" rather than "stages of evaluation."

The final version of the research instrument was administered to an elementary school faculty and several curriculum supervisors from one of the four school districts who participated in the study to ascertain the time required to complete it, and the clarity of the instructions. This group was able to respond to the research instrument without additional clarifications in approximately fifteen minutes. (See Appendix C p.160 for a summary of this group's background and their responses to the research instrument.)

Instrument Reliability

Instrument reliability was established in the following manner. The final version of the research instrument was administered a second time, one week later, to the same elementary school faculty and curriculum supervisors identified above. An analysis of this group's responses to the first and second administration of the research instrument showed minimal discrepancies between the results of each administration. There was no difference in the ranking of the relative importance of the four types of evaluation. There was high agreement (Spearman

rank correlation = .768) between the ranking of the relative importance of the twenty evaluation tasks in each administration of the research instrument. (See Appendix C pp. 162-163 for a summary of reliability data.)

Hypotheses to be Tested

The following null hypotheses were tested:

1. There is no difference in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the four types of evaluation and the mean rankings of evaluation tasks representative of each type of evaluation.

2. There is no difference in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the twenty weighted evaluation tasks.

3. There is no difference in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the four types of evaluation and evaluation tasks representative of each type of evaluation between school districts.

4. Classroom teachers' and curriculum supervisors' years of experience in their present positions is not significantly related to their perceptions of educational program evaluation.

5. Classroom teachers' and curriculum supervisors' highest level of education is not significantly related to their perceptions of educational program evaluation.

6. Classroom teachers' and curriculum supervisors' major area of graduate study is not significantly related to their perceptions of educational program evaluation.

The level of significance set for testing each of the hypotheses was .05. Results significant at the .01 level were also identified. All data analyses were conducted using the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner & Brent, 1975) using an IBM 370 installation of version M, Release 8.1.

CHAPTER IV

RESULTS OF THE STUDY

Return of Research Instruments

Research instruments were distributed to 160 classroom teachers, 60 reading program evaluation committee members, and 107 curriculum supervisors. Of these 327 instruments, 255 (78 percent) were returned. Nine instruments from the 255 were completed incorrectly and these were discarded. Thus, 246 instruments were analyzed in this study.

Population

Five groups of educational practitioners were included in this study. Three groups of classroom teachers and two groups of curriculum supervisors comprised the research population. Group one included all classroom teachers without prior experience in an educational program evaluation. Group two included all classroom teachers who reported some educational program evaluation experience other than the reading program evaluation discussed above. Group three included classroom teachers who participated in their school districts' reading program evaluation. These three classroom teacher groups were independent groups. Group four included all elementary school principals. Group five consisted of all curriculum personnel with such titles as

Table 1

Number of Cases for Groups

	Classroom Teachers			Curriculum Supervisors		
District	Teachers Group 1	Teachers Eval. Exp. Group 2	Teachers Eval. Com. Group 3	Principals Group 4	Curriculum Personnel Group 5	District Totals
A	16	11	10	12	22	71
	6.5%[a]	4.5%	4.1%	4.9%	8.9%	28.9%
B	16	15	16	8	8	54
	6.5%	6.1%	6.5%	3.3%	3.3%	22.0%
C	12	12	12	10	8	63
	4.9%	4.9%	4.9%	4.1%	3.3%	25.6%
D	19	12	11	6	10	58
	7.7%	4.9%	4.5%	2.4%	4.1%	23.6%
Group Totals	63	50	49	36	48	246
	25.6%	20.4%	20.0%	14.7%	19.5%	100%
		162		84		
		65.9%		34.1%		

Note. The rate of return of instruments by participants was as follows:

	Number Distributed	Number Returned	Percent of Return
Classroom Teachers	220	169	76.8%
Curriculum Supervisors	107	86	80.3%
Total	327	255	77.9%

Nine instruments from the 255 returned were completed incorrectly, seven from the classroom teacher groups and two from the curriculum supervisors. These were discarded.

[a]Percent of total N

supervisor, director, coordinator, consultant, etc.. The number of respondents comprising each of the four school districts' groups and their percentage of the total population are listed in Table 1.

Demographic Information

Several types of demographic information were gathered in this study. Included were years of experience in present position, highest level of education and major areas of graduate study were compiled for each of the groups.

Years of Experience in Present Position

The group profiles for years of experience in present position are listed in Table 2. These data illustrate that approximately one-half of the classroom teachers had more than 10 years experience and that the vast majority of these classroom teachers had 6 or more years of experience. The majority of curriculum supervisors reported 3-10 years of experience in their present positions. Curriculum supervisors included a greater percentage of respondents in the 1-5 years of experience range compared with the classroom teacher groups, and, a lower percentage of respondents in the more than 10 years category compared with the classroom teacher groups. (See Appendix D p.165 for a group profile for years of experience in present position for each of the four school districts.)

Table 2

Years of Experience in Present Position

Position	1-2 Years	3-5 Years	6-10 Years	>10 Years	
Teacher	2	6	31	24	
Teacher Eval. Exp.	1	5	19	25	
Teacher Eval. Comm.	1	12	13	23	
Principal	7	9	13	7	
Curriculum Personnel	8	17	14	9	
Experience Totals	19	49	90	88	N=246

Highest Level of Education

The group profiles for highest level of education are listed in Table 3. The data illustrate that the majority of classroom teachers held masters or higher level degrees. All elementary school principals and most curriculum personnel held masters or higher level degrees. (See Appendix D p. 166 for a group profile for highest level of education for each of the four school districts.)

Major Areas of Graduate Study

The group profiles for major areas of graduate study are reported in Table 4. These data illustrate that almost one-half of the educational practitioners sampled in this study held graduate degrees with a major concentration in the curriculum related areas of curriculum and instruction, education generalist, or subject area specialist. Approximately 25 percent of the practitioners in this study reported a graduate degree with a major concentration in administration and supervision.

The vast majority of the classroom teacher groups who reported a graduate major held degrees in the areas directly related to instruction rather than administration and supervision. In the curriculum supervisor groups, however, all but one respondent in the elementary school principal group and approximately 25 percent of the curriculum personnel group

Table 3

Highest Level of Education

Position	Bachelors Degree	Masters Degree	Certif. of Adv. Study	Doctorate	
Teacher	36	26	1	0	
Teacher Eval. Exp.	24	23	2	1	
Teacher Eval. Comm.	10	33	4	2	
Principal	0	28	6	2	
Curriculum Personnel	9	26	11	2	
Education Totals	79	136	24	7	N=246

Table 4

Major Areas of Graduate Study

Position	None[a]	Adm. or Suprv.	Curriculum/ Instruction	General	Subject Specialist	Special Education	Other
Major 1							
Teacher	36	4	4	15	3	1	0
Teacher Eval Exp.	26	4	12	5	3	0	0
Teacher Eval. Comm.	11	5	9	8	13	3	0
Principal	0	35	0	1	0	0	0
Curriculum Personnel	9	11	12	3	10	3	0
Major 1 Totals	83	58	37	32	29	7	0
Major 2							
Teacher	62	0	0	0	1	0	0
Teacher Eval. Exp.	44	0	2	1	1	1	1
Teacher Eval. Comm.	40	0	2	0	3	4	0
Principal	27	0	1	6	1	1	0
Curriculum Personnel	33	2	2	2	6	1	2
Major 2 Totals	206	2	7	9	12	7	3

[a]Due to Subjects' holding only Bachelor Degrees

reported major study in administration and supervision. Twenty-five percent of the elementary school principals and approximately 85 percent of the curriculum personnel group reported major study in curriculum related areas.

Research Findings

Classroom teachers and curriculum supervisors were asked to divide 100 value points among the four types of evaluation described by the CIPP Evaluation Model according to their perceptions of the relative importance of each type. These groups were also asked to rank five tasks representative of each type of evaluation according to their perceptions of the relative importance of the tasks within each of the four evaluation types. A summary of the results are reported in Tables 5, 6 and 7. The means reported for context, input, process and product evaluation represent educators' assessment of the importance of each type of evaluation described by the instrument in relationship to the other three evaluation types. The means reported for the evaluation tasks represent educators' assessment of the importance of each evaluation task in relationship to the other tasks within that particular type of evaluation.

Table 5 provides measures of central tendency for the classroom teacher and curriculum supervisor groups. These findings are further broken down in Table 6 to report the results

Table 5

Population Measures of Central Tendency

Type of Evaluation		Classroom Teachers Mean	Classroom Teachers Standard Deviation	Curriculum Supervisors Mean	Curriculum Supervisors Standard Deviation
Context		27.71	9.06	27.17	7.94
Input		24.79	6.59	22.25	5.50
Process		22.72	6.02	23.55	6.55
Product		24.69	8.05	27.27	8.64
Evaluation Tasks					
CONTEXT	1	4.33	.83	4.23	.80
	2	4.22	.92	4.25	.94
	3	2.40	1.08	2.33	1.02
	4	1.85	.96	1.92	1.03
	5	2.22	.94	2.26	1.08
INPUT	6	3.88	1.35	3.93	1.31
	7	2.28	1.38	2.00	1.26
	8	3.59	1.12	3.81	1.04
	9	2.43	1.22	2.36	1.10
	10	2.83	1.26	2.90	1.25
PROCESS	11	3.79	1.28	3.61	1.39
	12	3.00	1.32	3.17	1.26
	13	3.38	1.28	3.49	1.19
	14	1.87	1.18	1.98	1.25
	15	2.96	1.26	2.74	1.34

Evaluation Tasks		Mean	Standard Deviation	Mean	Standard Deviation
PRODUCT	16	4.13	1.15	4.04	1.28
	17	3.86	1.01	3.92	1.00
	18	2.91	1.21	3.08	.96
	19	2.37	1.03	2.26	1.05
	20	1.74	1.07	1.70	1.13

Table 6

Classroom Teacher Group Measures of Central Tendency

		Group 1		Group 2		Group 3	
Type of Evaluation		Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Context		25.94	8.40	29.14	11.70	28.53	6.10
Input		24.91	6.69	23.74	6.98	25.71	6.00
Process		23.57	5.64	21.94	7.03	22.61	5.34
Product		25.68	8.26	24.96	9.80	23.14	5.21
Evaluation Tasks							
CONTEXT	1	4.38	.91	4.32	.84	4.26	.73
	2	4.16	.77	4.18	.98	4.35	1.03
	3	2.56	1.13	2.38	1.01	2.20	1.06
	4	1.65	.83	1.84	1.08	2.10	.94
	5	2.29	.99	2.28	.93	2.08	.89
INPUT	6	4.05	1.29	3.84	1.30	3.69	1.46
	7	2.38	1.43	2.36	1.37	2.06	1.34
	8	3.54	1.03	3.64	1.29	3.59	1.08
	9	2.22	1.08	2.38	1.34	2.76	1.20
	10	2.81	1.33	2.78	1.09	2.90	1.36
PRODUCT	11	3.57	1.29	3.82	1.32	4.04	1.20
	12	2.89	1.21	3.02	1.39	3.12	1.39
	13	3.67	1.24	3.28	1.44	3.10	1.10
	14	1.73	1.18	2.10	1.22	1.82	1.11
	15	3.14	1.28	2.78	1.15	2.92	1.34

Evaluation Tasks		Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
PRODUCT	16	4.16	1.12	4.14	1.26	4.08	1.10
	17	3.81	1.13	3.90	1.00	3.90	.85
	18	2.86	1.28	2.78	1.18	3.10	1.14
	19	2.24	.98	2.36	.90	2.55	1.21
	20	1.94	1.10	1.86	1.21	1.37	.76

Table 7

Curriculum Supervisor Group Measures of Central Tendency

		Group 4		Group 5	
Type of Evaluation		Mean	Standard Deviation	Mean	Standard Deviation
Context		29.17	8.24	25.67	7.44
Input		23.33	6.21	21.44	4.81
Process		22.08	5.53	24.65	7.08
Product		25.69	7.48	28.46	9.32
Evaluation Tasks					
CONTEXT	1	4.22	.87	4.25	.76
	2	4.19	.89	4.29	.99
	3	2.28	.97	2.38	1.06
	4	2.11	1.16	1.77	.90
	5	2.19	1.19	2.31	.99
INPUT	6	3.89	1.37	3.96	1.27
	7	1.97	1.32	3.96	1.27
	8	3.67	1.07	3.92	1.01
	9	2.44	1.13	2.29	1.09
	10	3.03	1.25	2.81	1.25
PROCESS	11	3.44	1.59	3.73	1.22
	12	3.14	1.31	3.19	1.23
	13	3.33	1.24	3.60	1.14
	14	2.42	1.46	1.65	.96
	15	2.67	1.24	2.79	1.41

	Evaluation Tasks	Mean	Standard Deviation	Mean	Standard Deviation
PRODUCT	16	3.86	1.31	4.17	1.24
	17	3.92	1.08	3.92	.94
	18	3.33	.93	2.90	.95
	19	2.19	.98	2.31	1.11
	20	1.69	1.19	1.71	1.09

for the three classroom teacher groups and in Table 7 to report the results for the two curriculum supervisor groups.

Hypothesis I

The first hypothesis of the study stated that there is no difference in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the four types of evaluation and the mean rankings of evaluation tasks representative of each type of evaluation. Two-tailed t-tests were performed to compare responses for evaluation types and tasks between the classroom teachers and curriculum supervisor groups. Significant findings are reported in Table 8. (See Appendix E p. 171 for table listing t-values for all group comparisons.) Significant results were obtained for input evaluation $t = 3.03$ $p < .01$ and product evaluation $t = -2.33$ $p < .05$. There were no significant results for task ranking comparisons.

Classroom teacher group comparisons were also conducted. Although no significant findings were obtained in comparing evaluation types, several task comparisons were significant. In the comparisons of groups 1 and 3, significant results were found for task 4 $t = -2.70$ $p < .01$, task 13 $t = 2.50$ $p < .05$, and task 20 $t = 3.09$ $p < .01$. One task comparison was significant for groups 2 and 3, task 20 $t = 2.42$ $p < .05$.

Curriculum supervisor group comparisons indicated several

Table 8

Significant T-Test Results for Position Comparisons

	Teacher Groups			Supervisor Groups	Teachers & Supervisors
Type of[a] Evaluation	1,2	1,3	2,3	4,5	1+2+3,4+5
Context				*a	
Input					**
Product					*
Evaluation Tasks					
4		**			
13		*			
14				**	
18				*	
20		**	*		

[a]Types of evaluation and evaluation tasks omitted showed no significant differences.

*$p < .05$

**$p < .01$

significant results. Comparisons for context evaluation were significant $\underline{t} = 2.02$ $\underline{p} < .05$ as were comparisons for task 14 $\underline{t} = 2.92$ $\underline{p} < .01$ and task 18 $\underline{t} = 2.11$ $\underline{p} < .05$.

The findings from these comparisons indicated that classroom teachers perceived input evaluation to be more important than did curriculum supervisors and that curriculum supervisors perceived product evaluation to be more important than did classroom teachers. Although these two classifications did not differ in their ranking of tasks within each type of evaluation, individual group breakdowns did yield several significant findings.

Classroom teachers who served on the reading program evaluation committee perceived task 20 to be of less importance than did the other two classroom teacher groups. The remaining significant comparisons were between the classroom teacher groups with no evaluation experience and evaluation committee experience.

Elementary school principals perceived context evaluation and evaluation tasks 14 and 18 to be more important than did curriculum personnel. Overall, the significant task ranking comparisons for these groups, as well as the classroom teacher groups, were not concentrated in any of the four types of evaluation. No significant task ranking comparisons were found for input evaluation.

This first hypothesis of the study stated that there would be no differences in classroom teachers' and curriculum super-

visors' perceptions of program evaluation. Since several significant differences in these educators' perceptions of program evaluation were identified in the study, Hypothesis I was rejected.

Hypothesis II

The second hypothesis of the study stated that there is no difference in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the twenty weighted evaluation tasks. As a point of clarification, the previous hypothesis concerned differences in perceptions of the relative importance of the four types of evaluation and mean rankings within each type of evaluation. This second hypothesis examined comparisons of task rankings in terms of the importance associated with the type of evaluation which they represented. Task ranking were weighted by multiplying the individual task ranks by the value assigned to the type of evaluation in which it was found. Thus, the weighted values represented the value associated with each evaluation task within the total sphere of evaluation. For example, if a subject assigned 25 points to context evaluation and ranked task 1 as "2" (indicating he perceived task 1 to be the second most important task of the 5 context evaluation tasks identified by the research instrument), the weighted value of task 1 became 2 X 25, or 50 value points. These values were

used in all subsequent comparisons. The values are listed for the classroom teacher and curriculum supervisor groups in Table 19 and for the group breakdowns in Tables 20 and 21. (See Appendix E pp. 168-170 for these tables.)

Two-tailed t-tests were performed to compare weighted responses for evaluation task rankings between the classroom teacher and curriculum supervisor groups. Significant results are reported in Table 9. (See Appendix E p. 171 for table listing t-values for all group comparisons.) Significant results were obtained for task 7 $t = 2.68$ $p < .01$, task 17 $t = -2.02$ $p < .05$ and task 18 $t = -2.30$ $p < .05$.

Further classroom teacher group comparisons yielded several significant findings. Comparisons for groups 1 and 2 were significant for task 4 $t = -2.02$ $p < .05$ and task 13 $t = 2.04$ $p < .05$. Several comparisons of groups 1 and 3 were significant. These were task 2 $t = 2.05$ $p < .05$, task 4 $t = -3.53$ $p < .01$, task 9 $t = -2.28$ $p < .05$, task 13 $t = 2.44$ $p < .05$ and task 20 $t = 2.32$ $p < .05$. One comparison between groups 2 and 3 was significant, task 20 $t = 2.32$ $p < .05$. Curriculum supervisor group comparisons also resulted in only one significant task comparison, task 4 $t = 2.42$ $p < .05$.

The findings from these comparisons indicate that classroom teachers perceived task 7 as more important and tasks 17 and 18 as less important than did curriculum supervisors. Classroom

Table 9

Significant T-Test Results for Position Comparisons of Weighted Tasks

	Teacher Groups			Supervisor Groups	Teachers & Supervisors
Evaluation Tasks[a]	1,2	1,3	2,3	4,5	1+2+3,4+5
2		*			
4	*	**		*	
7					**
9		*			
13	*	*			
17					*
18					*
20		**	**		

[a]Evaluation tasks omitted showed no significant differences.

*$p < .05$

**$p < .01$

teachers without program evaluation experience perceived task 4 to be less important and task 13 to be more important than did teachers with program evaluation experience. Classroom teachers who served on a reading program evaluation committee perceived task 20 to be of less importance than did the other teacher groups. Committee teachers also perceived tasks 2 and 9 to be more important than did teachers without program evaluation experience. Elementary school principals perceived task 4 to be more important than did curriculum personnel. Significant comparisons were evident for tasks representing each of the four types of evaluation.

The second hypothesis of the study stated that there would be no differences in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the twenty weighted evaluation tasks. Several significant differences were identified in the study. Hypothesis II, therefore, was rejected.

Hypothesis III

The third hypothesis of the study stated that there is no difference in classroom teachers' and curriculum supervisors' perceptions of the relative importance of the four types of evaluation and evaluation tasks representative of each type of evaluation between school districts. The previous hypotheses concerned educators' perceptions of program evaluation treating

like cross-district groups as individual populations. Hypothesis III examined group differences between school districts.

Hypothesis III was initially tested by means of a one-way analysis of variance. A mean was found for each evaluation type and task within districts and these means were compared between districts. Secondly, an a posteriori contrast test was performed to identify significant contrasts between district means. Scheffee's Test (Hays, 1973 p. 606) was selected for this purpose because it is appropriate for examining all possible linear combinations of group means. These significant contrasts are reported in Table 10. (See Appendix E p.174 for a listing of all F-Value comparisons.)

Significant contrasts were found between Districts A and B for context evaluation $F = 4.51$ $p < .01$, and for product evaluation $F = 5.64$ $p < .01$. Significant contrasts were also found between districts for context and product evaluation tasks. In each instance, the significant contrasts involved District A and one or more of the other districts. Significant contrasts were found for context evaluation between Districts A and C for task 3 $F = 3.49$ $p < .05$, and between Districts B and Districts A and C for task 5 $F = 4.73$ $p < .01$. Significant contrasts were also found for product evaluation between Districts A and B for task 17 $F = 5.98$ $p < .01$, Districts A and D for task 18 $F = 3.72$ $p < .05$, and Districts A and B, C for task 19 $F = 5.24$ $p < .01$.

Table 10

Significant Cross-District Comparisons from Scheffee Test

Type of Evaluation [a]	Significant Cross-District [b] Comparisons
Context	A-B [c]
Product	A-B
Evaluation Tasks	
3	A-C
5	A-B
17	A-B
18	A-D
19	A-B

[a] Types of evaluation and evaluation tasks omitted showed no significant differences.

[b] $p < .05$

[c] Letters identify school districts.

These results indicate that classroom teachers and curriculum supervisors from District A differed from District B in their perceptions of the relative importance of context and product evaluation. These educational practitioners from District A also differed with those from the other three districts in their perceptions of the relative importance of context and product evaluation tasks.

Hypothesis III was further tested by performing a one-way analysis of variance on the means of each groups' perceptions between districts. Scheffee's Test was also performed to identify significant contrasts between districts. The significant Scheffee contrasts are reported in Table 11.

Two significant cross-district group contrasts were found In group 5, District A differed for Districts B and D for product evaluation $\underline{F} = 7.08$ $\underline{p} < .01$. Two significant cross-district contrasts were found for evaluation tasks representing context, input and process evaluation. Significant cross-district contrasts were found for all of the product evaluation tasks. In all but one instance, these significant contrasts involved groups 3 and 5, teachers who served on the reading program evaluation committee and curriculum personnel. As with the initial testing of hypothesis III, most of the significant cross-district comparisons involved District A.

Table 11

Significant Cross-District Group Comparisons from Scheffee Test

Type of Evaluation[a]	Group	Significant Cross-District[b] Comparisons
Product	5	A-B, A-D[c]
Evaluation Tasks		
2	5	A-B
5	1	A-B
7	3	A-D
8	5	A-C
11	3	A-C
14	4	D-A,D-B,D-C
16	3	A-B
17	1	B-A,B-D
	5	A-B
18	3	A-D
	5	A-B
19	3	A-B,A-C
20	3	C-D

[a]Types of evaluation and evaluation tasks omitted showed no significant differences.

[b]$p < .05$

[c]Letters identify school districts.

These results indicate that curriculum personnel from District A differed from Districts B and D in their perceptions of the relative importance of product evaluation. In the analysis of evaluation tasks, teachers who served on the reading program evaluation committees differed more often between districts than did any other group. Once again, these differences were most evident in District A indicating that teachers from the program evaluation committee from District A differed most often with those of other school districts.

The third hypothesis of the study stated that there would be no differences in classroom teachers' and curriculum supervisors' perceptions of program evaluation across the four school districts. Since several significant differences were identified in the study, Hypothesis III was rejected.

The first three hypotheses examined differences in classroom teachers' and curriculum supervisors' perceptions of educational program evaluation between groups and school districts. The remaining hypotheses examined the relationship of years of experience, level of education and major areas of concentration in graduate study on these educational practitioners' perceptions of educational program evaluation. The purpose of these analyses was to examine the effect of these variables on educators' perception of educational program evaluation in general, as well as for the individual groups. Consequently, these analyses were

performed on the population as a whole as well as within and between classroom teacher and curriculum supervisor groups.

Hypothesis IV

Hypothesis IV stated that classroom teachers' and curriculum supervisors' years of experience in their present positions is not significantly related to their perceptions of educational program evaluation. Four categories of years of experience were identified in the study: 1-2 years, 3-5 years, 6-10 years and more than 10 years.

This hypothesis was tested in two ways. First of all, two-tailed t-tests were performed on each of the four evaluation types and tasks between categories of years of experience. Secondly, Pearson Product-Moment Correlation Coefficients (Pearson Correlations) were calculated between groups 1 through 5, groups 1 through 3 together and groups 4 and 5 together. Thus, inferences were able to be made for comparisons between experience categories as well as within and between teacher and curriculum supervisor groups.

Significant results for the t-test comparisons are reported in Table 12. (See Appendix E p. 176 for a table listing T-Values for all comparisons.) Three experience category comparisons yielded significant results for context evaluation: Comparison 1, $\underline{t} = 2.87$ $\underline{p} < .01$; Comparison 4 $\underline{t} = -2.83$ $\underline{p} < .01$;

Table 12

Significant T-Test Results for Years of Experience Comparisons

	Comparisons					
	1	2	3	4	5	6
Type of Evaluation [a]	1-2,3-5	1-2,6-10	1-2, >10	3-5,6-10	3-5, >10	6-10, >10
Context	**			**	**	
Input	*					
Evaluation Tasks						
1				*	**	
2	**			*		
3					*	
5					*	
18		*				
19					*	

[a]Types of evaluation and evaluation tasks omitted showed no significant differences.

*$p < .05$

**$p < .01$

and Comparison 5, t = -2.86 $p < .01$. One comparison was significant for input evaluation: Comparison 1, t = -1.99 $p < .05$. Significant task comparisons were found in tasks representing context and product evaluation. Significant context task comparisons were found for task 1, Comparison 4, t = -2.10 $p < .05$ and Comparison 5, t = -2.78 $p < .01$; for task 2 in Comparison 1, t = 2.65 $p < .01$ and Comparison 4, t = -2.34 $p < .05$; task 3 Comparison 5, t = -2.32 $p < .05$. Two product evaluation tasks produced significant comparisons: task 18 Comparison 3, t = 1.98 $p < .05$ and task 19, t = 2.19 $p < .05$.

One pattern is evident in these findings. Significant comparisons were found in each of the context evaluation comparisons involving the 3-5 years of experience category. This finding was reinforced by similar findings in the significant context evaluation task comparisons. Thus, it appears that educational practitioners' perceptions of educational program evaluation were different after several years of experience in their positions; but, that this difference was no longer evident after five years of experience.

The significant results of the Pearson Correlations are reported in Table 13. (See Appendix E p. 178 for a table listing Pearson Correlations for all experience category comparisons.) There were no significant correlations for evaluation types in any of these analyses. Perceptions concerning several evaluation

Table 13

Significant Pearson Correlations for Years of Experience

	All Groups	Teacher Groups			Supervisor Groups		All Teachers	All Supervisors
Evaluation Tasks	1,2,3,4,5	1	2	3	4	5	1,2,3	4,5
1	*							
3					**			
10						*		
11			*					
15				**				
18	**							
19				**			*	
20							*	

[a]Types of evaluation and evaluation tasks omitted showed no significant differences.

*$p < .05$

**$p < .01$

tasks, however, did correlate significantly with years of experience. The All Groups analysis resulted in significant correlations for task 1 $\underline{r}$ = 0.13 $\underline{p} < .05$ and for task 2 $\underline{r}$ = -0.14 $\underline{p} < .05$. The breakdown into classroom teacher and curriculum supervisor groups resulted in significant correlations for task 19 $\underline{r}$ = -0.19 $\underline{p} < .05$, and task 20 $\underline{r}$ = 0.16 $\underline{p} < .05$ for classroom teachers. There were no significant correlations for the curriculum supervisor group. Further group breakdowns produced several significant correlations: group 2, task 11 $\underline{r}$ = 0.35 $\underline{p} < .05$; group 3, task 15 $\underline{r}$ = -0.40 $\underline{p} < .01$; and task 19, $\underline{r}$ = -0.41 $\underline{p} < .01$; group 4, task 3 $\underline{r}$ = 0.44 $\underline{p} < .01$; and group 5, task 10 $\underline{r}$ = -0.32 $\underline{p} < .05$.

These findings indicate that no discernable pattern of correlation was evident between educational practitioners' perceptions of educational program evaluation and years of experience in their present positions. No significant correlations were found in the analysis of evaluation types and the few significant task correlations were scattered across evaluation types and population groups.

Overall, the findings for years of experience indicate that there is somewhat of a tendency for classroom teachers and curriculum supervisors to place more importance on context evaluation during the first two years in their present positions and again after five or more years in their positions. No other patterns were evident in a breakdown analysis of population groups.

The fourth hypothesis of the study stated that classroom teachers' and curriculum supervisors' perceptions of educational program evaluation are not related to these educators' years of experience in their present positions. As some significant comparisons and correlations were identified for years of experience in this study, Hypothesis IV was rejected.

Hypothesis V

Hypothesis V stated that classroom teachers' and curriculum supervisors' highest level of education is not significantly related to their perceptions of educational program evaluation. Four categories of educational levels were identified in this study: bachelors degree, masters, certificate of advanced study (CAS) and doctorate. Practitioners representing each of these levels of education were compared in the same manner as in Hypothesis IV using two-tailed T-tests and Pearson Correlations.

The significant T-test results are listed in Table 14. (See Appendix E p. 180 for a table listing T-values for all group comparisons.) There were no significant comparisons between highest level of education and any of the evaluation types. Several significant task mean comparisons were found. Task 4 comparisons were significant in Comparison 1 $t = -2.20$ $p < .05$, Comparison 2 $t = -2.56$ $p < .05$ and Comparison 3 $t = 1.99$ $p < .05$. Task 18 was significant in Comparison 6 $t = -2.06$ $p < .05$ and task 20 was significant in Comparison 3 $t = -2.27$ $p < .05$ and

Table 14

Significant T-Test Results for Highest Level of Education Comparisons

	Comparisons					
	1	2	3	4	5	6
Evaluation[a] Tasks	Bachelors, Masters	Bachelors, CAS	Bachelors, Doctorate	Masters, CAS	Masters, Doctorate	CAS, Doctorate
4	*	*	*			
18						*
20			*		**	

[a]Types of evaluation and evaluation tasks omitted showed no significant differences.

*$p < .05$

**$p < .01$

Comparison 5 $t = -2.72$ $p < .05$.

These findings indicate that although no pattern was evident in educational practitioners' highest level of education and their perceptions of the four types of educational program evaluation, perceptions concerning three evaluation tasks appeared to be influenced by level of education. Practitioners with bachelors degrees perceived task 4 to be of less importance than did practitioners with graduate degrees. Practitioners with doctorates perceived task 20 to be of less importance than did practitioners with bachelors and masters degrees and task 18 to be of less importance than practitioners with certificates of advanced study.

Significant results for the Pearson Correlations are reported in Table 15. (See Appendix E p. 182 for a table listing Pearson Correlations for all category comparisons.) One significant correlation for evaluation types was found when all classroom teacher groups were analyzed for context evaluation $r = 0.15$ $p < .05$. Significant correlations were found for three evaluation tasks in the All Groups analysis. In each instance, the significant correlation was evident in only one group breakdown: task 4, All Groups $r = 0.18$ $p < .01$, All Teachers $r = 0.20$ $p < .01$; task 6, All Groups $r = -0.13$ $p < .05$, curriculum personnel $r = -0.28$ $p < .05$; and task 7, All Groups $r = -0.28$ $p < .05$. Three more tasks also resulted in significant correlations in

Table 15

Significant Pearson Correlations for Highest Level Of Education

	All Groups	Teacher Groups			Supervisor Groups		All Teachers	All Supervisors
Type of[a] Evaluation	1+2+3+4+5	1	2	3	4	5	1+2+3	4+5
Context							*	
Evaluation Tasks								
4	**						**	
6	*					*		*
7	*	*						
11		*						
19			*					
20				*				

[a]Types of evaluation and evaluation tasks omitted showed no significant differences.

*$p < .05$

**$p < .01$

the group breakdown: task 11, teachers $\underline{r}$ = -0.25 $\underline{p} < .05$; task 19 $\underline{r}$ = -0.31 $\underline{p} < .05$; and task 20 $\underline{r}$ = -0.30 $\underline{p} < .05$.

These findings indicate that the correlation between classroom teachers highest level of education and their perceptions of context evaluation was reflected in only one evaluation task, task 4. It appeared that classroom teachers attributed more importance to context evaluation as their level of education increased and that this tendency was also evident in their perceptions of the importance of task 4.

Two significant All Groups correlations were found for input evaluation tasks. The tendency for all groups to place less importance on task 6 as level of education increased was also significant in the correlations for all supervisors and curriculum personnel. This same tendency was evident for task 7 when broken down into the teacher group.

The three remaining significant correlations were evident in process and product evaluation tasks in the classroom teacher groups. In each instance, the correlation indicated that classroom teachers place less value on these evaluation tasks as their level of education increased.

The results for highest level of education indicate that there is somewhat of a tendency for classroom teachers to place more importance on context evaluation as their level of education increases. No other patterns were evident in analyses of curric-

ulum supervisors or in the group breakdowns.

The fifth hypothesis of the study stated that classroom teachers' and curriculum supervisors' perceptions of educational program evaluation are not significantly related to these educators' highest level of education. Several significant comparisons and contrasts were identified for highest level of education in this study. Hypothesis V, therefore, was rejected.

Hypothesis VI

Hypothesis VI stated that classroom teachers' and curriculum supervisors' major area of graduate study is not significantly related to their perceptions of educational program evaluation. Six categories of major areas of graduate study were identified: none (indicating that participant did not hold a graduate degree), administration/supervision, curriculum/ instruction, general education, subject or content area specialist, and other. If a second major was indicated, this information was also included in the analysis.

Hypothesis VI was initially tested by means of a one-way analysis of variance across education majors using graduate major as the dependent measure. Secondly, an a posteriori contrast was performed to identify significant contrasts between major area of graduate study means. Scheffee's Test was selected for this purpose because it is appropriate for examining all possible

linear combinations of groups, even for unequal size groups. (See Appendix E p. 184 for a table listing F-values for these comparisons.)

No significant comparisons were found for graduate major areas of study in either the first or second major comparisons for types of evaluation. One task, task 4, yielded a significant comparison for both first and second major areas of graduate study comparisons: Major 1, $F = 2.79$ $p < .05$; Major 2, $F = 2.52$ $p < .05$. Scheffee's Test did not identify any significant contrasts between major areas of graduate study indicating that the significance of the F-values was due to the interaction effect.

These results indicate that major area of graduate study was not significantly related to classroom teachers' and curriculum supervisors' perceptions of the education program evaluation process. Hypothesis VI, therefore, was not rejected.

Table 16

Summary of Significant Variables' Effects on Educators' Perceptions of Educational Program Evaluation

	Position	Program Evaluation Experience	Cross-District Differences	Years of Experience		Highest Level of Education	
				Comparisons	Correlations	Comparisons	Correlations
Context Evaluation	Principals assign greater importance than curriculum personnel.		A-B Overall	3-5, 1-2[a] 3-5, 6-10 3-5, >10			Positive for all teachers
1. Identify current and desired outcomes.				3-5, 6-10 3-5, >10	Positive Overall		
2. Identify unmet needs.		Evaluation committee teachers assign less importance than other teachers.	A-B Curriculum Personnel	1-2, 3-5 3-5, 6-10			
3. Identify human & material resources.			A-B Overall	3-5, >10	Positive for principals		
4. Gather information from outside sources.	Principals assign greater importance than curriculum personnel.	Teachers with evaluation experience assign greater importance than other teachers.				Bchlr, Mstr Bchlr, CAS Bchlr, Dctr	Positive for all groups Positive for all teachers
5. Explore other available programs.			A-B Overall A-B Teachers--no evaluation experience.	3-5, >10			

	Position	Program Evaluation Experience	Cross-District Differences	Years of Experience		Highest Level of Education	
				Comparisons	Correlations	Comparisons	Correlations
Input Evaluation	Classroom teachers assign greater importance than curriculum personnel.			1-2.3-5			
6. Assess present strategies.							Negative Overall
7. Identify costs & benefits of other available programs.	Classroom teachers assign greater importance than curriculum supervisor groups.	Evaluation committee teachers differ in districts A and D.	A-D Evaluation committee teachers				Negative Overall
8. Determine capabilities of present staff, facilities & resources.			A-C Curriculum Personnel				
9. Determine implementation logistics.		Evaluation committee teachers assign more importance than other teachers.					
10. Determine administration & evaluation plan.					Negative for Curriculum Personnel		

	Position	Program Evaluation Experience	Cross-District Differences	Years of Experience		Highest Level of Education	
				Comparisons	Correlations	Comparisons	Correlations
Process Evaluation							
11. Develop implementation plan.		Evaluation committee teachers differ in districts A and C.	A-C Evaluation committee teachers		Positive for teachers with evaluation experience		Negative for teachers without evaluation experience
12. Assess human and material resources during implementation.							
13. Determine piloting & implementation information needs.		Teachers with evaluation experience assign less importance than other teachers.					
14. Monitor concerned publics' understanding & agreement.			D-A, D-B, D-C Principals				
15. Design & assess communication channels.					Negative for Evaluation committee teachers		

	Position	Program Evaluation Experience	Cross-District Differences	Years of Experience		Highest Level of Education	
				Comparisons	Correlations	Comparisons	Correlations
Product Evaluation	Curriculum supervisors assign greater importance than teachers.						
16. Determine degree to which objectives are realized.		Evaluation committee teachers differ in districts A and B.	A-B Evaluation committee teachers				
17. Assess student achievement.	Curriculum supervisors assign greater importance than teachers.		A-B Overall B-A, B-D for Teachers--no evaluation experience A-B Curriculum personnel				
18. Assess attitudes regarding outcomes.	Curriculum supervisors assign greater importance than teachers.	Evaluation committee teachers differ in districts A and D.	A-D Overall A-D Evaluation committee teachers. A-B Principals	1-2,6-10	Negative for all groups	CAS,Dctr	
19. Identify unanticipated outcomes and their effects.		Teachers without evaluation experience differ with other teachers in districts A and B, A and C.	A-B Overall A-B, A-C Evaluation committee teachers	3-5, >10	Negative for evaluation committee and all teachers		Negative for teachers--some other evaluation experience

	Position	Program Evaluation Experience	Cross-District Differences	Years of Experience		Highest Level of Education	
				Comparisons	Correlations	Comparisons	Correlations
20. Conduct cost/benefit analysis.		Evaluation committee teachers assign less importance than other teachers.	C-D Evaluation committee teachers		Positive for all teachers.	Mstr,<u>Dctr</u>	Negative for evaluation committee teachers

Note: The variable of major area of graduate study is omitted from this table because no significant results for this variable were identified in this study.

[a]An underlined comparison category indicates category to which greater importance was assigned.

CHAPTER V

DISCUSSION

Summary

This study described classroom teachers' and curriculum supervisors' perceptions of the educational program evaluation process in terms of the CIPP Evaluation Model. These educators were asked to assess the importance of the four types of evaluation, and tasks representative of these types of evaluation, that were delineated in the CIPP Model. The study analyzed the effects of professional position, the experience of having served on an inservice evaluation committee, years of experience in present position, highest level of education and major areas of graduate study on these educators' perceptions of the educational program evaluation process. With the exception of major areas of graduate study, each of these variables was found to have influenced the perceptions of program evaluation.

Overall, the perceptions of context and product evaluation were more often affected by the above variables than were perceptions of input and process evaluation. This same phenomenon was true for tasks representative of each of these types of evaluation. Educators' responses concerning context evaluation were found to be affected by professional position, years of experience, and level of education. Perceptions of tasks repre-

sentative of context evaluation were also affected by these variables. Responses concerning one task regarding gathering information from sources outside the school district (such as research findings, consultants or subject area experts) were particularly affected by the variables of professional position, highest level of education and the experience of having served on an inservice evaluation committee.

Responses concerning product evaluation were found to be affected only by professional position. Perceptions of tasks representative of product evaluation, however, were found to be affected by all of the variables except major area of graduate study. Three product evaluation tasks were most often affected by these variables. These tasks concerned assessing concerned publics' attitudes regarding program outcomes, identifying unanticipated outcomes and their effects on students and faculty, and determining whether or not an educational program is cost-effective.

Classroom teachers' and curriculum supervisors' responses concerning input evaluation were affected by the variables of professional position and years of experience. These educators' perceptions of process evaluation, however, were not affected by any of the variables identified in this study. Tasks representative of these types of evaluation were found to be somewhat affected by each of the variables except major area of

graduate study, but, generally to a far lesser degree than were context and product evaluation tasks.

The variables of professional position, years of experience, highest level of education and major area of graduate study were found to affect classroom teachers' and curriculum supervisors' perceptions of the educational program evaluation process in different ways and to different degrees. This was anticipated due to the nature of their professional responsibilities. Curriculum supervisors' assigned greater importance to product evaluation. This group's concern with the overall implementation and outcomes of educational programs from a central perspective would indicate that product evaluation would be a higher priority than it would for classroom teachers. Input evaluation, however, was perceived to be more important by classroom teachers. This finding suggests that not only do classroom teachers recognize the importance of input evaluation, but that they attribute more importance to teacher contributions in this sphere of evaluation as the potential implementers of the program. It appears that curriculum supervisors place less importance on these evaluation activities.

The variable of years of experience in present position was also found to have an overall affect on both classroom teachers' and curriculum supervisors' perceptions of context and product evaluation. This finding reinforces the differences identified

in these groups' responses concerning program evaluation in terms of their professional position.

The experience of having served on an inservice evaluation committee for the purpose of selecting a new basal reading program was the variable which most often affected perceptions of the program evaluation process. Comparisons and contrasts involving teachers with this inservice evaluation experience were more often significant than were the findings concerning any other group.

The findings concerning the effect of highest level of education indicate that this variable had little influence upon classroom teachers' and curriculum supervisors' perceptions of the program evaluation process. Educators' responses concerning the importance of context evaluation and some product evaluation tasks increased somewhat overall, and especially for classroom teachers, as level of education increased. No other pattern of influence was found for level of education. Also, no evidence was found indicating that a particular major area of graduate study influenced perceptions of educational program evaluation. These findings were disappointing indicating that the science of evaluation remains a neglected area of graduate study in education. One would expect that graduate education programs designed to equip practitioners with the skills necessary for program development, implementation and evaluation would empha-

size evaluation theory and methodology. Evidence of such preparation was not found in this study.

Conclusions

The CIPP Evaluation Model was found to be an effective means through which to investigate educators' perceptions of the program evaluation process. Classroom teachers and curriculum supervisors acknowledged the importance of the four types of evaluation and representative evaluation tasks delineated in the CIPP Model. Furthermore, these educators were able to identify the relative importance of these evaluation types and tasks in the total sphere of evaluation as defined by the CIPP Model.

This study identified the effects of the variables of professional position, years of experience in present position, highest level of education, major area of graduate study and the experience of having served on a formal inservice program evaluation committee on classroom teachers' and curriculum supervisors' perceptions of the program evaluation process. A major finding concerned differences in perceptions of context and product evaluation. In a study investigating teachers' and administrators' perceptions of the availability of evaluation information in schools, Nevo and Stufflebeam (1975) found that classroom teachers and school administrators perceived context and product evaluation information to be the most often available.

Their findings suggest that the differences in educators' perceptions of program evaluation identified in the present study may be due to these educators' familiarity with context and product evaluation information and their experience using it in their work.

Several other factors support this contention. Context and product evaluation tasks more closely resemble instructional and administrative tasks than do input and process evaluation tasks. Furthermore, there has been a major emphasis in the past decade to specify educational objectives and to demonstrate that students achieved the stated objectives. Identifying objectives and assessing student achievement are tasks which generate context and product evaluation information. The higher frequency of significant results concerning context and product evaluation may be due to educators' familiarity in dealing with these kinds of evaluation information whereas the lesser frequency of significant results concerning input and process evaluation may reflect their unfamiliarity with such evaluation information.

The variable which most often affected responses concerning program evaluation was the experience of having served on a formal inservice program evaluation committee. Teachers who served on such committees demonstrated more significant differences in their perceptions of program evaluation than did any other group. These differences were particularly evident in

cross-district comparisons indicating that the trends identified for the population as whole were reinforced by further breakdown of the results.

Analyses involving the curriculum supervisor groups, however, did not demonstrate as many significant findings as did those involving classroom teachers. It would be expected that curriculum supervisors would perceive the process of program evaluation differently from classroom teachers due to the supervisors' professional responsibilities which include the development, implementation and evaluation of instructional programs. Such findings were only somewhat evident in this study. Instead, the experience of being part of a formal program evaluation was found to most often affect perceptions of the educational program evaluation process--not professional position, years of experience, highest level of education or major area of graduate study.

The educators in this study demonstrated a recognition of the importance of the four types of evaluation and tasks representative of each type of evaluation identified in the CIPP Evaluation Model. These findings suggest that classroom teachers and curriculum supervisors have the perceptual base, or readiness, necessary to pursue comprehensive educational program evaluations. The discrepancy identified between the sophistication of evaluation theory and methodology and the

practices of school systems in their evaluation of educational programs apparently is not due to the inability of practitioners to recognize the importance of a wide range of evaluation procedures.

Ten years ago Guba and Stufflebeam (1970) identified areas of need in the practice of educational program evaluation. These needs were: the need for evaluation theory from which methodology and instrumentation could be developed, and, a need for trained evaluators. Since that time, there have been many contributions to the literatrue of evaluation. Today evaluators have a fund of evaluation theory and technology to draw upon in the design of comprehensive program evaluations. The results of this study, however, suggest that the need for trained evaluators remains unsatisfied. In this study, the variables which most often influenced educators' perceptions of evaluation were not those related to training--position, level of education or major area of graduate study. Instead, the crucial variable was the experience of participating in a comprehensive program evaluation. It appears, therefore, that the preparation of educational leaders responsible for evaluating educational programs has not sufficiently emphasized the theory and practice of evaluation.

Recommendations for Future Studies

1. This study examined educators' perceptions of program evaluation in four school districts and a companion study (Smith, 1980) examined evaluation practices in these same school districts. Although both studies were based upon the CIPP Evaluation Model, the studies were independent. Future studies of educators' perceptions of the program evaluation process should include a behavioral component to examine the relationship of perceptions to practices of evaluation in the context of a single research design.

2. The need for educational leaders trained in the science of evaluation suggests several avenues for future research. There is a need to determine the status of evaluation in graduate education course content and to identify the areas where practitioners can gain exposure to evaluation theory and practice. This need is present in formal graduate education programs where future leaders are being trained as well as in institute and inservice programs to reach the present leadership in the field.

3. Similar studies of evaluation could describe the perceptions and practices of evaluation held by practitioners in smaller school systems. Such studies could focus on the availbility of evaluation information in smaller school districts and the impact of curriculum supervisory personnel on the practice

of program evaluation.

4. The CIPP Evaluation Model has been demonstrated to be an effective evaluation research tool in a variety of contexts. Other evaluation models should also be employed in future studies to examine their usefulness in evaluation research.

5. The successful participation of practitioners in this study demonstrated that these educators can make important contributions to evaluation research. Future research should include practitioners as active participants as well as subjects for observation.

REFERENCES

Alkin, M.C. Evaluation theory 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., Hastings, J., & Madus, G. Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.

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)

Center for the Advanced Study of Educational Administration. CASEA Progress Report. Eugene, Oregon: University of Oregon, 1971 & 1972.

Clifford, G. A history of the impact of action research on teaching. In R.M. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

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)

Corey, S. Action research to improve school practices. New York: Bureau of Publications, Teachers College, Columbia University, 1953.

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

Cronbach, L.J. Evaluation in a contest 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.

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 Scoiety 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 147934)

Forgione, P. Evaluation of compensatory education programs: problems, promising strategies and recent trends. San Francisco: Paper presented at the Annual Meeting of the American Educational Research Association, 1979. (ERIC Document Reproduction Service No. ED 175933)

Gess, E. 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.

Gowin, R. Can meta-evaluation give a direction for research on evaluation. Toronto: Paper presented at the Annual Meeting of the American Educational Research Association, 1978. (ERIC Document Reproduction Service No. ED 161938)

Grasso, J. Impact Evaluation in Vocational Education: The State of the Art. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, 1979. (ERIC Document Reproduction Service No. ED 173547)

Grobman, H. Evaluation activities of curriculum projects: a starting point. In R.E. Stake (Ed.), AERA monograph series on curriculum evaluation. Chicago: Rand McNally & Company, 1968.

Grobman, H. Developmental curriculum projects: decision points and processes. Itasca, Illinois: Peacock, 1970.

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.

Hays, W. Statistics for the Social Sciences. New York: Holt, Rinehart and Winston, 1973, 606-607.

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).

Hogkinson, H. Action Research: a critique. Journal of Educational Sociology, 1957, 31, 137-153.

Jenkins, W., & Lehman, G. Nine pitalls of PPBS. School Management, 1972, 16, 2-6.

Kunder, L. Procedures for textbook and instructional materials selection: ERS report. Arlington, Virginia: Educational Research Service, 1976. (ERIC Document Reproduction Service No. ED 132677)

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)

Merkel, C. Multi-dimensional planning/evaluation schema for community education. Paper presented at the Annual Meeting of the American Educational Research Association, 1979. ERIC Document Reproduction Service No. ED 177216)

Metfessel, N., & Michael, W. A paradigm involving multiple criterion measures for the evaluation 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: Illinois University, 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)

Nie, N., Hull, C., Jendins, J., Steinbrenner, K., & Brent, D. SPSS: Statistical package for the social sciences. New York: McGraw-Hill, 1975.

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)

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. Educational evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Provus, M. Discrepancy evaluation. Berkley, 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. 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. (University Microfilms No. 69-8796).

Rose, D., & 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. An introduction to meta-evaluation. Educational Product Report, February, 1969, 2(5).

Scriven, M. Evaluation perspectives and procedures. In W. J. Popham (Ed.), Evaluation in education: current applications. Berkley, California: McCutchan, 1974.

Scriven, M. The methodology of evaluation. In A. Bellach & 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).

Smith, K. An analysis of the practice of educational program evaluation in terms of the CIPP Evaluation Model. Unpublished doctoral dissertation, Loyola University, Chicago, 1981.

Stake, R. The countenance of educational evaluation. In A. Bellack & H. Kliebard (Eds.), Curriculum and evaluation. Berkley, California: McCutchan, 1977.

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, R. A third party evaluation of the Appalachian Maryland experienced-based career education project: 1976-1977. Morgantown, West Virginia: West Virginia University, 1977. (ERIC Document Reproduction Service No. ED 147510)

Stufflebeam, D. The relevance of the CIPP Evaluation Model for educational accountability. Journal of Research and Development in Education, 1971, 5, (1), 19-25.

Stufflebeam, D. Toward a technology for evaluating evaluation. Chicago: Paper presented at the Annual Meeting of the American Educational Research Association, 1974. (ERIC Document Reproduction No. 090319)

Stufflebeam, D. Meta-evaluation. Evaluation Center, Occasional Paper, 1977, 3.

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, 1972, 73(1), 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 Service No. ED 163059)

Thorndike, R., & Hagan, 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: Univerisyt 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)

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, Pennsylvannia: Pittsburgh University, 1972. (ERIC Document Reproduction Service No. ED 064376)

Worthen, B., & Sanders, J. Educational evaluation: theory and practice. Wrothington, Ohio: Charles A. Jones, 1973.

APPENDIX A

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. Annonymity 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 this Study	Time Required
ge 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
e 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

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. Annonymity 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.

EDUCATOR VARIABLES

PLEASE GIVE THE INFORMATION BELOW:

1. Years of full time teaching experience (including this year).

   _____ 1-2 years

   _____ 3-5 years

   _____ 6-10 years

   _____ More than ten years.

2. Highest level of education.

   _____ Bachelor's degree

   _____ Master's degree

   _____ Certificate of Advanced Study

   _____ Doctorate

3. If you hold a graduate education degree, indicate your major area of study.

   _____ Administration/Supervision

   _____ Curriculum and Instruction

   _____ Other, please specify _____

4. Have you participated in a program evaluation as a member of an evaluation committee?

   _____ Yes

   _____ No

EDUCATOR VARIABLES

PLEASE GIVE THE INFORMATION BELOW:

1. Present position.

   _____ principal

   _____ curriculum consultant/coordinator/supervisor

   _____ assistant superintendent (curriculum)

   _____ other, please specify _____

2. Number of years in your present position.

   _____ 1-2

   _____ 3-5

   _____ 6-10

   _____ more than ten years

3. Highest level of education.

   _____ Bachelor's degree

   _____ Master's degree

   _____ Certificate of Advanced Study

   _____ Doctorate

4. If you hold a graduate education degree, indicate your major area of study.

   _____ Administration/Supervision

   _____ Curriculum and Instruction

   _____ other, please specify _____

DIRECTIONS

1. Four types of evaluation are described below. Each of the four types of evaluation is comprised of appropriate tasks to be completed in that sphere of evaluation. Within each type of evaluation, rank the five tasks according to your perception of their relative importance. Rank the most important task as "1" and the remaining tasks as "2" through "5" accordingly.

2. Assume that the process of educational program evaluation can be represented by 100 points. Assuming that each point represents an equal measure of value, or importance, divide the 100 points among the four types of evaluation described below according to your perception of the relative importance of each type.

CONTEXT EVALUATION

Purpose: to provide a rationale for determining program objectives.

TASKS

_____ 1. Identify learning outcomes of current program and identify learning outcomes desired from a new program.

_____ 2. Identify needs not being served by current program.

_____ 3. Identify potential human and material resources available to implement a new program.

_____ 4. Gather information from sources outside the school district such as research findings or outside consultants.

_____ 5. Explore other available programs in terms of the impact of change on students, faculty, parents and community.

INPUT EVALUATION

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

TASKS

_____ 1. Determine what is already being done to meet a new set of objectives.

_____ 2. Identify potential costs and benefits of other available programs.

_____ 3. Determine how existing staff, facilities and resources can be utilized to implement the new program.

_____ 4. Determine a specific schedule of events and activities to guide the new program's implementation.

_____ 5. Determine how the new program should be administered, evaluated and reviewed at various levels.

PROCESS EVALUATION

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

TASKS

_____ 1. Develop an implementation plan for the new program.

_____ 2. Determine the adequacy of resources, facilities, staff and time schedules during implementation of the new program.

_____ 3. Determine the kinds of feedback needed during piloting and implementation.

_____ 4. Monitor the various publics (teachers, students, administrators, parents) understanding of and agreement with the program.

_____ 5. Design and assess communication channels between teachers, consultants, administrators.

PRODUCT EVALUATION

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

TASKS

_____ 1. Determine whether or not the program is achieving its objectives.

_____ 2. Assess gains (or losses) in pupil achievement.

_____ 3. Asses the attitudes of students, staff, parents and community regarding the outcomes of the program.

_____ 4. Identify unanticipated outcomes and their effects on students and faculty.

_____ 5. Determine whether or not the program results justify the finances and efforts needed to maintain it.

100 TOTAL POINTS

APPENDIX B

Initial Draft PART I: EDUCATOR VARIABLES (TEACHERS)

PLEASE GIVE THE INFORMATION BELOW.

1. Years of teaching experience (including this year).
   ___ 1-2 years
   ___ 3-5 years
   ___ 6-10 years
   ___ more than ten years

2. Level of education.
   ___ Bachelor's degree
   ___ Bachelor's degree plus 15 graduate education credit hours
   ___ Master's degree
   ___ Master's degree plus 15 graduate education credit hours
   ___ Master's degree plus 30 graduate education credit hours
   ___ Doctorate

3. If you hold a graduate education degree, indicate your major area of study.
   ___ Administration/Supervision
   ___ Curriculum and Instruction
   ___ Other, please specify

4. Have you participated in a curriculum evaluation as a member of an evaluation committee?
   ___ Yes, the reading program evaluation identified above
   ___ Yes, another curriculum evaluation
   ___ No

Initial Draft PART I: EDUCATOR VARIABLES (SUPERVISORS)

PLEASE GIVE THE INFORMATION BELOW.

1. Present position.
   ___ principal
   ___ curriculum consultant/coordinator/supervisor
   ___ assistant superintendent (curriculum)
   ___ resource teacher
   ___ other, please specify _____

2. Number of years in your present position.
   ___ 1-2 years
   ___ 3-5 years
   ___ 6-10 years
   ___ more than ten years.

3. Level of education.
   ___ Bachelor's degree
   ___ Bachelor's degree plus 15 graduate education credit hours
   ___ Master's degree
   ___ Master's degree plus 15 graduate education hours
   ___ Master's degree plus 30 graduate education hours
   ___ Doctorate

4. If you hold a graduate education degree, indicate your major area of study.
   ___ Administration/Supervision
   ___ Curriculum and Instruction
   ___ other, please specify _____

Initial Draft

PART II: CIPP EVALUATION MODEL

DIRECTIONS

1. Assume that the process of educational program evaluation can be represented by 100 points. Assuming that each point represents an equal measure of value, or importance, divide the 100 points among the four stages of evaluation described below according to your perception of the relative importance of each stage.
2. Each of the four stages of evaluation is comprised of appropriate tasks to be completed in that sphere of evaluation. Within each stage of evaluation, divide the number of points you assigned to each stage among the tasks listed for each stage.

____ Stage 1: Context Evaluation

Purpose: to provide a rationale for determining program objectives.

TASKS

_____ 1. Identify learning outcomes of current program and identify learning outcomes desired from a new program.

_____ 2. Identify needs not being served by current program.

_____ 3. Identify potential human and material resources available to implement a new program.

_____ 4. Gather information from sources outside the school district such as research findings or outside consultants.

_____ 5. Explore other available programs in terms of the impact of change on students, faculty, parents and community.

____ Stage 2: Input Evaluation

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

TASKS

_____ 1. Determine what is already being done to meet a new set of objectives.

_____ 2. Identify potential costs and benefits of other available programs.

_____ 3. Determine how existing staff, facilities and resources can be utilized to implement the new program.

4. Determine a specific schedule of events and activities to guide the new program's implementation.
5. Determine how the new program should be administered, evaluated and reviewed at various levels.

Stage 3: Process Evaluation
Purpose: to provide periodic feedback to persons responsible for implementing the new program.

TASKS

1. Develop an implementation plan for the new program.
2. Determine the adequacy of resources, facilities, staff and time schedules during implementation of the new program.
3. Determine the kinds of feedback needed during piloting and implementation.
4. Design and assess communication channels between teachers, consultants, administrators.
5. Monitor the various publics (teachers, students, administrators, parents) understanding of and agreement with the program.

Stage 4: Product Evaluation
Purpose: to measure and interpret attainments during the implementation and duration of the program.

TASKS

1. Determine whether or not the program is achieving its objectives.
2. Determine whether or not students' needs are being met by the program.
3. Assess gains (or losses) in pupil achievement.
4. Identify unanticipated outcomes and their effects on students and faculty.
5. Determine whether or not the program results justify the finances and efforts needed to maintain it.

00 TOTAL POINTS

Second Draft

PART I: EDUCATOR VARIABLES (TEACHERS)

PLEASE GIVE THE INFORMATION BELOW.

1. Years of full time teaching experience (including this year).
   ____ 1-2 years
   ____ 3-5 years
   ____ 6-10 years
   ____ More than ten years

2. Level of education.
   ____ Bachelor's degree
   ____ Bachelor's degree plus 1-15 graduate education credit hours
   ____ Master's degree
   ____ Master's degree plus 1-15 graduate education credit hours
   ____ Master's degree plus 16 or more graduate education credit hours
   ____ Certificate of Advanced Study
   ____ Doctorate

3. If you hold a graduate education degree, indicate your major area of study.
   ____ Administration/Supervision
   ____ Curriculum and Instruction
   ____ Other, please specify _____

4. Have you participated in a program evaluation as a member of an evaluation committee?
   ____ Yes, the recent reading program evaluation identified above
   ____ Yes, another program evaluation
   ____ No

PART I: EDUCATOR VARIABLES (SUPERVISORS)

Second Draft

PLEASE GIVE THE INFORMATION BELOW.

1. Present position.
   ____ principal
   ____ curriculum consultant/coordinator/supervisor
   ____ assistant superintendent (curriculum)
   ____ other, please specify _____

2. Number of years in your present position.
   ____ 1-2 years
   ____ 3-5 years
   ____ 6-10 years
   ____ more than ten years.

3. Level of education.
   ____ Bachelor's degree
   ____ Bachelor's degree plus 1-15 graduate education credit hours
   ____ Master's degree
   ____ Master's degree plus 1-15 graduate education hours
   ____ Master's degree plus 16-30 graduate education hours
   ____ Certificate of Advanced Study
   ____ Doctorate

4. If you hold a graduate education degree, indicate your major area of study.
   ____ Adminstration/Supervision
   ____ Curriculum and Instruction
   ____ other, please specify _____

Second Draft

PART II: CIPP EVALUATION MODEL

DIRECTIONS

1. Assume that the process of educational program evaluation can be represented by 100 points. Assuming that each point represents an equal measure of value, or importance, divide the 100 points among the four stages of evaluation described below according to your perception of the relative importance of each stage.
2. Each of the four stages of evaluation is comprised of appropriate tasks to be completed in that sphere of evaluation. Within each stage of evaluation, divide the number of points you assigned to each stage among the tasks listed for each stage.

____ Stage 1: Context Evaluation
Purpose: to provide a rationale for determining program objectives.

TASKS

_____ 1. Identify learning outcomes of current program and identify learning outcomes desired from a new program.

_____ 2. Identify needs not being served by current program.

_____ 3. Identify potential human and material resources available to implement a new program.

_____ 4. Gather information from sources outside the school district such as research findings or outside consultants.

_____ 5. Explore other available programs in terms of the impact of change on students, faculty, parents and community.

____ Stage 2: Input Evaluation
Purpose: to provide information for determining how to utilize resources to meet program goals.

TASKS

_____ 1. Determine what is already being done to meet a new set of objectives.

_____ 2. Identify potential costs and benefits of other available programs.

_____ 3. Determine how existing staff, facilities and resources can be utilized to implement the new program.

PROCESS EVALUATION

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

TASKS

_____ 1. Develop an implementation plan for the new program.

_____ 2. Determine the adequacy of resources, facilities, staff and time schedules during implementation of the new program.

_____ 3. Determine the kinds of feedback needed during piloting and implementation.

_____ 4. Monitor the various publics (teachers, students, administrators, parents) understanding of and agreement with the program.

_____ 5. Design and assess communication channels between teachers, consultants, administrators.

PRODUCT EVALUATION

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

TASKS

_____ 1. Determine whether or not the program is achieving its objectives.

_____ 2. Assess gains (or losses) in pupil achievement.

_____ 3. Asses the attitudes of students, staff, parents and community regarding the outcomes of the program.

_____ 4. Identify unanticipated outcomes and their effects on students and faculty.

_____ 5. Determine whether or not the program results justify the finances and efforts needed to maintain it.

100 TOTAL POINTS

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

	Position	Employer
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 C

Summary of elementary school faculty's response to research instrument: Educator Variables

Classroom Teachers 12

Curriculum Supervisors 3

Years of Experience in Present Position

1 1-2 years

3 3-5 years

4 6-10 years

7 More than ten years

Highest Level of Education

9 Bachelor's degree

5 Master's degree

0 Certificate of Advanced Study

1 Doctorate

Major Area of Graduate Study

2 Administration and Supervision

2 Curriculum and Instruction

2 Other

Classroom Teacher's Experience on a Program Evaluation Inservice Committee

3 Yes

9 No

Summary of elementary school faculty's initial response to research instrument: Types of Evaluation

Rank in terms of importance	Evaluation Type	Mean	Range	Standard Deviation
First	Product	27	25	7
Second	Context	26	20	7.5
Third	Input	24	30	8.2
Fourth	Process	22	40	9.8

Summary of elementary school faculty's second response to research instrument: Types of Evaluation

Rank in terms of importance	Evaluation Type	Mean	Range	Standard Deviation
First	Product	30	25	6.5
Second	Context	26	20	6.6
Third	Input	24	20	7.7
Fourth	Process	21	15	5.7

Summary of elementary school faculty's initial and second response to research instrument: Ranking of Tasks

Task Number	Initial Response	Second Response
CONTEXT		
1	1.5	4
2	3	9
3	14.5	18
4	19	20
5	14.5	16
INPUT		
6	12	7
7	16	13
8	11	10
9	9	14
10	5	11
PROCESS		
11	7	5
12	8	3
13	13	8
14	10	17
15	7	5

Summary of elementary school faculty's initial and second response to research instrument: Ranking of Tasks

Task Number	Initial Response	Second Response
PRODUCT		
16	6	2
17	1.5	1
18	4	6
19	20	15
20	17	19

Spearman Rank Correlation Coefficient

$$r_s = 1 - \left[\frac{6(\sum_i D_i^2)}{N^2(N - 1)}\right]$$

$$r_s = .768$$

APPENDIX D

Table 17

District Breakdown of Years Of Experience in Present Position

Years of Experience	Group 1	Group 2	Group 3	Group 4	Group 5
District A					
1-2	0	0	0	2	4
3-5	1	2	4	1	10
6-10	8	5	3	6	6
10	7	4	3	3	2
District B					
1-2	11	7	4	0	2
3-5	4	7	11	4	6
6-10	1	1	0	3	0
10	0	0	1	1	0
District C					
1-2	1	0	1	1	2
3-5	2	3	0	5	3
6-10	6	4	3	1	3
10	3	5	8	3	0
District D					
1-2	1	0	0	2	1
3-5	0	0	5	2	4
6-10	9	4	3	2	3
10	9	8	3	0	2

Table 18

District Breakdown for Highest Level of Education

Highest Level of Education	Group 1	Group 2	Group 3	Group 4	Group 5
District A					
Bachelors	7	6	1	0	7
Masters	9	3	9	11	10
CAS	0	1	0	0	4
Doctorate	0	1	0	1	1
District B					
Bachelors	11	7	4	0	2
Masters	4	7	11	4	6
CAS	1	1	0	3	0
Doctorate	0	0	1	1	0
District C					
Bachelors	9	6	5	0	0
Masters	3	6	7	9	5
CAS	0	0	0	1	3
Doctorate	0	0	0	0	0
District D					
Bachelors	9	5	0	0	0
Masters	10	7	6	4	5
CAS	0	0	4	2	4
Doctorate	0	0	1	0	1

APPENDIX E

Table 19

Population Measures of Central Tendency for Weighted Task Comparisons

	Evaluation Tasks	Classroom Teachers Mean	Classroom Teachers Standard Deviation	Curriculum Supervisors Mean	Curriculum Supervisors Standard Deviation
CONTEXT	1	120.6	47.9	115.9	41.3
	2	117.0	45.7	116.7	43.7
	3	64.2	32.6	61.5	28.5
	4	51.2	31.0	52.9	36.5
	5	63.0	38.6	60.5	32.3
INPUT	6	96.4	41.7	87.5	38.6
	7	56.3	38.8	43.5	27.8
	8	89.0	37.4	84.6	30.5
	9	61.7	39.6	53.6	31.5
	10	68.5	33.6	64.6	31.8
PROCESS	11	85.5	37.1	84.6	39.8
	12	67.0	31.7	73.4	33.7
	13	77.1	37.3	82.1	36.9
	14	43.5	32.9	47.1	35.5
	15	68.5	36.8	65.5	40.6
PRODUCT	16	101.0	42.9	110.3	50.1
	17	95.9	43.1	107.7	43.1
	18	72.5	40.0	85.2	42.6
	19	58.1	30.4	58.7	26.9
	20	43.1	32.3	47.3	38.4

Table 20

Classroom Teacher Group Measures of Central Tendency

For Weighted Task Comparisons

	Evaluation Tasks	Group 1 Mean	Group 1 Standard Deviation	Group 2 Mean	Group 2 Standard Deviation	Group 3 Mean	Group 3 Standard Deviation
CONTEXT	1	114.9	46.1	126.3	60.5	122.0	33.9
	2	108.3	42.4	120.8	53.5	124.4	39.7
	3	63.5	30.1	67.2	36.6	62.1	31.1
	4	42.2	23.1	54.1	38.7	59.6	28.8
	5	60.8	35.4	68.7	47.4	59.9	32.2
INPUT	6	102.4	44.2	91.0	41.0	94.3	39.1
	7	58.5	38.3	55.1	35.2	54.6	43.2
	8	89.4	38.6	85.8	39.3	91.7	34.4
	9	55.8	33.2	59.4	44.3	71.7	40.9
	10	67.6	32.7	64.8	30.8	73.3	37.4
PROCESS	11	83.2	35.0	82.3	39.2	91.6	37.6
	12	68.0	32.3	63.5	32.0	69.1	31.0
	13	86.6	36.4	71.7	41.2	70.4	32.2
	14	41.2	33.8	48.5	35.8	41.3	28.4
	15	74.3	36.5	63.1	38.2	66.7	35.5
PRODUCT	16	106.4	45.6	102.7	49.9	92.3	28.7
	17	98.4	46.7	98.9	50.5	89.8	27.3
	18	73.1	40.0	70.8	43.4	73.6	37.1
	19	57.6	30.2	57.2	27.3	59.6	33.9
	20	49.7	34.3	45.8	37.5	31.9	18.9

Table 21

Curriculum Supervisor Group Measures of Central Tendency

For Weighted Task Comparisons

	Evaluation Tasks	Group 4 Mean	Group 4 Standard Deviation	Group 5 Mean	Group 5 Standard Deviation
CONTEXT	1	124.6	46.5	109.4	36.0
	2	121.9	41.3	112.8	45.4
	3	64.4	29.1	59.2	28.2
	4	63.8	46.1	44.8	24.7
	5	62.8	35.3	58.8	30.0
INPUT	6	91.0	44.9	84.9	33.3
	7	44.6	29.7	42.7	26.5
	8	88.5	35.8	83.1	26.1
	9	58.2	32.1	50.2	32.2
	10	69.7	31.2	60.7	32.1
PROCESS	11	75.1	39.8	91.7	38.7
	12	70.0	34.5	76.0	33.2
	13	73.5	32.4	88.5	39.0
	14	54.5	38.2	41.4	32.6
	15	58.1	32.5	71.0	45.2
PRODUCT	16	99.3	46.0	118.5	51.9
	17	102.0	44.0	119.8	42.4
	18	85.3	34.6	85.1	48.1
	19	54.7	24.9	61.7	28.2
	20	44.2	34.1	49.6	41.5

Table 22

T-Test Results for Position Comparisons

		Teachers			Supervisors	Teachers & Supervisors
Type of Evaluation		1,2	1,3	2,3	4,5	1+2+3,4+5
Context		-1.69	-1.81	0.32	2.04*	0.46
Input		0.90	-0.66	-1.51	1.58	3.03**
Process		1.36	0.90	-0.54	-1.80	-0.93
Product		0.43	1.88	1.15	-1.46	-2.33*
Evaluation Tasks						
CONTEXT	1	0.37	0.73	0.34	-0.16	0.81
	2	-0.13	-1.11	-0.82	-0.47	-0.22
	3	0.86	1.67	0.85	-0.43	0.43
	4	-1.06	-2.70**	-1.29	1.51	-0.54
	5	0.03	1.13	1.09	-0.49	-0.30
INPUT	6	0.85	1.36	0.53	-0.24	-0.29
	7	0.08	1.20	1.10	-0.17	1.54
	8	-0.46	0.26	0.20	-1.10	-1.52
	9	-0.69	-2.46	-1.47	0.62	0.47
	10	0.13	-0.35	-0.48	0.78	-0.46
PROCESS	11	-1.01	-1.96	-0.87	-0.93	1.03
	12	-0.54	-0.95	-0.37	-0.17	-0.95
	13	1.53	2.50*	0.69	-1.04	-0.66
	14	-1.63	-0.39	1.21	2.92**	-0.66
	15	1.57	0.90	-0.55	-0.42	1.30

	Evaluation Tasks	1,2	1,3	2,3	4,5	1+2+3,4+5
PRODUCT	16	0.88	0.36	0.25	-1.09	0.58
	17	-0.44	-0.46	0.01	0.00	-0.39
	18	0.33	-1.05	-1.38	2.11*	-1.16
	19	-0.68	-1.51	-0.89	-0.51	0.78
	20	0.35	3.09**	2.42*	-0.06	0.26

*$p < .05$

**$p < .01$

Table 23

T-Test Results for Position Comparisons of Weighted Tasks

		Teachers			Supervisors	Teachers & Supervisors
Evaluation Tasks		1,2	1,3	2,3	4,5	1+2+3,4+5
CONTEXT	1	-1.13	-0.90	0.44	1.69	0.76
	2	-1.39	2.05*	-0.37	0.95	0.06
	3	-0.58	0.25	0.74	0.82	0.66
	4	-2.02*	-3.53**	-0.79	2.42*	-0.40
	5	-1.01	0.14	1.07	0.56	0.50
INPUT	6	1.41	1.01	-0.42	0.72	1.64
	7	0.48	0.50	0.06	0.31	2.68**
	8	0.49	-0.32	-0.80	0.51	0.93
	9	-0.50	-2.28*	-1.44	-1.16	1.63
	10	0.45	-0.87	-1.24	1.29	0.88
PROCESS	11	0.13	-1.22	-1.20	-1.92	0.17
	12	0.74	-0.18	-0.89	-0.80	-1.47
	13	2.04*	2.44*	0.16	-1.88	-0.99
	14	-1.12	-0.02	1.11	1.70	-0.79
	15	1.59	1.11	0.48	-1.46	0.59
PRODUCT	16	0.42	1.90	1.27	-1.76	-1.52
	17	-0.06	1.15	1.12	-1.05	-2.02*
	18	0.29	-0.07	-0.34	0.01	-2.30*
	19	0.09	-0.33	-0.40	-1.18	-0.15
	20	0.58	3.27**	2.32*	-0.64	-0.89

*p < .05 **p < .01

Table 24

F-Values for Cross-District Comparisons

Type of Evaluation		All Groups	Groups 1	2	3	4	5
Context		4.51**	2.04	1.02	0.71	0.40	3.70*
Input		1.19	0.54	1.70	2.22	0.74	1.03
Process		0.80	1.15	0.80	3.04*	0.86	0.29
Product		5.64**	1.50	0.23	1.70	1.24	7.08**
Evaluation Tasks							
CONTEXT	1	3.64*	1.74	2.01	0.08	0.64	1.50
	2	3.65*	1.16	0.21	0.44	0.68	4.00*
	3	3.49*	0.26	3.00*	1.91	0.84	1.18
	4	0.87	2.16	0.60	1.37	0.89	2.70
	5	4.73**	4.61**	1.44	0.11	0.10	1.83
INPUT	6	2.69	1.10	1.31	2.56	1.51	2.05
	7	2.00	0.86	2.63	3.11*	0.97	1.64
	8	2.08	1.03	1.28	0.49	0.43	4.70**
	9	1.81	1.97	0.30	1.40	0.62	2.72
	10	0.84	1.04	0.58	0.76	0.98	2.22
PROCESS	11	1.96	0.34	0.57	4.05*	0.05	2.19
	12	0.70	0.46	1.39	0.70	1.38	1.58
	13	2.47	0.54	0.42	3.19*	1.71	0.51
	14	1.59	0.58	0.29	1.14	6.31**	0.42
	15	0.24	1.28	0.29	0.69	0.37	0.28

Evaluation Tasks		All Groups	Groups 1	2	3	4	5
PRODUCT	16	0.68	0.66	0.86	4.23**	0.86	1.85
	17	5.98**	4.41	0.08	1.24	0.05	5.28**
	18	3.72*	1.15	0.72	4.23**	0.87	4.98**
	19	5.53**	0.65	0.13	9.91**	0.73	2.27
	20	2.48	0.56	1.05	4.31**	2.16	2.99*

*$p < .05$

**$p < .01$

Table 25

T-Test Results for Years of Experience Comparisons

Type of Evaluation		Years of Experience 1-2,3-5	1-2,6-10	1-2, 10	3-5,6-10	3-5, 10	6-10, 10
Context		2.87**	0.80	0.27	-2.83**	-2.86**	-0.69
Input		-1.99*	-0.92	-1.07	1.41	1.01	-0.37
Process		-1.12	-0.57	-0.32	1.31	1.43	0.34
Product		0.08	0.35	0.57	0.40	0.72	0.39
Evaluation Tasks							
CONTEXT	1	1.70	0.21	-0.47	-2.10*	-2.78**	-1.19
	2	2.65**	1.10	0.93	-2.34*	-1.82	0.10
	3	1.81	0.70	-0.03	-1.60	-2.11*	-1.03
	4	0.39	-0.13	0.07	-0.88	-0.54	0.38
	5	2.22	0.76	0.39	-1.65	-2.32*	-0.65
INPUT	6	-1.51	-0.45	-0.59	1.50	1.38	-0.19
	7	-0.18	-0.10	-0.82	0.14	-0.91	-1.30
	8	-1.44	-0.32	-0.08	1.52	1.79	0.37
	9	-1.52	-1.31	-1.60	0.36	-0.45	-0.91
	10	-0.15	-0.22	0.48	-0.11	0.95	1.91
PROCESS	11	-0.14	-0.36	-1.03	-0.27	-1.24	-1.21
	12	-1.43	-1.62	-1.44	0.02	0.15	0.16
	13	-1.28	0.77	-0.24	0.96	1.55	0.81
	14	0.65	1.83	1.83	1.36	1.30	-0.12
	15	-0.98	-0.49	0.16	1.06	1.96	1.17

Evaluation Tasks		Years of Experience 1-2,3-5	1-2,6-10	1-2, 10	3-5,6-10	3-5, 10	6-10, 10
PRODUCT	16	0.26	0.25	-0.22	-0.07	-0.68	-0.79
	17	0.65	0.37	1.05	-0.43	0.56	1.14
	18	0.58	0.84	1.98*	0.41	1.75	1.41
	19	-1.12	0.19	0.21	1.97	2.19*	-0.01
	20	-0.68	-0.90	-0.99	-0.44	-0.76	-0.46

*$p < .05$

**$p < .01$

Table 26

Pearson Correlation for Years of Experience

Type of Evaluation	All Groups	Teachers 1	Teachers 2	Teachers 3	Supervisors 4	Supervisors 5	All Teachers	All Supervisors
Context	0.10	-0.11	0.10	0.33	0.11	0.13	0.08	0.12
Input	0.01	-0.03	-0.06	-0.01	-0.01	-0.25	-0.04	-0.12
Process	-0.04	-0.06	0.24	-0.16	-0.03	-0.08	0.01	-0.07
Product	-0.05	0.18	-0.16	-0.21	0.01	0.07	-0.03	0.04
Evaluation Tasks								
CONTEXT 1	0.13*	-0.01	0.17	0.24	0.13	0.13	0.12	0.13
CONTEXT 2	0.03	-0.15	-0.01	0.23	-0.09	0.16	0.01	0.06
CONTEXT 3	0.09	-0.09	-0.03	0.17	0.44**	-0.02	0.02	0.19
CONTEXT 4	0.02	-0.22	0.20	-0.01	0.08	-0.01	0.01	0.05
CONTEXT 5	0.07	0.05	0.03	0.24	-0.15	0.10	0.10	-0.02
INPUT 6	-0.03	0.09	-0.26	-0.07	-0.19	-0.01	-0.06	-0.09
INPUT 7	0.07	-0.21	0.10	0.27	-0.05	-0.08	0.05	-0.07
INPUT 8	-0.07	-0.01	-0.18	-0.10	0.02	-0.23	-0.09	-0.10
INPUT 9	0.09	0.10	0.16	-0.08	0.15	0.02	0.05	0.08
INPUT 10	-0.06	-0.06	0.01	-0.08	0.12	-0.32*	-0.06	-0.12
PROCESS 11	0.09	-0.07	0.35*	0.08	0.23	-0.09	0.10	0.09
PROCESS 12	0.06	-0.01	0.03	0.12	0.27	0.12	0.04	0.18
PROCESS 13	-0.05	-0.13	0.08	-0.10	-0.23	-0.07	0.03	-0.13
PROCESS 14	-0.12	-0.12	0.01	-0.07	-0.21	-0.20	-0.54	-0.21
PROCESS 15	-0.08	-0.07	0.19	-0.40**	-0.17	-0.06	-0.11	-0.10

Evaluation Tasks	All Groups	Teachers 1	2	3	Supervisors 4	5	All Teachers	All Supervisors
16	0.04	0.22	-0.17	0.04	0.13	0.15	0.04	0.13
17	-0.07	0.24	-0.11	-0.18	-0.14	-0.06	-0.03	-0.10
18	-0.14*	-0.16	-0.19	-0.06	-0.06	0.01	-0.14	-0.02
19	-0.09	0.04	-0.19	-0.41**	-0.10	0.20	-0.19*	0.07
20	0.08	0.20	0.10	0.19	0.15	-0.02	0.16*	0.05

*$p < .05$

**$p < .01$

Table 27

T-Test Results for Highest Level of Education Comparisons

Type of Evaluation		Highest Level of Education: Bachelors, Masters	Bachelors, CAS	Bachelors, Doctorate	Masters, CAS	Masters, Doctorate	CAS, Doctorate
Context		-1.47	-1.39	-0.89	-0.65	-0.42	-0.03
Input		1.26	1.58	1.01	0.90	0.65	0.18
Process		0.11	-0.36	0.48	-0.96	0.24	0.81
Product		-0.37	0.38	-1.29	0.60	-1.07	-1.13
Evaluation Tasks							
CONTEXT	1	-0.65	-0.11	-0.96	0.30	-0.83	-0.90
	2	-1.87	-1.31	-0.67	-0.30	-0.02	0.12
	3	0.05	0.52	0.41	0.52	0.42	0.15
	4	-2.20*	-2.56*	-1.99*	-1.24	-0.85	-0.05
	5	0.44	-1.58	-0.15	-1.37	0.02	0.58
INPUT	6	0.53	1.51	1.75	1.40	1.80	0.93
	7	1.96	1.20	1.44	0.01	0.76	0.73
	8	0.98	0.62	0.38	0.08	0.07	0.03
	9	-0.57	-0.32	-1.47	0.03	-1.33	-1.42
	10	0.53	1.04	0.51	0.72	0.31	-0.10
PROCESS	11	1.06	-0.49	0.47	-1.14	0.06	0.70
	12	-0.13	-0.05	-0.67	0.03	-0.63	-0.55
	13	0.26	-0.67	0.31	-0.90	0.24	0.82
	14	0.33	0.08	1.43	-0.14	1.41	1.81
	15	0.52	0.29	-0.22	0.02	-0.48	-0.38

Evaluation Tasks	Highest Level of Education: Bachelors, Masters	Bachelors, CAS	Bachelors, Doctorate	Masters, CAS	Masters, Doctorate	CAS, Doctorate
16	-0.13	-1.02	0.70	-0.97	0.73	1.04
17	-1.01	0.07	-0.86	0.73	-0.54	-0.85
18	-0.60	1.17	-1.17	1.55	-0.92	-2.06*
19	-0.01	0.89	-1.08	0.87	-0.99	-1.34
20	0.92	0.57	-2.27*	0.01	-2.72**	-1.82

*$p < .05$

**$p < .01$

Table 28

Pearson Correlation for Highest Level of Education

Type of Evaluation	All Groups	Teachers 1	Teachers 2	Teachers 3	Supervisors 4	Supervisors 5	All Teachers	All Supervisors
Context	0.11	0.10	0.22	0.01	0.26	-0.05	0.15*	0.92
Input	-0.12	-0.17	-0.03	0.05	-0.03	-0.23	-0.03	-0.12
Process	-0.02	-0.16	0.09	0.04	-0.01	-0.09	-0.02	-0.08
Product	0.04	0.18	-0.23	-0.11	-0.01	0.19	-0.10	0.11
Evaluation Tasks								
CONTEXT 1	0.05	0.06	0.13	-0.06	0.21	-0.04	0.07	0.08
CONTEXT 2	0.11	0.05	0.18	0.09	0.21	-0.04	0.15	0.06
CONTEXT 3	-0.04	-0.03	0.12	-0.09	0.25	-0.26	-0.01	-0.06
CONTEXT 4	0.18**	0.05	0.20	0.14	0.22	0.09	0.20**	0.17
CONTEXT 5	0.07	0.23	0.21	-0.05	-0.10	0.08	0.12	0.02
6	-0.13*	-0.01	0.06	-0.05	-0.22	0.35*	-0.18	-0.28*
INPUT 7	-0.13*	-0.27*	0.17	-0.04	0.13	0.28	-0.06	-0.12
INPUT 8	-0.58	-0.10	-0.15	0.10	-0.15	0.04	-0.04	-0.04
INPUT 9	0.08	0.02	0.01	0.25	0.23	-0.04	0.14	0.07
10	-0.06	-0.13	-0.20	-0.15	0.05	0.09	0.12	0.09
PROCESS 11	-0.02	-0.25*	0.16	-0.02	0.07	-0.08	0.01	-0.05
PROCESS 12	0.03	-0.06	0.06	-0.06	0.09	-0.03	-0.01	0.01
PROCESS 13	0.01	-0.10	0.12	0.15	0.01	-0.03	0.01	-0.04
PROCESS 14	-0.06	0.11	-0.09	-0.20	-0.14	-0.20	-0.05	-0.14
PROCESS 15	-0.01	-0.08	-0.01	0.19	-0.06	0.04	0.01	-0.01

Evaluation Tasks		All Groups	Teachers 1	2	3	Supervisors 4	5	All Teachers	All Supervisors
PRODUCT	16	0.12	0.18	-0.27	-0.17	0.08	0.20	-0.12	0.14
	17	0.05	0.24	-0.09	-0.15	-0.09	0.10	-0.01	0.02
	18	0.01	0.02	-0.20	-0.01	-0.04	-0.01	-0.06	-0.02
	19	0.01	0.06	-0.31	0.18	-0.26	0.12	0.01	-0.02
	20	0.05	0.01	0.13	-0.30*	0.20	0.22	-0.08	0.20

*$p < .05$

**$p < .01$

Table 29

F-Values for Major Areas of Graduate Study Comparisons

Type of Evaluation		Graduate Major 1 F-Value	Graduate Major 2 F-Value
Context		0.93	0.70
Input		0.77	0.89
Process		0.19	0.54
Product		1.01	1.60
Evaluation Tasks			
CONTEXT	1	0.53	0.38
	2	1.76	0.88
	3	0.78	0.36
	4	2.79*	2.52*
	5	0.45	0.56
INPUT	6	1.27	1.69
	7	1.28	1.20
	8	0.42	0.64
	9	0.43	0.53
	10	0.28	1.07
PROCESS	11	0.29	1.90
	12	1.02	0.54
	13	0.29	0.64
	14	1.11	1.48
	15	0.71	1.33

Evaluation Tasks		Graduate Major 1 F-Value	Graduate Major 2 F-Value
PRODUCT	16	0.59	1.69
	17	1.09	1.42
	18	0.81	1.28
	19	1.23	1.06
	20	0.39	1.14

*$p < .05$

APPROVAL SHEET

The dissertation submitted by Michael Palmisano has been read and approved by the following committee:

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

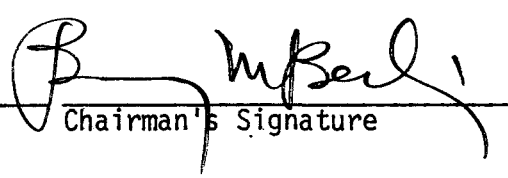
Dr. Robert C. Cienkus
Associate Professor, Department of
Curriculum and Instruction, Loyola

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

The final copies have been examined by the chairman 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 Education.

12/5/80
Date


Chairman's Signature