DOCUMENT RESUME

ED 076 682 TM 002 690

AUTHOR Hinkle, Dennis E.
TITLE Methodological Consideration in Evaluating Year-Round Educational Programs.
PUB DATE 1 Mar 73
NOTE 36p.; Paper presented at symposium of annual meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; *Models; *Program Evaluation; Research Design; Speeches; *Year Round Schools
IDENTIFIERS CIPP Evaluation Model

ABSTRACT

The evaluation methodologies used in the 1960s to evaluate various ESEA programs were shown to be inappropriate and inadequate for the evaluation of year-round educational programs. Due to their complexity and the various levels of decision-making, YRE programs require consideration of contemporary evaluation methodology. Due consideration is the use of the CIPP Evaluation Model in the development of the evaluation design. Several of the thirty steps of the CIPP Model are particularly important in the evaluation of YRE programs--the identification of the various levels of decision-makers, the writing of objectives stating performance criteria, the determination of the value of each objective, and subsequently, the determination of the priority of each objective. When each of these steps are considered in context with the remaining steps, the resultant evaluation design would provide the various levels of decision-makers with the appropriate information at the proper time in order to make responsible decisions regarding the effects of YRE programs. Responsible decision-making based upon the availability of appropriate information should be the goal of all evaluation efforts. (Author/KM)

METHODOLOGICAL CONSIDERATION IN EVALUATING YEAR-ROUND EDUCATIONAL PROGRAMS

Dennis E. Hinkle
Virginia Polytechnic Institute
and State University

Multi-Disciplinary Symposium on the
Possible Effects of Year-Round Education
on Community, School, Family, and Child

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

March 1, 1973

Introduction

Year-round education has become an increasingly prominent innovation in American education, public and private. This prominence is illustrated in the current literature which reflects the interest in year-round education as a means for improving the educational opportunities of children as well as for effectively utilizing the many resources of public school systems throughout the United States. In a time of overcrowded conditions in the schools coupled with the pressing financial conditions and defeated bond issues, year-round education is being proposed as an alternative to the traditional nine-month school calendar. The interest in year-round education is also reflected by the number of feasibility studies presently being conducted by several states and local school districts, the implementation of year-round programs in several states, legislation and financial support for such efforts in several states, and the recent formation of the National Council for Year-Round Education which seems approximately 1000 educators and lay citizens.

An outgrowth of this interest in year-round education is the need for data concerning the pros and cons of implementing a year-round program. The literature to date is filled with debates regarding the advantages and disadvantages, but there is little substantive data to support the arguments. These data are needed not only by teachers, administrators and school board members, but also by medical, legal, social and governmental professionals; businesses (such as moving and real estate); social and community organizations; parents; and other taxpayers.

The process of providing such data should not be viewed lightly. Evaluating any educational program, especially of this magnitude, requires considerable time and effort. In this paper, methodological considerations in developing a design for evaluating year-round educational programs will be discussed. The basis for this

discussion will be the Stufflebeam CIPP Evaluation Model (10). However, before the CIPP Model is defined and illustrated, the evaluation process, specifically related to education, will be discussed.

Educational Evaluation - What is it?

Evaluation in education received considerable attention during the past decade as a result of the bonanza of federal aid to education during the mid 1960's. These federal monies brought with them, however, the stipulation that all programs funded with these monies must be evaluated, i.e., a pit came with every plum. The purpose of these evaluations was to provide information that would guide future thinking and action in support of education. Legislators along with other laymen and professional educators were "seeking to understand more fully the relations between the various 'inputs' into [the] schools and the progress of education (23:13)."

Although formal evaluation was emphasized with the advent of increased federal aid to education, informal evaluation of educational programs and methodologies has been continuous process. The purpose of such evaluations have been the following:

1) to add to the substantial knowledge of educational processes;
2) to provide information in order to adjust, discard or otherwise change the application of an on-going educational process;
3) to provide justification for political-social-economic action relating to education;
4) to provide instruments which may be used to carry information on the success of the process to the educational community; and
5) to create a production (usually of paper) which can move through educational bureaucratic systems and thus keep these systems operative (5:15).

These five purposes are not mutually exclusive and do not necessarily operate in a discrete fashion, i.e., an evaluation of an education program and/or methodology can have more than one purpose. Also, when the purpose of the evaluation is to create a production that moves through the bureaucracy to keep it operative (purpose #5), considerable caution should be used. It is a recognized fact that evaluation reports are necessary for the proper functioning of the different decision-making groups within the bureaucracy. However, "a careful distinction must be made between required evaluation which is necessary and has an effect on operations and decisions, and that which only serves the life function of the bureaucracy itself. The first needs improvement; the second needs to disappear (5:17)."

Another need for evaluation arises when innovation and change in educational programs and/or methodologies proceed without an appreciably relevant theoretical basis or without careful planning. The resultant of such action dictates the need for a thorough evaluation procedure, i.e., these trial and error programs can only be rationalized through evaluation. Many times

> . . .pressure for innovation is frequently so great that change is introduced for its own sake with no adequate basis for hypothesizing improvement as a result. Empirical validation through evaluation becomes increasingly important under these circumstances (12:2).

Educational Evaluation Defined

The preceding discussion suggests that relevant information from an educational program would be gathered, compiled and interpreted in the evaluative process. The specific information gathered would be determined by the purposes of the evaluation in light of the objectives of the program. It was then assumed that this information would be provided for and used by those in positions of responsibility to make the necessary decisions relative to the objectives of the program evaluated. Thus, educational evaluation can be operationally defined as the process of providing information for the purpose of decision-making.

Expanding this definition in the context of education, the role of evaluation is to assist in the development and construction of new curricula methods and materials, the redevelopment and improvement of existing methods and materials, and/or the prediction of student academic achievement. The goal of evaluation is to obtain and provide information for decision-making relevant to the selection, adoption, support, and worth of educational materials and activities (6). The procedures in any evaluation effort consist of two basic steps. The first of these is to establish a set of descriptive, appraisal-related contexts or categories that appropriately order the particular curricular phenomena under study; the second is to establish a set of specific normative rules and procedures that make possible the appraisal of the curricular rationales and practices (24).

Present Approaches to Meeting the Evaluation Requirement.

Previously mentioned was the fact that the bonanza of federal aid to education in the late 1960's brought with it explicit evaluation requirements. Even though evaluation was not a new phenomenon in education, guidelines for evaluations were. Prior to this time, many evaluations were poorly planned and executed, and the results offered little service relative to decision-making. The methods often used in many of these evaluation efforts have been satirically described by Wolf (25) in his "5-C" Model. The five C's stand for cosmetic, cardiac, colloquial, curricular and computational. Three of these methods have particular relevance as they caricaturize many of the above-mentioned efforts in evaluation; they are presented below.

Cosmetic Method

This method is easily applied. Essentially, it involves taking a cursory look at a program and deciding if it looks good. Some of the things worth noting about a program when using this method include whether: Students look busy and involved, student projects emanating from the program can be easily and attractively displayed on bulletin boards, and one can easily develop an assembly or PTA

presentation based on activities of the program. When using the cosmetic method, one need not be concerned about objectives or gathering evidence about student learning. All such questions can be easily dealt with by showing an inquiring person the program in action and saying, "Look at all the wonderful things that are happening here. Who needs any more evidence to know we're doing a good job!"

Cardiac Method

The cardiac method is often used in conjunction with a systematic empirical approach. The use of planned evaluation procedures often results in showing that students enrolled in a new program learn no more than students in a conventional program, or that the new program did not attain its objectives. This can often present a dilemma since one always wants to claim beneficial results for a new program. The cardiac method resolves this dilemma. All one must do is dismiss the data and believe in his heart that the new program is indeed a good one. This method is quite similar to the use of "subclinical findings" in medical research.

Colloquial Method

This method is somewhat easier to apply than the cardiac method. Social psychological research has demonstrated that decisions arrived at by a group will achieve greater acceptance than decisions arrived at by an individual. This finding is the basis of the colloquial method. In applying this method, one need merely assemble a group of people who have been associated with a particular program to discuss its effectiveness. After a brief discussion, the group will usually conclude that the program has been indeed successful. This conclusion can then be transmitted to funding agencies and other school personnel. It is unlikely that such evaluations will be challenged since they have been arrived at by a group (25: 107-108).

Many times evaluation efforts such as the ones satirically described above lead to

a) inconclusive results;

b) evaluation reports which have no effect on administrative decisions, either because of bad timing or lack of relevance, or both;

c) lack of appreciation of the roles which evaluative activity can and should play in the many-factored war on social and educational problems (12:3).

It is very doubtful whether the results of such evaluators would be of much use to anyone responsible for the progress of the educational program. However, results such as these "are likely to fit well into the conventional schoolman's

stereotype of evaluation: something required from on high that takes time and pain to produce but which has very little significance for action (9:127)."

The micro-utility of this type of evaluation report dictated an urgent need for new and improved evaluation methodology. Such methodology is necessary for making evaluation reports concise yet thorough, but more importantly, useful for decision-making purposes. When this need became apparent, it was discovered that personnel trained in evaluation, evaluation designs and instruments, and overall experience in evaluation were essentially all lacking. Educators faced with deadlines for evaluation reports turned to the educational research methodologists for help in developing more adequate evaluation methodology. However,

> . . . the efforts of educational research methodologists to respond to these needs errupted in controversy when factions recommended opposing approaches for accomplishing the needed evaluation (18:121).

This controversy obviously did not resolve the pressing need for new and improved evaluation methodology; rather, the urgent need for evaluating current evaluation methodology was further emphasized.

The necessary and logical first step taken in the evaluation of present-day evaluation methodologies was the determination of what, in fact, the purposes and the general methodologies were versus what they should be. For the most part, it was found that they were summative (15) in nature. Summative evaluation was defined as

> . . . terminal evaluation concerned with the comparative worth of effectiveness of competing programs. The results of summative evaluation are not intended to serve directly in the revision, improvement or formation of a program; rather they are gathered for use in making decisions about support and adoption (8:12).

It was also found that many evaluators used the Campbell-Stanley chapter on experimental design in the Handbook on Research in Teaching (1) as a model for evaluation designers to follow as they made an effort to devise generalizable

evaluation designs. The "evaluators . . . noted, with envy, the tremendous help this chapter [provided] the researcher who [was] in need of . . . [designs for]. . . experimental and quasi-experimental research (26)." The crucial problem underlying this usage of the designs in this chapter for evaluation purposes was that

> . . . evaluators as a group [were] erudite enough to realize that experimental design [i.e., Campbell-Stanley] per se is generally inapplicable in attempts to solve evaluation problems, but the intrinsic appeal of rigor and parsimony inherent in experimental design still [seemed] to influence evaluators' efforts to come to grips with their own design problems (26:3).

In using such experimental designs, it is necessary for the evaluator to attempt to control extraneous variables while manipulating experimental variables. However, many present-day decision situations are much too complex to be dealt with in this traditional experimental variable-control variable manner. The evaluator must recognize that, in many situations, he does not exercise experimental control over the situation, nor does he manipulate it in any way. He must accept it as it is and as it evolves, and monitor the total situation by focusing his most sensitive noninterventionist data collection techniques on the most crucial aspects of the project. Such evaluations are multivariate and require the evaluator to focus his attention on theoretically important variables while remaining alert to any other important variables which were not, and could not have been specified at the initiation of the project (21). These situations dictate the need for not only end-of-the-project evaluation, but also for continuous monitoring throughout the project. Thus present summative methodologies must be supplemented with formative (15) ones; those which provide diagnostic or process data during the development and operation of the project or educational program.

Development of Models for Evaluation

The need for more adequate evaluation methodologies, e.g., formative methodologies to supplement existing summative ones, resulted in many serious

investigations by prominent research and evaluation methodologists into the problems underlying this need. They began, as was previously mentioned, by determining what evaluation was, as it existed then, and what it should have been. They then concentrated on what remained to be done in order to make evaluation what it should be, namely, the process of providing valid, reliable and timely information for the purpose of decision-making. The result of these investigations was the development of several evaluation models. Operationally defined, an evaluation model is a set of generalizable steps which can be followed in the development of an efficient and effective evaluation design. The CIPP Evaluation Model developed by Stufflebeam at the Ohio State University Evaluation Center is one such model. This model contains twenty-two steps for developing a design for an evaluation study.

It is important at this juncture to emphasize the distinction between, as well as the relationships between, an evaluation model, an evaluation design, and an experimental (or research) design. For this purpose, these terms will be defined as follows:

1) An evaluation model is a set of generalizable steps from which an evaluation design is developed.

2) An evaluation design is a set of specific procedures to be employed in accomplishing the purpose and/or objectives of a particular evaluation.

3) An experimental (in the context of research) design is a preconceived plan of systematically varying one or more variables for the purpose of determining the effects of this variation while exercising control over the remaining sources of variation. (In the concept of this paper, this term is used to include the experimental and quasi-experimental designs, i.e., Campbell-Stanley, as well as other more advanced designs.)

Since no two decision situations requiring evaluation are exactly the same, a different evaluation design is developed from the model for each situation. Furthermore, one or more experimental designs may be included in an evaluation design depending upon the objectives of the evaluation, but experimental design

must not be interpreted as being synonymous with evaluation design. Likewise, the term, evaluation design, should not be interpreted as being synonymous with evaluation model.

Some Theoretical Considerations

Following the development of the models, evaluation methodologists have attended to practical and theoretical considerations underlying the application of the models. Glass (7) identified five such problems; the solutions of which he said "could substantially advance the theory and application of evaluation (7:1)." These five problems are listed below and then discussed individually:

1) The validity of Judgment
2) Generality-Specificity of Evaluation Data
3) Models of Summative Decision-Making
4) Priorities on Evaluation Data
5) Units of Observation

The Validity of Judgment. "Personal value-commitments, educational aims, goals, objectives, priorities, perceived norms, and standards--in one form of expression or another--are judgment data (17:181-182)." The use of such judgment data has been legitimized by Stake (17) and Scriven (15), and evaluation has profited by their use. Evaluators are now more willing to "exploit the incomparable ability of humans to collect, store, and integrate information and to render judgments (7:1)." Unfortunately, up to this time, evaluators have depended upon psychometric theory for methods of measuring judge-agreement and describing judgmental data; they have, however, done little beyond

> . . . arguing that judgments are valuable data and that psychometric theory can help describe them . . . . Evaluators presently have no methodology for assessing the pre-eminant quality of judgments, namely their validity (7:1).

Therefore, in some instances, it may be necessary to consider only the utility of judgments; while in others, the validity of these judgments must also be considered.

Generality-Specificity of Evaluation Data. Present-day educational evaluation is an extremely complex undertaking; there are innumerable levels of specificity to be considered. It is, therefore, important that evaluators consider the entire forest while examining the individual trees. Many times, evaluators become so engrossed in the fine details of the data that they lose the ability to generalize, intuitively or otherwise, about the overall program. "Evaluation methodologists have yet to suggest any means for determining whether one should observe general or specific phenomena. Without the guidance of explicit methodology, too many efforts become either absurdly reductionistic or worthlessly global (7:5)."

Models of Summative Decision-Making. Many times, an evaluation design may require only summative methodologies which would "involve the measurement of competing programs of performance or goal scales and the integration of the data into a conclusion of superiority for one program (7:5)." A serious problem sometimes results if the evaluator fails to take the next step, that of integrating the information into a summative judgment. Consider a comparative situation in which several measuring instruments are used to assess the value of two programs; one program is inexpensive and traditional and the other, innovative but expensive. If the traditional program is superior to the innovative program on two of the four measuring scales and inferior on the other two, then a judgment must be rendered. The question as to which of these scales is more important must be raised and thus each of the measuring instruments must be given a ranking or weighting based upon the goals of the program. This weighting of the scales requires a summative judgment on the part of the evaluator. Assuming that the problem of weighting measurement scales can be adequately solved, a second problem appears.

Assume as a result of the weighting procedure, the innovative program is shown to be superior. Now the question must be raised as to whether the superiority of the innovative program adequately compensates for the additional cost. Again a judgment must be rendered. This now becomes a problem of "utility (2)." With regard to this latter problem,

> . . . management science has recently adopted Bayesian decision models for application in business. These models are a meld of information and human judgment into decision-making situations (16)

A thorough investigation into these models may significantly advance evaluation methodology in integrating objective information and judgment into summative decisions (7:6).

Priorities on Evaluation Data. "Evaluation methodologiests have adopted the notion (explicityly and by example) that practically all data merit collection and analysis (7:8)." The surprising and impressive fact is the number and diversity of variables considered worthy of measurement. Even though present-day computer systems are capable of handling vast quantities of data, there is a realistic limit to the amount of data that an evaluator can collect, analyze and interpret. For example, an evaluation design may call for a k-factorial ANOVA in the analysis procedure. If the data can be feasibly collected, a computer program can be written (if not already available) for the analysis; but there is a not-too-remote possibility that a resultant, significant fourth order interaction cannot be interpreted. Therefore, a judgment must be made by the evaluator concerning the data that are most relevant to the decision situation, i.e., priorities in data collection must be determined.

In spite of the widespread curiousity in other disciplines concerning "cost-benefit analysis, cost-utility analysis, program planning and budgeting, etc., such methods have influenced education only at a macro-economic level (7:11-12)."

influenced education only at a macro-economic level (7:11-12)." In other words, evaluators in education have shown little concern with regard to the assessment of costs and the relationships of costs to utilities. A temporarily workable methodology for estimating on the collection of data is presented below:

1) the costs of gathering different data;

2) estimates of the prior probabilities that each alternative embodies in a decision will be supported by data--if they were to be gathered; and

3) the costs of implementing each of the alternatives of the decision (7:8).

Units of Observation. A basic assumption underlying the experimental designs, i.e., a la Campbell-Stanley or more advanced designs, is either the random selection of, or random assignment of individual subjects to a unit or sample. Meeting this assumption enhances statistical as well as intuitive generalizing from the observed unit to other units. However, in educational evaluation, practical considerations often inhibit the meeting of this assumption; the unit of observation is generally a school or a class instead of individual students. To do otherwise "may require as many as 200 classes and more than 5,000 students. . . [which] . . . for most evaluation efforts . . . is prohibitively expensive (7:8)." At the present time no solution to this problem of cost has been advanced probably no solution appears possible. It appears likely that before feasible methods are available for handling these problems of units of observation with limited resources, the nature of school instruction will change from teaching groups to teaching individual students. Thus, the changing system may make the problem irrelevant (7:12-13).

The preceding five problems do not exhaust the possibilities of additional considerations in evaluation methodology. Others include 1) methods for justifying

goal scales, i.e., "the activity which distinguishes evaluation from accreditation (7:12)," 2) the effective utilization of behavioral objectives in the specification of the performance criteria to be measured, and 3) the application and use of multivariate statistical techniques in the evaluation designs. However,

> . . . like any complex human fabrication, evaluation methodology has no real genotype; its only genotype is a plan for its future growth in the minds of its builders. The architects of evaluation methodology must attend to planning its future as well as fostering its present growth (7:12).

The CIPP Evaluation Model

Introduction

The need for new and better evaluation methodologies was discussed in the previous section, along with a rationale for considering the practical aspects as well as several theoretical developments. That discussion is particularly important when considering the development of a design for evaluating year-round education (YRE) programs. The magnitude of such a design eliminates the possibility of utilizing the previously mentioned methodologies. The research methodologies of the educational researcher-turned-evaluator are inappropriate and the cosmetic and cardiac methodologies of the public school personnel are totally unacceptable.

The development of an efficient and effective design for evaluating YRE programs would be facilitated through the application of the CIPP Evaluation Model. The twenty-two steps in the original model (expanded for explicitness to thirty steps in a expanded model (10)) provide the framework for such development. The acronym CIPP refers to four types of evaluation strategies: Context, Input,

Process and Product. Each of these strategies corresponds to a specific decision situation encountered by the evaluator and/or decision-maker. The rationale underlying each of these four types of evaluation strategies is summarized in the following chain of reasoning:

1. the quality of programs depends upon the quality of decision in and about the program;

2. the quality of decisions depends upon decision-makers' abilities to identify the alternatives which comprise decision situations and to make sound judgments of these alternatives;

3. making sound judgments requires timely access to valid and reliable information pertaining to the alternatives;

4. the availability of such information requires systematic means to provide it; and

5. the processes necessary for providing this information for decision-making collectively comprise the concept of evaluation (19:6).

Given this chain of reasoning, stufflebaeam defined evaluation as "the provision of information through formal means, such as criteria, measurement, and statistics, to serve as rational bases for making judgments in decision situations (19:6)".

The general logic and cyclic nature of the CIPP Evaluation Model is shown in Figure 1. The figure illustrates that program operations are evaluated to influence decisions which result in actions designed to improve program operations, which in turn are evaluated, etc. Implicit in this logic are four tenets which form the basis for the model. They are:

1. The purpose of evaluation is to provide information for decision-making and to evaluate; therefore, it is necessary to know the decisions to be served.

2. Since evaluation studies should answer questions posed by decision-makers, designs for such studies should satisfy criteria both of scientific adequacy and of practical utility. Specifically, evaluation studies should meet criteria of validity, reliability, and objectivity, as should any scientific study. But, to be useful, they should also provide concise and meaninful information.

THE RELATIONSHIP OF EVALUATION TO DECISION-MAKING

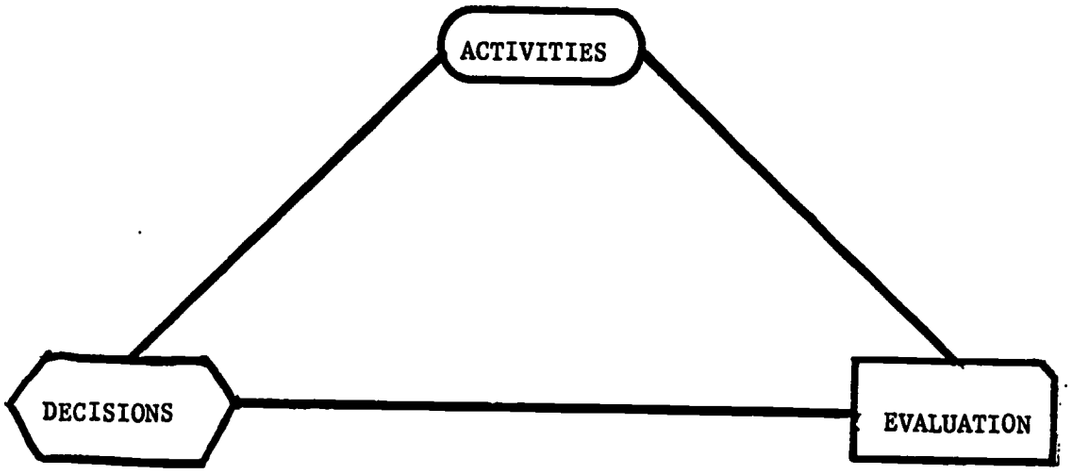


Figure 1

If extensive scientific rigor is employed in an evaluation study and results in irrelevant, meaningless and useless information, then the employment of such rigor is wastefully absurd. On the other hand, if the study is conducted haphazardly with little scientific rigor, useless information may also result. Thus, evaluation designs must satisfy both criteria, i.e., scientific adequacy and practical utility.

3. Since different types of decisions require different types of evaluation, a generalizable and efficient model of evaluation should be based upon a generalizable and parsimonious conceptualization of types of decisions and evaluation.

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 study (3:210).

The above rationale and basic tenets, which underlie the CIPP Evaluation Model, serve as an introduction to the following discussion of the Model.

The CIPP Model

In the CIPP Model, the type of evaluation strategy (context,input, process or product) to be carried out is dependent upon the type of decision situation in which the evaluators and decision-makers are involved. More specifically, there is a one-to-one relationship between the type of decision to be served and the evaluation strategy to be used. Generally, four types of decision situations occur in education: 1) planning, 2) structuring, 3) implementing and 4) recycling decision.

1. Planning decisions specify that changes are needed in a program. The need for such decisions arises from one of two sources: (a) awareness of a lack of agreement between what the program was intended to be and what it actually is (congruence evaluation), and (b) awareness of lack of agreement between what the program could become and what it is likely to become (contingency evaluation). In either case, a decision to change the intentions and/or the actualities in a program could be made.

2. Structuring decisions specify operationally defined objectives; general program strategies; and method, personnel, facilities, budget, schedule, organization, and context--for use in affecting desired changes. These decisions arise from three sources: (a) awareness of planning decisions which specify that the program is to be changed, (b) awareness that there are alternative means available to bring about the specified changes, and (c) awareness of the relative strengths and weaknesses of the available alternatives. Given these three conditions, an action plan to achieve the desired changes in a program can be structured.

3. Implementing decisions are those used in carrying through the action plan. These decisions arise from two sources: (a) knowledge of the procedural specifications, and (b) continuing knowledge of the relationship between procedural specifications and the actual procedures. These two kinds of information aid in process control.

4. Recycling decisions are those used in determining the relation of outcomes to objectives and in determining whether to continue, terminate, evolve, or drastically modify the activity. These decisions require information about: (a) specified outcomes, (b) actual outcomes, and (c) relation of the outcomes to the context within which the activity exists (3:213).

In turn, each of these decisions demands an accompanying evaluation process. Context evaluation provides information for planning decisions; input evaluation--structuring decisions; process evaluation--implementing decisions; and product evaluation--recycling decisions. In other words, context evaluation indicates the need for change within the program; input evaluation helps determine how the change is to be affected; process evaluation aids in the day-to-day implementation of the change; and product evaluation identifies the outcomes of the change effort.

Developing and Evaluation Design

As mentioned above, the type of evaluation strategy (i.e., context, input, process, product) is dependent upon the decision situation to be served. Once the strategy is determined, the evaluator must begin the difficult and laborious task of developing the design for the implementation of this strategy. In designing the strategy, the evaluator must be continually cognizant of the decision points in the evaluation. He must also have available alternatives so that the total set of decisions generated "will yield information which will meet the specified evaluation criterion of validity, reliability, pervasiveness, timeliness, and credibility (18:3)." These criteria are defined by the following questions:

1. Validity - is the information what the decision-maker needs?
2. Reliability - is the information reproducible?
3. Pervasiveness - does the information reach all the decision-makers who need it?

4. Timeliness - is the information available when the decision-makers need it?

5. Credibility - is the information trusted by the decision-makers and those he must serve (19:6)?

In general, the development of an evaluation design involves the preparation of a set of decision situations which occur at critical periods during the planning, structuring and implementation stages of the project or program. The decisions made at these critical periods determine the course(s) of action that must be taken in order to achieve the specified program objectives. Three procedures are generally undertaken by an evaluator in developing an evaluation design. The evaluator must first identify the specific evaluation objective(s) to be achieved through the implementation of the evaluation design (Note the distinction between program objectives and evaluation objectives). For example, in attempting to meet the general program objective of increasing the achievement level of inner-city children, two evaluation objectives in the evaluation design may be to determine if the newly-purchased second grade reading materials and the new method of teaching fourth grade arithmetic have been effective (as defined by appropriate criteria). Secondly, the evaluator must identify and define the decision situations in the procedure for achieving the evaluation objective. Thirdly, the evaluator must be prepared to make a choice among the available alternatives for each identified decision situation. Thus the completed evaluation design would contain a set of decisions as to how the evaluation is to be conducted and what instruments are to be used.

One of the previously mentioned, basic tenets which should underlie any evaluation model used in developing an evaluation design was that the model should contain a single set of generalizable steps. In keeping with this tenet, Stufflebeam proposed twenty-two generalizable steps which reflect decision situations

common to most evaluation designs (10). Hinkle (10) expanded the model to thirty steps; this expansion not only explicated many of the original steps, but also included many of the theoretical considerations previously mentioned. These thirty steps or decision situations are grouped into six general headings which basically outline the procedures for developing an evaluation design. They are: 1) focusing the evaluation, 2) collection of the information, 3) organization of the information, 4) analysis of the information, 5) reporting the information and 6) administration of the evaluation. These six headings and thirty steps are found in Figure 2. It must be emphasized that the figure illustrates only a general guide for developing context, input, process or product evaluation designs. In developing a design for each type of evaluation strategy, these twenty-two steps should be considered, but each design would, in fact, be developed de novo dependent upon its objectives. In the following paragraphs, the sixteen steps under the first two headings will be discussed in detail. The remaining fourteen steps under the last four headings are administrative in nature and generally self-explanatory; thus they will not be further explicated.

Focusing the Evaluation. The general purpose of the four steps grouped under this heading is to determine the goals for the evaluation and to define the policies within which the evaluation will be conducted. These four steps are extremely crucial and should not be taken lightly. Poorly performed, these steps could result in an evaluation design that fails to meet the previously mentioned criteria of validity, reliability, pervasiveness, timeliness, and credibility. Failure to meet these criteria could result in inadequate information for decision-making purposes.

The first step in the model (Figure 2) may be considered rather obvious, but very important; the overall purpose of the evaluation should be defined. When properly defined, there will be less tendency to proceed tangentially during the actual implementation of the design. The second step is to identify the major

Figure 2

Procedures for Developing Evaluation Designs

Major components for Context, Input, Process or Product Evaluation

Focusing the Evaluation

1. Define the overall purpose of the evaluation.
2. Identify the major level(s) of decision-making to be served, e.g., local, state, and national.
3. Determine the objectives of the procedures (context and input evaluation) and/or program (process and product evaluation).
4. Define the contexts or categories that appropriately order the phenomena under investigation, i.e., the focus of the evaluation.
5. Write the objectives in behavioral terms (when appropriate) which explicitly specify the performance criteria.
6. For each level of decision-making, describe the decision situations in terms of their focus, timing and composition of alternatives.
7. Consider the VALUE of the objectives.
8. Determine the PRIORITY of the objectives.
9. Identify potential outcomes of the evaluation which do not deal with those objectives specified above.
10. Define the policies within which the evaluation must operate.

Collection of Information

1. Determine the information needs, i.e., the performance criteria.
2. Specify the source of the information to be collected.
3. Specify the instruments and methods for collecting information.
4. Specify the standards to be used in the analysis.
5. Specify the sampling procedure to be employed.
6. Specify the conditions and schedule for information collection.

Organization of Information

1. Specify a format for the information which is to be collected.
2. Specify a means for coding, organizing, storing, and retrieving information.

Analysis of Information

1. Specify the analytical procedures to be employed.
2. Specify a means for performing the analysis.

Reporting of Information

1. Define the audiences for the evaluation reports.
2. Specify means for providing information to the audiences.
3. Specify the format for evaluation reports and/or reporting sessions.
4. Schedule the reporting of information.

Administration of the Evaluation

1. Summarize the evaluation schedule.
2. Define staff and resource requirements and plans for meeting these requirements.
3. Specify means for meeting policy requirements for conduct of the evaluation.
4. Evaluate the potential of the evaluation design for providing information which is valid, reliable, credible, and timely.
5. Specify and schedule means for periodic updating of the design.
6. Provide a budget for the total evaluation program.

levels and decision-making to be served. It is important to consider all relevant levels since each level may require different information and/or the information at a different time. In the third step, the overall purpose is further specified, in light of step two, through determining the objectives of the procedures (usually in context and input evaluation) and/or program (usually in process and product evaluation). In this step, the objectives need not be written in behavioral terms (when appropriate) but rather in more general terms. This leads directly into the fourth step, that of defining the contexts or categories that appropriately order the phenomena under investigation relative to the types of decisions to be made. In other words, the design variables and their interaction(s) would be specified, i.e., defining the decision situations in terms of <u>focus</u>.

Step five requires the above stated objective to be written in terms which explicitly specify performance criteria. The evaluation of any program, e.g. a YRE program, is no better than the information collected and the information collected is no better than the validity of the criteria measured.

The sixth step may be one of the most important and most difficult. It involves defining the decision situations to be served at each level of decision-making, specifying those responsible for making the decisions (teachers, principals, board of education members, state legislators, etc.), and identifying the type of decision to be make (appropriational, allocational, approval, continuation, etc.), i.e., the <u>locus</u> of the decision situation. In addition, the <u>timing</u> of the decision situations to be served and the <u>alternatives</u> which may reasonably be considered in reaching a decision must be determined. This will involve insuring that all the relevant cost and benefit information is available at the time when the decisions must be made.

The <u>VALUE</u> of the above objectives is to be considered in step seven. There are two bases for this value analysis which are not necessarily mutually exclusive.

The first involves a logical basis which checks the reasonableness of the selected objectives given a certain value position (11). For example, a behavioral objective written in step five may be concerned with increasing the achievement level of inner-city children, as measured by specified criteria. However, if the philosophy of the school program is that the increase in the level of achievement is a resultant of a positive change in the children's attitudes toward school and the learning process, the objective may need to be rewritten in terms of measuring an attitude change rather than a change in the level of achievement. The second basis is an empirical one. It involves an empirical analysis to see to what extent certain value positions are widely held among those individuals who will be directly involved in the program or project, i.e., to see how desirable certain objectives are perceived to be (13). In the former, the value position is assumed; in the latter, the value position is determined.

The purpose of step eight is to determine the PRIORITY of the objectives. Priorities on evaluation data were previously discussed; this discussion is pertinent in this step in the model since all the objectives have been written in behavioral terms which specify the performance criteria (i.e., evaluation data). In this discussion, a temporarily workable methodology was presented. At the risk of being redundant and in order to establish a framework for determining a priority of the objectives, this methodology is presented again. It involves the determination of

1. the costs of gathering different data;
2. estimates of the prior probabilities that each alternative embodies in a decision will be supported by data--if they were to be collected; and
3. the costs of implementing each of the alternatives of the decision.

If there were

> . . . unlimited resources or if all the objectives were attainable in the time available, it would not be important to specify the priorities. In actuality, it is important not only to choose the objectives to be pursued but to allocate scarce resources to each of the several objectives (17:184).

Thus this step might result in a delimitation of the purpose of the evaluation defined in the initial step of the procedures.

In step nine, potential outcomes of the evaluation which do not deal specifically with the stated objectives are identified. In the discussion in the previous section, formative evaluation (as distinct from summative evaluation) was defined as the process of providing evaluation data during the development and operation of the program under consideration. Identifying potential outcomes will assist the evaluator in remaining alert during the evaluation to any unanticipated but significant events.

The final step is to determine the policies in which the evaluation must operate. Basically, this involves decisions to be made concerning 1) those who will conduct the evaluation (from within or from without), 2) the accessibility of the evaluation team to the data, and 3) those who are to receive the reports.

Collection of the Information - The six steps listed under this heading are closely related to step five above, the stating of performance criteria. Step one is a potential recycle step which determines whether the information needs, i.e., the performance criteria for each decision, have been adequately specified in step five above. If not, those responsible for the design must rewrite the behavioral objectives, this time insuring that the performance criteria have been adequately specified.

Steps two and three indicate that the source of the information is defined (students, teachers, administrators, parents, etc.) followed by the specification

of the appropriate instruments and methods for collecting the information. Step four follows and involves the specification of standards to be used in the analysis. The word "Standard" used here refers to

> . . . a desired level or quality of something as cited by an authority. Standards answer the question, "How much is good?" . . . and can be considered . . . another form of objective: those seen by outside authority--figures who know little or nothing about the specific program being evaluated but whose advice is relevant to many programs (17:185).

The last two steps involve the determination of the sampling procedures to be employed and the development of a master schedule for data collection. This schedule should reflect the interrelations between the sampling procedure(s), the measuring instruments, and the overall time schedule (i.e., specified in the sixth heading, Administration of the Evaluation).

Up to this point, educational evaluation has been defined and explicated. The inadequacies of former methodologies have been discussed along with the CIPP Evaluation Model. It was indicated that through the utilization of the CIPP Model, adequate evaluation designs would potentially be developed. In this way, the resulting evaluation provide useful information for decision-making purposes. In the final section, the development of an design for evaluating YRE programs will be considered through the use of the CIPP Evaluation Model.

The CIPP Model for Evaluating YRE Programs

The four types of evaluation strategies, Context, Input, Process and Product, considered in the CIPP Evaluation Model were previously discussed in relation to the respective decision situation in which the evaluator and/or decision maker is involved, i.e., planning, structuring, implementing or recycling. In developing a design for evaluating an ongoing YRE program it is assumed that the initial context and input evaluations were conducted and that the appropriate planning and structuring decisions were made. In other words, the result of these planning and structuring decisions was the decision to implementation of a YRE program.

As the YRE program is implemented, there is a need to make daily decisions regarding the program. A process evaluation, properly designed and executed would provide the data necessary for making such decisions (i.e. implementing decisions). However, a process evaluation resulting in day-to-day changes is not sufficient in these days when educational accountability is in vogue; a product evaluation, at the end of some specified time, is necessary. (Note the similarity between the terms process and product evaluation, a la Stufflebeam and the terms formative and summative evaluation, a la Scriven).

The remainder of this paper will be concerned with the product evaluation of YRE programs; however, the reader should keep in mind the cyclic nature of evaluation as defined through the utilization of the CIPP Model. That is, a product evaluation is not the end -- the decisions resulting are recycling in nature. In other words, the decisions made at the product evaluation stage result in recycling decisions which not only look at the outcomes at that particular point in time, but also provide input for future planning, structuring and implementing decisions in future context, input and process evaluations.

As previously mentioned, the CIPP Model, when used in developing a context, input, process or product evaluation design, is a set of generalizable steps to be followed. While all of these steps are relevant to the development of the design and should be considered by the evaluator, certain setps are more important than others. This is particularly true for the sixteen steps listed under the "Focusing the Evaluation" and "Collection of the Information" (See Figure 2). For example, a product evaluation design for a YRE program, due to its magnitude and complexity, would give special attention to several of these steps listed under these headings.

Step A2, indentify the major levels of decision making, should be given careful consideration. YRE programs have implications regarding not only education, but also a wide range of other institutions in our modern and complex society. Each of these institutions, i.e. family, political, social, economic, etc., will or have been affected by the implementation of a YRE program and should be considered in the final decision-making process. Therefore, the evaluation design should identify these levels of decision-making in order to insure that the most appropriate information will be provided for these levels. An evaluation design for a YRE program should indentify the following levels of decision makers:

A. Educational

1. Students
2. Teachers
3. Building Administrators and Supervisors
4. Central Administrators and Supervisors
5. School Board Members
6. State Education Officials
7. Federal Education Officials

B. Family

1. Parents or Guardians

C. Economic Institutions

1. School Board Members
2. Local Government Officials
3. State Government Officials
4. Business Officials - regarding hiring practices, vacations, volume, etc.
   a. Chamber of Commerce
   b. Labor Unions
   c. Professional Associations

D. Religions and Social Institutions

1. Council of Churches
   a. Individual Church Councils
2. Board of Director of Various Social Organizations
   a. Adult Organizations
   b. Student Organizations
      1. intramural
      2. extramural

E. Political Institution

1. Local Government
2. State Government
3. Federal Government
4. Political Parties (Major and Minor)

The identification of the above levels of decision-making is difficult without concurrently considering steps A3 and A4, i.e. determining the objectives of the YRE program and the identification of the context or categories for evaluations. If A3 is adequately completed and the general objectives of the YRE program are stated, these objectives should define the contexts or categories which will provide the focus for the evaluation. The objectives stated for YRE programs have generally been educational in nature and have concerned with the following:

A. Year-round utilization of the educational facilities

1. buildings
2. library and media centers
3. equipment
4. textbooks and other materials

B. More efficient utilization of the facilities

1. flexible scheduling
2. innovative space utilization
   a. reduction of overcrowded conditions

C. Professional Development of Staff
   1. inservice training
   2. 12 month contracts
   3. retention of professional staff

D. Curriculum Development Opportunities

E. Attitude of Professional Staff

F. Attitude and Achievement of Students
   1. Individualization of Instruction
   2. Special Programs - academic and special interest

G. Year-round job placement opportunities

H. Family Life
   1. adjustment of schedules
   2. vacations

While the above are the most often cited contexts of objectives in the literature, other objectives relating to other social, economic and political institutions need to be considered. For example, consideration should be given to the effects of YRE programs upon the seasonal aspects of many businesses and professions, e.g., department store sales, resorts, medical establishments, employment opportunities, etc.

The objectives referred to above are written for the purpose of ordering the contexts for the evaluation and not specifically for stating the performance criteria. However, following the ordering of the contexts (i.e. focusing the evaluation), the objectives need to be expanded and stated in explicit terms which specify the performance criteria and thus make them amenable to evaluation. The importance of this step cannot be over-emphasized. One of the basic problems of past evaluation efforts discussed earlier was the lack of program objectives which explicitly state the performance criteria. While there is an on-going debate regarding the extent to which the objectives must be explicated, there is a need for stating the performance criteria. If this is not done during the initiation of the YRE program, it has to be done at the product evaluation stage.

Once the objectives have been stated in terms amenable to evaluation, the value and priority of the objectives need to be considered. It is unfortunately

commonplace for the evaluation of most educational programs to be focused upon student achievement scores, i.e. placing high value on student achievement scores due to the relative convenience of data collection, analysis and interpretation. However, these data are often over-emphasized at the expense less tangible, yet meaniful, data which require more effort to collect and are more difficult and less sophisticated to analyze and interpret. In these cases, the value of the objectives was ignored and high priority was placed upon the objectives which had easily accessible and assessible data.

In step A7 of the CIPP Model, the value of each objective is initially determined irregardless of the cost-effective data collection priorities. These priorities are subsequently determined independent of value of the objectives; this is step A8 of the CIPP Model. If an objective has a rather high value position in the design, and is also rather costly and time-consuming to evaluate, a decision would have to be made regarding the implementation of that portion of the evaluation strategy in light of alternate strategies. At other extremes, an objective with low value and high evaluative cost would probably not be considered, while an objective with high value and low evaluative cost would be ideal.

One of the concerns of this writer regarding the evaluation of YRE program to date is that once again the evaluators have tended to over-emphasize student achievement data and under-emphasize other relevant aspects of the program. Granted that the ultimate goal of any new and innovative program is the improvement of educational opportunities for the students, this goal should not be assessed only in terms of student achievement. Such assessment should be considered long-ranged with other variables considered over the shorter periods of the program. For example, YRE programs have great potential with regard to professional and curriculum development, flexibility in programs and scheduling, facility utilization, attitudes of professional staff and students, etc. Rather than just look at student achievement scores and certain cost factors in YRE programs attention

should be given to the value and priority of the objectives relating to these other variables.

The preceding discussion has involved several steps of the CIPP Model which are of primary importance in developing a design for evaluating YRE programs. It was mentioned that these steps along with the other of the thirty steps are not necessarily sequential and/or mutually exclusive. The sequencing of the steps is dependent upon the objective of the design as well as the objectives of the program being evaluated. Step F1 under the heading, "Administration of the Evaluation", requires a summary of the evaluation schedule. A suggested summary schedule for developing a design for evaluating YRE program is illustrated in Figure 3. This PERT chart depicts the interrelationship among the steps along a time continuum and places the total evaluation effort in perspective.

It should now be evident that using the CIPP Evaluation Model in developing a design for evaluating YRE or any other educational program is not a simple task. As the complexity of the program increases, so does the evaluation design for the program. However, in order to insure adequate and appropriate evaluation of complex programs, such as the YRE program, each of the steps in the CIPP Model needs to be considered. In this way, the proper information will be provided at the proper time to the proper people at the various level of decision-making so that proper decisions can be made. This is the goal of all evaluations.

Figure 3

PERT Chart Illustrating the Total Evaluation Schedule (F1

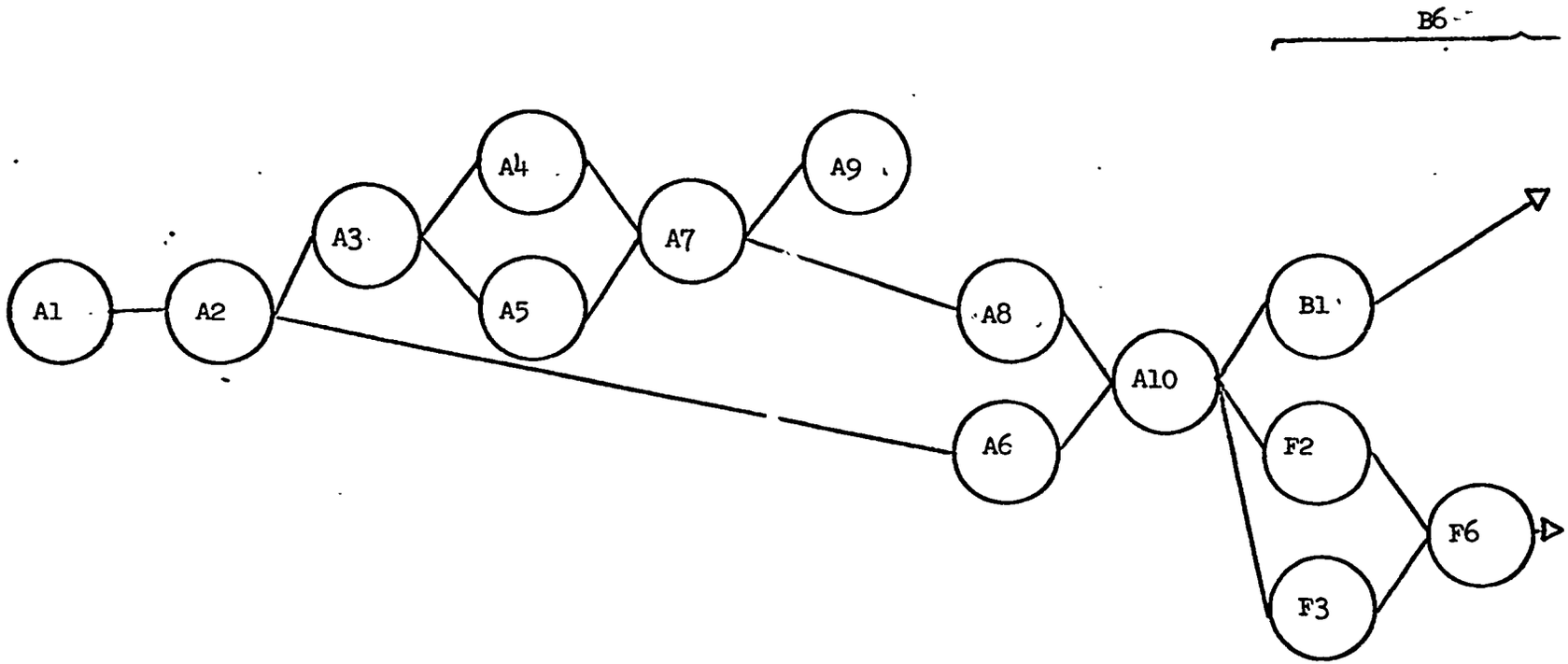
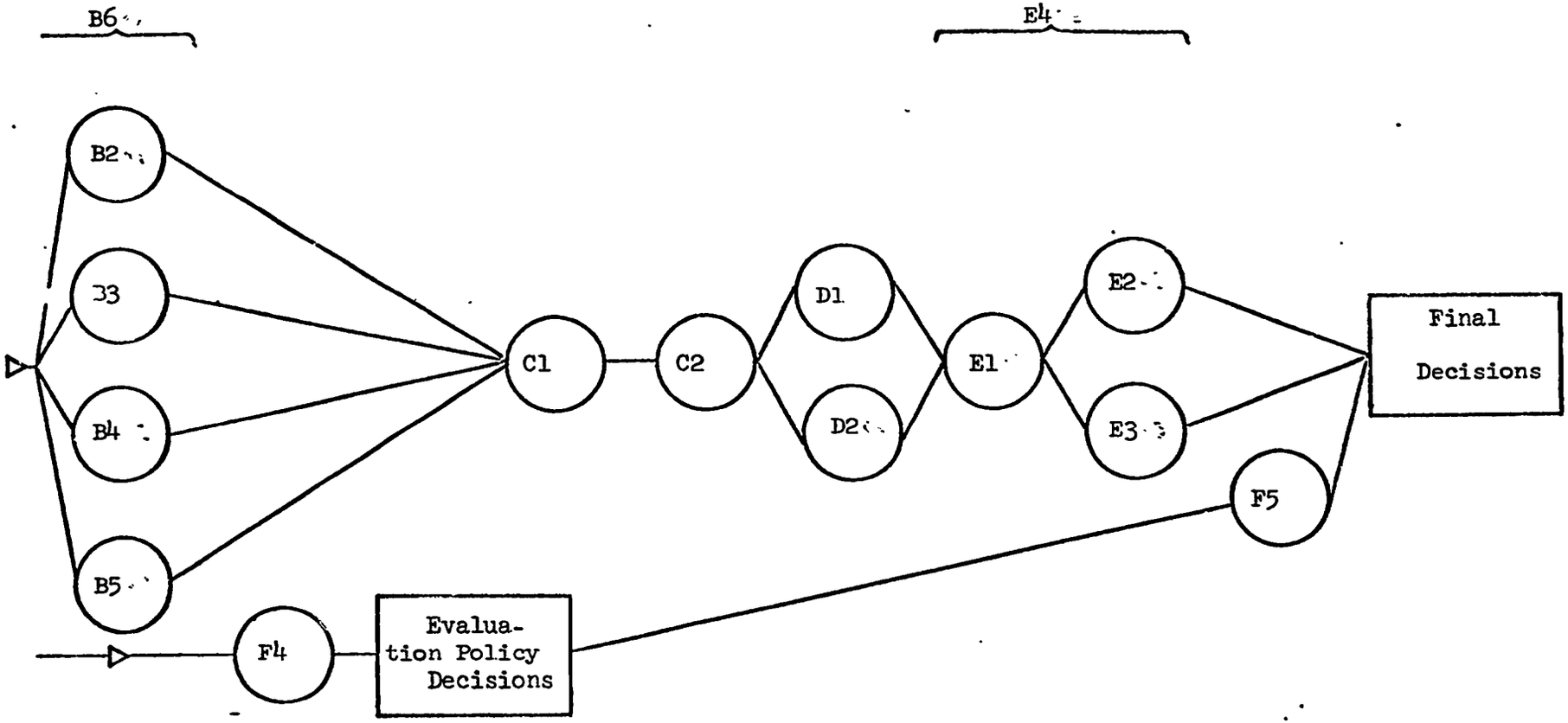


Figure 3 (continued)



Summary

The evaluation methodologies used during the middle and late 1960's in evaluating the various ESEA programs were shown to be inappropriate and inadequate for the evaluation of year-round educational programs. Due to their complexity and the various levels of decision-making, YRE programs require consideration of contemporary evaluation methodology. Due consideration is the use of the CIPP Evaluation Model in the development of the evaluation design. The implication was made that through the use of the CIPP Model, the evaluator could overcome many of the inadequacies of previous evaluation efforts. These inadequacies were characterized by insufficient and inappropriate information being available to decision-makers during the decision-making process.

Several of the thirty steps of the CIPP Model are of particular importance in the evaluation of YRE programs. They include the identification of the various levels of decision-makers, the writing of objective stating performance criteria, the determination of the value of each objective, and subsequently, the determination of the priority of each objective. When each of these steps are considered in context with the remaining steps, the resultant evaluation design would provide the various levels of decision-makers with the appropriate information at the proper time in order to make responsible decisions regarding the effects of YRE programs. Responsible decision-making based upon the availability of appropriate information should be the goal of all evaluation efforts.

BIBLIOGRAPHY

1. Campbell, Donald T., and Stanley, Julian C., "Experimental and Quasiexperimental Design for Research on Teaching," in Nathan L. Gage, ed. Handbook on Research in Teaching (Chicago: Rand McNally, 1963).

2. Chernoff, Herman, and Moses, Lincoln E., Elementary Decision Theory (New York: John Wiley & Sons, Inc., 1959).

3. Dickson, George E., et al., Educational Specifications for a Comprehensive Elementary Teacher Education Program, Volume I, The Basic Report, Chapter V - Evaluation; Toledo, Ohio: Research Foundation, 1968, pp. 209-235.

4. Elam, Stanley, Performance-Based Teacher Education: What is the State of the Art?, American Association of Colleges for Teacher Education, PBTE Series: No. 1, December, 1971.

5. Finn, James D., "Institutionalization of Evaluation," Educational Technology, December 1969, p. 15.

6. Glass, Gene V., "Comments on Professor Bloom's Paper entitled, 'Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment,'" CSEIP Occasional Paper No. 11, (Los Angeles: University of California, Center for the Study of Evaluation of Educational Programs, 1969).

7. Glass, Gene V., "The Growth of Evaluation Methodology," Research Paper No. 27, (Boulder: University of Colorado, Laboratory of Educational Research, 1969).

8. Glass, Gene V., "Designs for Evaluation Studies," Paper presented at the Council for Exceptional Children Special Conference on Early Childhood Education, New Orleans, December, 1969. (Boulder: Laboratory at Educational Research, University of Colorado).

9. Guba, Egon G., "Study of Title III Activities: Report on Evaluation," October 1, 1966 (mimeo) cited in Daniel L. Stufflebeam, "The Use and Abuse of Evaluation in Title III." Theory into Practice, June, 1967, p. 127.

10. Hinkle, Dennis, "The Conceptualization of the Stufflebeam CIPP Model in a Multivariate Context," Unpublished Doctoral Dissertation, The University of Toledo, Toledo, Ohio, 1971.

11. Jensen, Gail E., The Validation of Aims for American Democratic Education (Minneapolis: Burgess, 1960).

12. Johnson, George A., "The Purpose of Evaluation and the Role of the Evaluation," Evaluative Research: Strategies and Methods, (Pittsburgh: American Institutes of Research, 1970). p.2-3.

13. Larkin, Grey A., and Shaver, James P., "Hardnosed Research and the Evaluation of Curriculum" (Logan: Utah State University, College of Education, 1969).

14. Popham, W. James, "Objectives '72," Phi Delta Kappan, Vol. LIII, No. 7, March, 1972.

15. Scriven, Michael, "The Methodology of Evaluation," Perspectives of Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation, No. 1 (Chicago: Rand McNally, 1969).

16. Schlaifer, Robert, Probability and Statistics for Business Decisions, (New York: McGraw-Hill, 1959).

17. Stake, "Objectives, Priorities and Other Judgment Data," Review of Educational Research, April, 1970, pp. 181-185.

18. Stufflebeam, Daniel L., "A Depth Study of the Evaluation Requirement," Theory into Practice, Volume 5, 1966, p. 121.

19. Stufflebeam, Daniel L. "Toward a Science of Educational Evaluation," Educational Technology, July 30, 1968, p. 6.

20. Stufflebeam, Daniel L., et. at., Educational Evaluation: Decision-making, (Itasca, Illinois: F. E. Peacock, 1971).

21. Stufflebeam, Daniel L., and Westerland, Stuart R., "The Evaluation of Context, Input, Process and Product in Elementary and Secondary Education," Commissioned by: The Bureau of Elementary and Secondary Education, United States Office of Education, February 28, 1967, pp. 21-22.

22. Thorndike, Robert L., and Hagen, Elizabeth, Measurement and Evaluation in Psychology and Education, (New York: John Wiley & Sons, Inc., 1969).

23. Tyler, Ralph W., "Assessing the Progress of Education," Phi Delta Kappan, September, 1965, p. 13.

24. Urmson, J. D., "On Grading," Philosophical Essays on Teaching, (Philadelphia: Lippincott, 1969), pp. 194-217.

25. Wolf, Richard, "A Model for Curriculum Evaluation," Psychology in the Schools, Volume 6, 1969, pp. 107-108.

26. Worthen, Blaine R., "Toward a Taxonomy of Evaluation Designs," Educational Technology, August 15, 1968, p. 3.