A CIPP-BASED MODEL FOR THE EVALUATION OF INSTRUCTIONAL MATERIALS

Francisco J. Islas-Guzmán

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements
for the degree of Master of Arts at
Concordia University
Montréal, Québec, Canada

May, 1980

Abstract

A CIPP-BASED MODEL FOR THE EVALUATION

OF INSTRUCTIONAL MATERIALS

Francisco J. Islas-Guzmán

The purpose of this thesis is to develop a model for the evaluation of instructional materials in educational settings. The central notion of the model is that there are discernible evaluation stages in the development of instructional materials. A review of four evaluation frameworks is presented to sinthetize the state of the art in the evaluation theory. These include: the Countenance Model, the Discrepancy Model, the CIPP Model, and the Scriven's Model. The structure of the proposed model follows the definition of evaluation and the rationale of the CIPP Model. Context evaluation encompasses activities such as needs assessment and goals and objectives establishment; Input evaluation entails activities focused on the evaluation of plans and strategies, including the process of selection of instructional materials; Process evaluation directs the activities in the formative evaluation; and Product evaluation determines the instructional value of the final product. The validity of the model is discussed within a systems approach framework and finally a plan for a field testing of the model is proposed.

Acknowledgments

The author wishes to express his gratitude to Dr. Dennis Dicks for his assistance and guidance in the preparation of this thesis. The author would also like to acknowledge the assistance from all the members of the Committee.

Table of Contents

List of Figures

List of Tables

CHAPTER I

INTRODUCTION

A. Problem Statement

The evaluation of instructional materials is an extremely important yet often overlooked component of the total instructional process. The term "instructional materials" refers to all equipment and materials traditionally called "audiovisual" and all the so-called "new media" such as television, films, textbooks, filmstrips, transparencies, laboratory manuals, programmed instruction, computer assisted instruction and other such materials.

Often the evaluation of instructional materials is operationally defined as a "checklist" by many curriculum especialists who perform a weak form of input evaluation on instructional materials, using some variation of a survey form or checklist as an unique means (e.g. Felter 1978; Hug, 1973; McLaughlin & Trlica, 1976; and Teague, 1976).

The validity and reliability of these "materials evaluation checklists" and "materials evaluation guidelines" are invariably questionable. The result is that most school systems are basing what amounts to major instructional decisions on simple descriptive information that has been more or less idiosyncratically organized from school to school. Even in state systems that attempt to evaluate materials there is little recourse to empirical evaluation either before or following adoption of materials (Webster, 1976). Here too the

reliance on the ubiquitous checklist is heavy.

A multitude of instructional materials are marketed to or produced in schools and other educational and training institutions. It is important to know how effective these materials are in helping teachers teach and learners learn. The users and suppliers of instructional materials are constantly making decisions about the relative value of various products, whether on a structured formal basis, or on a highly informal and limited one. Then teachers and media specialists are extensively engaged in the evaluation of instructional materials.

Media specialists and teachers evaluate instructional materials for different purposes, but share common problems in arriving at decisions. Teachers evaluate materials to determine whether they meet specific needs of students whose learning activities they direct. Media specialists are concerned additionally with the extent to which materials are consistent with general curriculum needs and whether materials being considered for purchase will complement existing holdings.

Every evaluator of instructional materials needs a well developed procedure that will give direction to the evaluative process and result in an evaluative decision as accurate and appropriate as can be achieved. Rather than calling for broad descriptive statements about the potential instructional values of materials, the procedure should directly reflect specific criteria and force the evaluator to apply appropriate criteria to the materials being considered. There is a widespread need to determine among the increasing quantity of instructional materials which ones are of high quality.

Evaluation, to be accurate, must involve trying instructional materials under controlled conditions with a wide variety of learners. Seldom either teachers or media specialists follow an elaborate evaluative model involving large number of students randomly selected. At best, teachers and media specialists examine, with great care, the materials being considered for use with learners.

Two of the most serious faults with current instructional materials are either 1) that their development and evaluation are too far removed from the classroom teacher and 2) when the operational learning situation is purported to be controlled, the research methods are ineffectual.

The lack of models for evaluating instructional materials has been attributed to three major factors (Armstrong, 1973); 1) the lack of trained evaluators, 2) a lack of appropriate evaluation instruments and procedures, and 3) a lack of adequate evaluation theory (Stufflebeam, 1969). Although this comment on the "state of the art" is over ten years old, these same factors undoubtedly still apply.

The purpose of this thesis is to present a model for the evaluation of instructional materials. The central notion of the model is that there are discernible evaluation stages in the life of instructional materials, and that it is possible to specify criteria that should be met before the material is advanced to the next stage of development.

B. Evaluation and Evaluation of Instructional Materials.

Evaluation and evaluation of instructional materials have been some of the most frequently talked about and written about phenomena in recent educational experience. This is evidenced by the large number of documents which have been published on educational evaluation during the past years (e.g. Bunda, 1976; Naegle, 1970). One of the major contributions of this body of literature has been the definitions of evaluation and related terminology.

Evaluation has been defined and characterized in many different ways. Evaluation has been described as "the discovery of the nature and worth of something" (Stake & Denny, 1969). Evaluation has been characterized by what it is supposed to do, (Scriven, 1967); for example, evaluation attempts to answer certain types of questions about certain entities; the entities of concern are the various educational instruments and the types of questions include questions of the form: How well this instrument perform? Does it perform better that some other instrument? What does this instrument do? Is this instrument worth what it is costing?

Evaluation has also been characterized by being both descriptive and judgmental. Evaluation is seen to both "describe something and to indicate its percived merits and shortcomings (Stake & Denny, 1969)", or as a process aiding to decision making, for instance, educational evaluation was defined as "the process of obtaining and providing useful information for making educational decisions" (Guba & Stufflebeam, 1970). Evaluation

also has been defined as the "process of ascertaining the decisions to be made, selecting related information and collecting and analyzing that information in order to report summary data useful to decision-makers in selecting among alternatives" (Alkin, 1969). In yet another instance, evaluation has been defined as the process of examining certain objects and events in the light of specified value standards for the purpose of making adaptive decisions (Paulson. 1970).

The roles and goals of evaluation have also been used as definers (Scriven, 1967). The goals of evaluation are to determine the value or worth of an object. The identified roles of evaluation are: formative evaluation which is referred to the gathering of data while a program is being developed for the purpose of guiding the development process; e.g. continuous data collection and feedback in instructional systems (Glaser, 1967; Lindvall & Cox, 1969); and summative evaluation which refers to the assessment of the overall value of an educational program (Scriven, 1967).

In order to organize and understand the different approaches to evaluation, several classifications has been advanced, among them the classifications from Gardner (1977); House (1978); Popham (1975); Steele, (1973); and Worthen and Sanders (1973).

Gardner utilizes as classification criterion the definitions of evaluation, identifying five categories of evaluation frameworks. The five definition-based categories of evaluation are: 1) evaluation as professional judgment, 2) evaluation as measurement, 3) evaluation as the assessment of congruence between performance and objectives (or

standars of performance), 4) decision-oriented evaluation, and 5) goal-free/responsive evaluation.

House (1978) bases his taxonomy in the comparison of the underlying theoretical assumptions on which the models are based. The result are eight categories of evaluation models; these are: 1) systems analysis models, 2) behavioral objectives models, 3) decision-making models 4) goal-free models, 5) art-criticism models, 6) accreditation models, 7) adversary models, and 8) transactional models.

Popham (1975) groups the models under the most general descriptive rubric reflective of a particular orientation, obtaining four descriptive categories: 1) goal-attainment models, 2) judgmental models emphasizing intrinsic criteria, 3) judgmental models emphasizing extrinsic criteria, and 4) decision-facilitation models.

Steele (1973) grouped 50 approaches to evaluation based on the primary focus they have in common, obtaining the following six groupings: 1) evaluation as input into decision making, 2) evaluation of program parts, 3) evaluation-kinds of data and types of activities, 4) evaluation processes, 5) results-attainment of objectives, and 6) results-evaluation of outcomes and effects.

Finally, Worthen and Sanders (1973) developed their classification attending to the analysis of the similarities and differences in the evaluation strategies suggested by several models, resulting three categories: 1) judgmental strategies, 2) decision-management strategies, and 3) decision-objective strategies.

As can be seen the models and approaches to evaluation can be

classified applying different types of criteria, however the main point at least for the practitioner of evaluation, is not to classify the models but to select that or those which would satisfy the requirements of a specific evaluation problem.

On the other hand, some of the most crucial problems facing all persons who evaluate materials center around procedures and instrumentation. Educational literature is flooded with forms and instruments for evaluating materials. Many individuals and groups have developed their own forms for specific localized applications (see for example: Antonoplos, 1977; Nystrom, 1977; Thompson, 1974).

An analysis of evaluative forms in current literature reveals that several factors tend to limit their effective use on a broad scale: 1) some instruments are tied specifically to only one type of material (e.g., Martin, 1973); 2) some instruments provide only for recording broad, general conclusions about materials (e.g., Hug, 1973); 3) some instruments include little or no reference to specified evaluative criteria (e.g., Cowles, 1976); 4) some instruments ask for excessive amounts of information of a nonevaluative nature (e.g., Penta & Cavero, 1976); 5) some instruments are too detailed and lengthy to be of practical use.

A review of over 30 instruments (Lathan, 1973) used throught the United States has revealed a variety of approaches to assessing the value of instructional materials for educational purposes. Althought all the instruments have a number of common elements a common defect is that they do not measure the educational value of instructional material. In other words, in no way can the data from such instruments

be interpreted to identify a given instructional material as "good" or "poor", "effective" or "ineffective" as an aid to learning. What they appear to measure is teacher's perceptions of the value or worth of instructional materials as aids to learning; these findings have prompted a suspicious view toward the accuracy of the activities and instruments in the field of instructional materials evaluation.

CHAPTER II

FOUR EVALUATION FRAMEWORKS

Evaluation is a pervasive feature of all schools. School personnel continually are called upon to make both overt and covert judgments about individuals, programs, and policies. Teachers evaluate student learning; principals evaluate teacher competence; school boards evaluate administrative efficiency. Apart from the evaluation of people, the school setting calls for the evaluation of new curricula, programs, policies, schedules, and practices. The impact of a media center, team teaching, flexible scheduling, or a new science textbook are a few examples of innovations and practices that require evaluation in school settings. Evaluative decisions are needed at all levels of school organization, and they affect personnel and practices at each level.

To help sort out and systematize the multiplicity of roles of evaluation can play in educational decision making, numerous conceptual models outlining various types and functions of evaluation have been advanced (see, for example, Bunda, 1976; Carter, 1975; House, 1978; Steele, 1973). The aim of such models is to identify crucial decision-making points within the educational process. It is important to note that evaluation models themselves are not blue prints for conducting an evaluation study (Airasian, 1974). Instead, they identify the informational or decision-making needs to be considered in planning an evaluation. Once a particular model is selected as a guide, the

evaluator is faced with such practical problems as identifying appropriate data-gathering instruments, selecting samples, and analyzing data. In sum the many evaluation models suggest what decisions need to be made but not how they should be made.

The four models to be briefly described have been selected in terms of their relevance to the broad field of educational evaluation, and their utility in the development of a model for the evaluation of instructional materials.

A. The Countenance Model

Created by Robert Stake (1967), the Countenance Model is so named because of the title of his article describing it ("The Countenance of Educational Evaluation"). This model is based on the notion that judgment and description are both essential to the evaluation of educational programs. Accordingly, Stake distinguishes between three bodies of information that are elements of evaluation statements that should be included in both descriptive and judgmental acts. These elements are: antecedents, transactions, and outcomes.

Antecedents refers to conditions existing prior to implementation of the program that may relate to outcomes. Transactions are the "succession of engagements" that constitute the process (the instructional process). Films, examinations, homework, class discussions and teachers' comments on student papers are all examples of transactions. Outcomes, as conceived by Stake, refer to much more than traditional student outcomes. They include immediate, long-range, cognitive, affective, person, and societal outcomes. Outcomes also include the program's impact on teachers, administrators, and others.

Descriptive information is classified either as intents or observations. Intents include program objectives, not only intended student outcomes, but also the planned-for environmental conditions as well. The judgment matrix includes both the standards used to reach judgments and the actual judgments themselves. A graphic representation of Stake's layout is presented in Figure 1.

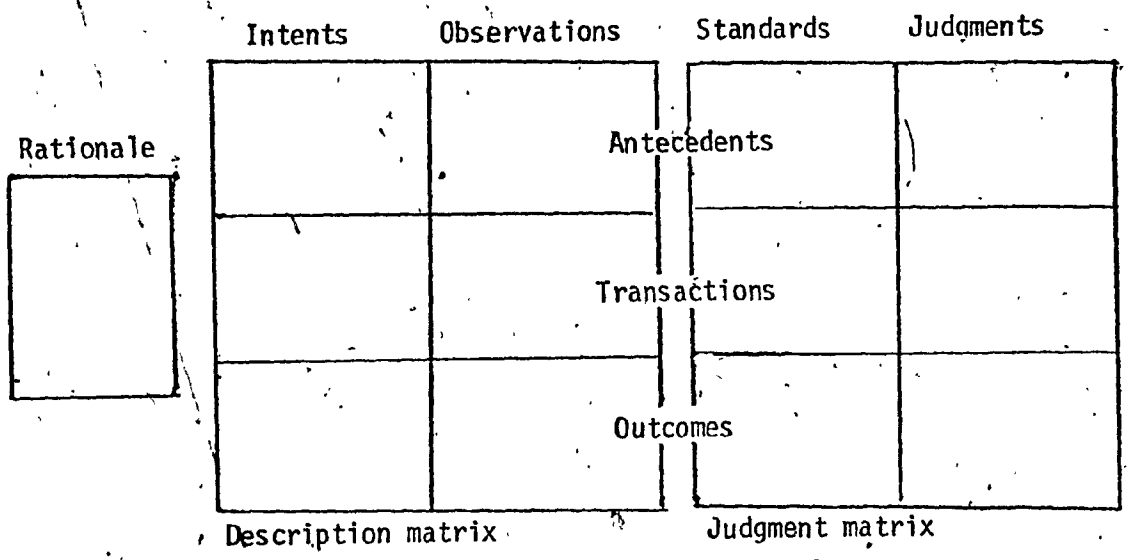


Figure 1 Layout of the Countenance Model (Worthen & Sanders, 1973)

Note that a separate box depicted to the left of the layout is labeled rationale. According to Stake, an evaluation is not complete without an statement of the program's rationale. This statement indicates the philosophical background and basic purposes of the program and provides a basis for evaluating intents.

There are two principal ways of processing descriptive evaluative date: finding the contingencies among antecedents, transactions, and outcomes, and finding the congruencies between intents and observations.

The data for a program are congruent if what was intended actually happened, although Stake admits that it is unlikely that all of the intended antecedents, transactions, and outcomes come to pass exactly as intended even in the best of programs. With reference to transaction data, Stake insists that the evaluator carefully observe and record data emerging from the transactional and interactional classroom processes. He broadens the general concept of outcome data to include future application, transfer. and the effect of process on outcomes.

The contingencies among the variables are of special importance to the evaluator. In the sense that evaluation is the search for relationships that facilitate educational improvement, the countenance evaluator's task is to identify outcomes that are contingent upon particular antecedents conditions and instructional transactions.

As it has been stated the foundation for a model's orientation derives from the author's definition of evaluation. In this case Stake is a proponent of the value-judgment school, the model is judgmental and the process of judging the merit of a program is an integral part of the model.

B. The Discrepancy Model

The Discrepancy Model proposed by Malcom Provus (1971), so named because the discrepancy between performance and standards is a key point in his definition of evaluation. Provus (1971) defines evaluation as: "...the process of 1) defining program standards; 2) determining whether a discrepancy exists between some aspects of program performance and the standards governing that aspects of the

program; and 3) using discrepancy information either to change performance or to change program standards."

Depending upon the information yielded as a result of the evaluation there are four possible decisions to be made. The program can be terminated; it can be modified; it can continue or be repeated as is; or the standards can be changed.

The Discrepancy Model involves five stages, each of which involves a comparison between reality, or performance, and standards. Discrepancies are determined by examining the three content categories (input, process and output) at each stage and comparing the program performance information with these defined standards at each stage.

The design of the program is compared with design criteria; program operations are compared against the input and process selections of the program design; the degree to which interim objectives are achieved is compared with their specification in the program design; and, finally, the cost of the program is compared against the cost of other programs with similar goals.

The first stage focuses on the design and refers to the nature of the program -its objectives, students, staff and other resources required for the program, and the actual activities designed to promote attainment of the objectives. The program design that emerges becomes the standard against which the program is compared in the next stage.

The second stage, installation, involves determining whether an implemented program is congruent with its implementation plan.

Process is the third stage, in which the evaluator serves in a

formative role, comparing performance with standards and focussing on the extent to which the interim or enabling objectives have been achieved. The fourth stage, product, is concerned with comparing actual attainments against the standards (objectives) derived during stage 1 and noting the discrepancies. The fifth and final stage is concerned with the question of cost. A cost-benefit analysis is made of the complete program and compared to other programs similar in nature.

C. The CIPP Model

Probably the most comprehensive of existing evaluation models is the CIPP Model developed by Stufflebeam et al (1971). CIPP is an acronym that stands for the four types of evaluations for which the model is appropriate: context evaluation, input evaluation, process evaluation, and product evaluation.

As noted earlier, the foundation for the development of a model is the author's definition of evaluation, and for Stufflebeam et al., (1971) evaluation is the process of delineating, obtaining and providing useful information for judging decision alternatives.

This definition contains three important points. First, evaluation is a systematic, continuing process. Secondly, the process includes three basic steps: 1) delineating the questions to be answered; 2) obtaining relevant information so that the questions may be answered; and 3) providing the information for decision makers. Thirdly, evaluation serves decision making. Although there is a judgmental component, the primary emphasis in this model is on decision making.

Basically, the CIPP model answers four questions: 1) What objectives should be accomplished? 2) What procedures should be followed in order to accomplish the objectives? 3) Are the procedures working properly? and 4) Are the objectives being achieved?

The CIPP model distinguishes between four types of evaluation: context evaluation to feed planning decisions; input evaluation to feed programming decisions; process evaluation to feed implementing decisions; and product evaluation to feed recycling decisions.

Context evaluation provides a rationale for determining educational objectives by defining relevant environment, describing desired and actual conditions of the environment, identifying unmet needs, and diagnosing problems that prevent needs from being met. Input evaluation assesses relevant capabilities of responsible agencies, identifies strategies for achieving the objectives determined through context evaluation, and suggests designs for implementing selected strategies. Once a strategy has been selected, process evaluation provides periodic feedback to help predict or detect faults in procedural design or implementation so that interim adjustments may be made. Finally product evaluation provides interim and final assessment of the effects of the strategy selected through input evaluation to meet the need identified by context evaluation. Such assessment is completed in light of process evaluation data.

D. Scriven's Model

Scriven (1967) has conceptualized an extremely straighforward and widely accepted evaluation framework. Not nearly as comprehensive as the CIPP model, it is largely concerned with the process-product

portion of Stufflebeam's Model. According to Scriven, the major goal of evaluation is to credibly judge the merits of educational programs. To accomplish the goal, he introduces the concepts of formative and summative evaluation.

The focus of formative evaluation is upon program improvement. Thus, formative evaluation attempts to provide feedback to program personnel in order to upgrade or improve an educational program while it is in the developmental stage. In the CIPP, interim product and process data provide formative evaluation information to program personnel.

The focus of summative evaluation is upon the determination of the ultimate worth of a program or project. This type of evaluation should be implemented when a program has reached some stability. Summative data feed recycling decisions; as a result of summative evaluation information, a program may be terminated, restructured, continued, or expanded. In the CIPP, final product evaluation information, interpreted in consideration of context, input, and process data, is used to draw summative conclusions about the merits of an educational program to feed recycling decisions.

Discussion

The four evaluation frameworks described are representative of the main approaches to evaluation currently in use.

In order to highlight the important points in each one of the models, a descriptive chart of the way these authors view different aspects of the evaluation process is presented as Table 1. Following, a discussion of the models is presented.

Table 1 Comparison of Four Evaluation Models. (Adapted from Worthen and Sanders, 1973)

	Scriven	Stake	Provus	Stufflebeam
Definition of Evaluation	Gathering and combining performance data with weighted set of goals s scales.	Describing and judging an educational program.	Comparing performance against standards.	Defining, obtaining and using information for decision-making.
Purpose	To establish and justify merit or worth. Evaluation plays many roles.	To describe and judge educational programs based on a formal inquiry process.	To determine whether to improve, maintain, or terminate a program.	To provide relevant information to decision-makers.
Principal Focus	Comparison of performance or product with previously stated standards of performance, goals or objectives.	Collection of descriptive and judgmental data from various audiences.	Identifying discrepancies between standards and performance using team approac. approach.	Production of evaluation reports used for decison-making.
Types of Evaluation	Formative-Summative Comparative	Formal-Informal	Design, Installation Process Product Cost.	Context, Input, Process, Product

Table 1 (continued)

	Scriven	Stake	Provus	Stufflebeam
Implications for the Evaluation of Instructional Materials	Discriminates between formative and summative evaluation. Require use of scientific methodology.	Provides a view of the use of instructional materials as a transaction.	Acknowledges alternative procedures in adjusting objectives and in changing treatment. Forces explicit statement of standards.	Provide continuous evaluation (feedback loops) Allows for evaluation to take place at any stage of the program.

Scriven (1967) has elaborated on the functions of evaluation by noting that, while evaluation can play many roles in education, the evaluation process has only one functional goal, that of determining the worth or merit of something. By making this distinction, Scriven (1967) has emphasized that no study of any program can be labeled as evaluation unless some judgment is made. In other words, values as standards are a central consideration in evaluation studies. The useful distinction between formative evaluation (evaluation used to improve a program while it is still fluid by providing feedback to the developer) and summative evaluation (evaluation of a completed product, aimed at the potential consumer) is the contribution taken for the development of an instructional materials evaluation model.

Stake (1967) has elaborated on the distinctions made by Scriven (1967) to formalize evaluation into a systematic procedure. Stake has discriminated between formal (objective) and informal (subjective) evaluation procedures and has suggested that educators must abandon informal procedures if rational judgments are to be made. The Stake's model is a useful mnemonic device for planning an evaluation study . Like the other models it is not an evaluation recipe but it is an organizational framework which emphasizes the two most important components of program evaluation: description and judgment.

Provus, like Stufflebeam, has suggested focusing evaluation on four stages of program development. Although Provus's and Stufflebeam's conceptualizations are slightly different, both serve to focus the attention of the evaluator on several different stages in the development of a program. The emphasis which Provus puts on a "team

approach" to program evaluation is noteworthy. The cooperation among the members of the team such as the evaluator and the program director can lead to a more informed evaluation. Provus also suggests that involvement of the program staff in decision making will establish a feeling of rapport with the evaluator and may lead to the collection of relevant information which is not always available to the evaluator.

Finally, the approach suggested by Stufflebeam presents several advantages when used in the context of the evaluation of instructional materials: 1) It is noteworthy that the CIPP model is cyclical in that feedback is continuously being provided to the decision-maker and new information may lead to reexamination of earlier decisions in a sequence which resembles the process of development of instructional materials; 2) the CIPP model makes evaluators aware of both the variety and range of evaluative information that is necessarily a part of the different types of decisions that have to be made in the development of instructional materials and the different settings in which those decisions have to be made; and 3) as Stufflebeam has pointed out, the evaluation studies are closely related to management procedures and decision-making and the nature of evaluation is such that it does not go on a vacuum: instead, is influenced greatly by many diverse contextual factors. Particularly, in the evaluation of instructional materials it is essential that the evaluator be fully aware of these factors when evaluating them.

CHAPTER III

AN INSTRUCTIONAL MATERIALS EVALUATION MODEL

A. Purpose of the Model

The purpose of this model is to provide a strategy for guiding and monitoring efforts within the instructional materials development and evaluation. The potential users of this model are those educational practitioners responsibles of the development and evaluation of instructional materials, who are not trained or experienced developers and evaluators, who have neither the time nor the money to carry out elaborate evaluation studies, and whose development activities are modest in scope.

The model presents a system in which every phase of the development of instructional materials is feedback by formative evaluations implemented throught the process, the presentation of the model is concentrated upon developing and understanding of the elements of the CIPP model (Stufflebeam, 1971): Context evaluation, input evaluation, process evaluation, and product evaluation.

B. Rationale of the Model

Guba and Stufflebeam (1970) identified four categories of decisions to be served by evaluation, those categories are:

1. Planning decisions. This category encompasses those decisions that set the boundaries of the system. In particular, they

include two classes: decisions to make changes and decisions to maintain existing operational elements as they are.

2. Structuring decisions. Those decisions that determine project designs are included in this category. Such decisions are considered as structuring if a planning decision to change a part of the operation has been made. In other words, if a planning decision is made to mantain the status quo, no new structures need to be established. All that is needed is to continue the operation as is. If a decision to change is made, a series of decisions must then be made to implement the proposed change.
3. Implementing decisions. Once a designed course of action has been approved and initiated, an array of decisions are encountered relating to implementation. Included here are assessments that determine whether the actual procedures coincide with those that were planned and decisions about modifying the projected course of action on the basis of circumstances that appear after the project has started.
4. Recycling decisions. As a project reaches its termination point, decisions are faced regarding what to do next. Are the results close enough to those expected that the project can be installed as a continuing part of the overall system? Are the results such that a second trial run is necessitated? Or are the results so bad that the whole effort ought to be scrapped?

For each of these categories of decision making a different type of evaluation activity is called for. Stufflebeam et al. (1971)

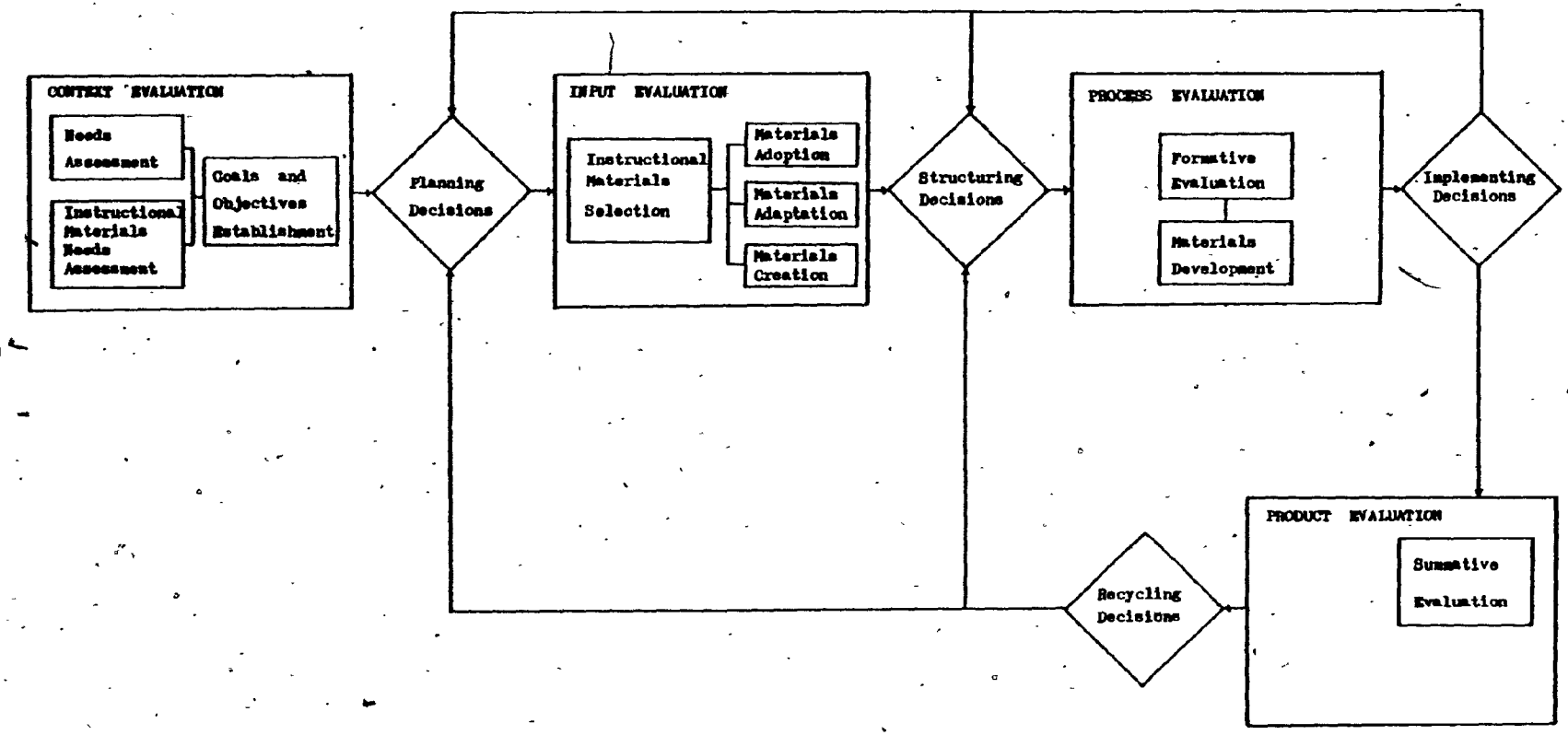
describe these as follows: "Context evaluation serves planning decisions to determine objectives; input evaluation serves structuring decisions to determine project designs; process evaluation serves implementing decisions to control project operations; and product evaluation serves recycling decisions to judge and react to project attainments." (p. 218)

The proposed model for evaluation of instructional materials follows the rationale of the CIPP model. There are three levels in the four categories of decisions. Planning decisions are the first level. They are continually encountered and are fed by context evaluation which includes assessments of curricular needs and instructional materials needs as well as the specific goals and objectives proposed to meet the needs (see Figure 2).

At the second level among the four decision types are structuring decisions. They come into being when and if a decision is made to change. Here the function of input evaluation is to provide information on which to base decisions on the type of instructional material to be selected and developed.

The third level of the model encompasses both implementing and recycling decisions. Context and product evaluation are conducted concurrently to provide information on what the implemented strategies for test the instructional materials consisted of, and what effects resulted. The combination of information from these two types of evaluation helps to clarify the alternatives at the end of a trial. If the trial is successful and the materials are incorporated as an ongoing component of the system, the process and product evaluations

Figure 2 A Model for the Evaluation of Instructional Materials



are used to adjust the continuous context evaluation effort. Finally, if the results indicate that the project is unsatisfactory for achieving the desired objectives, the descriptions should illustrate some of the details to be attended to at input evaluation level as different alternatives are sought.

In summary then, the function of evaluation of instructional materials, as proposed by the model, is the provision of information that will assist in making programmatic decisions. To do this, the model developed displays context evaluation, a continuous effort examining the degree to which the instructional materials are accomplishing its objectives and what objectives they might take on; input evaluation, the effort that starts with a decision to make a change or to include instructional materials and that assesses alternative ways of effecting the desired change; and process and product evaluation, the effort that tests proposed instructional materials, when information about their effects is not available in advance.

CHAPTER IV

CONTEXT EVALUATION

A. General Principles

Context evaluation is defined by Stufflebeam et al.(1971) as the provision of baseline information that delineates the environment of, interest, describes desired and actual conditions pertaining to the environment, identifies unmet needs and unused opportunities, and diagnoses the problems that prevent needs from being met and opportunities from being used.

The major objective of context evaluation is to determine the curricular needs, specify the population and sample of individuals to be served, and devise objectives designed to meet these needs. The procedures for context evaluation include: 1) defining and describing the environment in which the action will take place; 2) identifying the needs; 3) identifying sources of problems or deficiences in meeting these needs; and 4) predicting future discrepancies by considering the desirable, expected, possible, and probable outcomes. In other words, context evaluation provides the rationale for justifying a particular type of program . This stage in the model starts when the development team has been assigned to the task of developing instructional materials for a course on a defined or semi defined topic. At the end of this stage the topic of the instructional materials has to be clearly defined, the subject matter to be covered by the material has to be determined, instructional methods to be applied, besides the instructional

materials, have to be chosen; the general organization of the course has to be defined, and a working plan for the team has to be developed. Decisions of the development team of this stage are facilitated by various kinds of information, such as information on the instructional needs of students for whom the instructional materials are being developed, general educational goals that must be served, alternative teaching methods, and patterns of instructional materials. Such information can be obtained from the literature, educational consultants or from direct observations in classrooms and interaction with teachers and students. Obviously, a systematic instructional needs assessment study is necessary.

The major evaluation activities at this stage of development of the model include the following:

a) Review of the literature. This is a joint effort of the development team and the evaluation staff. The focuss is mainly on documents related to the subject matter.

b) Pilot observations in classrooms. The evaluation staff assists the development team in designing and conducting initial observations of classrooms occupied by students of the target population.

c) Interviews of experts and practitioners. These interviews are focussed mainly to the identification of the goals of the course.

d) A systematic instructional needs assessment study to delineate the needs of students and teachers to be served by the instructional materials. This topic is discussed separately.

B. Needs Assessment

Needs assessment is a large class of activities that are aimed at determining the needs that exist among certain group of people. In the context of education and training programs, a need may be defined as a condition in which there is a discrepancy between an acceptable state of affairs and an observed state of affairs (Anderson et al., 1975). A need is usually taken to be measurable discrepancy between what is and what should be (Kaufman, 1972). This process by which one identifies needs and decides upon priorities among them may be applied to individuals, groups, or institutions.

Needs assessment is of particular relevance in the context of instructional materials evaluation; it is ridiculus to put much credence in the evaluation of a program that might have illegitimate justification for its existence. For curriculum workers, the concept of needs assessment can be considered essentially equivalent to the process of identifying educational goals (Popham, 1972).

The extent of the discrepancy may be either objectively measured or subjectively estimated (Anderson et al. 1975). In the first case, the level of measured performance is compared with the level judged acceptable. In the second case, selected judges are asked to indicate the extent to which needs exist in a given area. The proposed needs assessment in this stage of the model, pertains to the second category, since the needs to be evaluated correspond only to a part of the curriculum design process, the instructional materials development, and because there is not a measured performance to be

compared with the standards.

A needs assessment usually consist of the following steps:

1) Identify the goal areas considered important to the educational system.

2) Develop a rating scale for judging the degree to which present performance in the goal area is acceptable. Goal areas could be rank-ordered according to the acceptability of performance, or the acceptability of performance in each goal area could be rated on an interval scale.

3) Obtain the ratings from a group of judges and average the ratings to obtain indices of need. It is useful to keep points of view separate here if the responses of some judges are clearly different from those of other judges.

4) The next step is to assess the priorities among needs for the purpose of developing instructional materials.

Several factors affect decisions about need priorities and may be considered either when establishing needs or at a later time. One factor is the judged importance of the goal areas. Given two goal areas, discrepancies in the high-importance areas may be attacked before discrepancies in the low-importance areas. The number of persons demonstrating a need may also be taken into consideration. The decision maker must determine whether it is preferable to attack intense needs demonstrated by only a few persons or to deal with less intense needs that are, however, pervasive. For example, it is better to design a set of instructional materials directed to a few students who are failing a course of study, or instructing many others, through instructional materials, in useful skills that they would not

otherwise learn? A third factor that affects the ordering of need priorities is the feasibility of initiating programs to eliminate the need. Even though a need might be important and pervasive, if the decision maker cannot see a ready way to attack it, it might be assigned a lower priority than a need considered less vital but about which something can be done. An example quoted by Anderson et al.. (1975) illustrates this; he mentions that it is certainly more feasible for a school to try to improve children's low reading scores than to try to improve the educational levels of their parents -even though the latter maybe recognized as a major source of the children's reading problem. In some cases, it may be possible to conduct research to find ways to deal with important but stubborn needs. However, the delays attendant on a long-term research effort must be weighed against the short-term effect of directly attacking a less important need.

C. Instructional Goals and Objectives Establishment

Having completed the instructional needs assessment the development team can establish goals and objectives. Just as there were many instructional needs to be satisfied and the instructional needs assessment procedures were employed to identify the prioritary instructional needs, there are potentially many goals for a course; however, for purposes of planning, the priorities identified in the instructional needs assessment should now form the basis for establishing the course goals.

Like the instructional needs assessment, goals can focus on learner outcomes, desired staffing levels, improved use of facilities, and other desired resources. However, at this stage of evaluation of the planning, it is important to limit goal statements to explanations of what students should have in terms of knowledge, skills, abilities, or attitudes, when they complete the course.

Goal statements are a natural out growth of the instructional needs assessment, and in fact, goals are often just reworded needs statements. For example "students demonstrate low reading comprehension levels" might be a needs assessment translated in the following way as a goal statement: "Student reading comprehension level should be improved."

Goals are statements of intentions or purposes that have the characteristics of being future oriented, broadly directed, and timeless; as opposed to objectives which specify desired levels of accomplishment, which can be measured within a given time under specific conditions.

However, the terms goals and objectives are often used interchangeable. The distinction rests in the level of generality of each kind of statement. Goals are logically stated first; they are concerned with ultimate outcomes and are usually phrased in general or global terms. Objectives are narrower and usually short-range; they are statements of student behavior that, taken together are tought to contribute to the envisioned final goals. Learner outcome objectives are behavioral statements of all or some aspects of a goal statement.

The rationale for using behavioral objectives is throughly

treated in a number of documents (i.e. Gagné & Kneller, 1972; Mager, 1962; Popham, 1970). Briefly, however, the essence of instruction is behavioral modification and therefore, objectives for instruction cannot be otherwise stated than in behavioral terms. Unless there is a clear, acceptable statement of instructional outcomes, students, teachers, administrators, parents, will never know if programs are successful or unsuccessful.

Goals and objectives can be established on any level in the school. They can refer to all students in the school, students in specific grade levels, students in particular classrooms, or ultimately individual students.

Steps in setting goals and objectives

1. Convert the priority instructional need statement into program or course wide goals.
2. Decide what level performance objectives will be written for each program or course (program, course, classroom. individual student).
3. Develop performance objectives for each goal that specifies learner outcomes that can be assessed at some specific time.
4. Check to make sure that the instructional needs identified, the goal statements based on the needs, and the objectives statements based on the goals, all logically follow one another.
5. If there are goals for which performance objectives cannot be stated, other planning approaches can be developed.

CHAPTER V.

INPUT EVALUATION

A. General Principles

The purpose of input evaluation is to provide information for determining how to utilize resources to meet program goals (Stufflebeam et al., 1971). The end product of input evaluation is an analysis of the strategies for implementing the selected instructional materials.

Input evaluation helps determine the best resources to accomplish program goals. It logically follows the context evaluation stage. The development team can use three major sources of information in selecting instructional materials to meet specific needs: 1) Previous summative product evaluation information, 2) research information, and 3) non-research information.

1) Summative product evaluation concerns the extent to which specific product or program goals are achieved. When product evaluation information is available on a given program with goals similar to those identified in response to context evaluation information, that information helps decision makers determine the probability that the program would reduce the identified discrepancy between desired and existing conditions.

2) Research information pertains to fundamental relationships that affect student learning and the interaction between student characteristics, teacher characteristics, and instructional system.

Before making a decision to implement or to produce a given program, decision makers should know if that program is or is not consistent with the principles established by basic research in learning and development.

3) Non-research information also enters into any materials selection decision. The influence of materials distributors, costs, political feasibility of material adoption in given communities, capabilities and attitudes of staff members and existing facilities are among the many other considerations that influences such decisions.

Input evaluation provides information for decision-making relative to answering the questions about strategies and plans for meeting the needs and goals. If one posed all of the possible questions relative to the viability of a strategy or plan for meeting a given need, these questions could all be grouped under one or the other (or possible a combination of the two) of two general criteria, desirability considerations and feasibility considerations. Desirability refers to the establishment of a sound rationale for the commitment of the resources necessary to fulfill a consumer need, and feasibility to questions of management, cost, and alternative development possibilities (Whrigth & Hess, 1974).

While criteria classes are useful in the generation of key questions, desirability and feasibility are too broad to be of much value. However, by applying desirability and feasibility considerations to strategies in terms of what the strategy is for, where it will be implemented, by whom it will be implemented, and in comparison with

other alternative strategies, it might be possible to generate more specific and more useful criteria. A set of criteria has been presented by Caldwell (1968) that could be applied in the assessment of the proposed strategies and plans. These criteria are: relevance, legality, congruence, legitimacy, compatibility, balance, practicability, and cost/effectiveness. These criteria are not totally discrete and there is some overlap in certain areas. Caldwell (1968) mentions that these criteria are useful in terms of their primary purpose, that is, to assist in the generation of crucial questions which the decision maker should consider. Table 2, from Caldwell (1968), could serve as a framework for conducting a plan assessment in the selection of instructional materials.

Relevance. This criterion refers to the degree to which the need will be met if the purposes of the strategy or action plan are achieved. The criterion of relevance serves to keep planners "on-track" in developing strategies and plans.

Legality. This criterion grows out of the juxtaposition of the strategy or plan and the context in which is to be implemented. It takes in account the legal aspects of the implementation.

Congruence. The criterion of congruence relates to the degree to which a proposed strategy or plan is consistent with the value system of the context in which it is to be implemented. A strategy can be considered highly desirable in terms of the need, but it could be in opposition to certain values in the value system of the context in which is pretended to implement it.

Legitimacy. This criterion grows out of juxtaposing the strategy

Table 2. Purpose of Applying the Criteria and Illustrative Questions for the Evaluation of Plans. (Caldwell, 1968)

	RELEVANCE	LEGALITY	CONGRUENCE	LEGITIMACY
PURPOSE OF APPLYING THE CRITERION	to determine the relevance of the proposed solution strategy to the identified need areas	to determine the legal status of the proposed solution strategy relative to the context within which it is to be implemented	to determine the congruence of the solution strategy with the value system(s) of the context within which it is to be implemented	to determine if the solution strategy is within the purview of the agency charged with implementation
ILLUSTRATIVE TYPES OF QUESTIONS	Are the purposes of the solution strategy directly related to the statement of need? How much of the need would be filled if the end-products of the strategy were to become inputs in the need situation? How reasonable are the assumptions which relate intermediate steps, e.g., teaching ability, to ultimate objectives, e.g., student achievement?	Are there laws in the context within which the strategy is to be used which would prohibit its implementation? If the legality of the strategy and/or its constituent activities is unclear, what are the chances that it might be judged to be illegal?	What value assumptions are embodied in the actions necessary to implement the proposed strategy? Which of these value assumptions are most likely to conflict with the value system(s) of the context? What is the flexibility level of the context in terms of each of the identified areas of possible value conflict?	Is this agency authorized to implement such a set of actions? From whom should clarification be sought regarding the limits of the agency's authority?

	COMPATIBILITY	BALANCE	PRACTICABILITY	COST/EFFECTIVENESS
PURPOSE OF APPLYING THE CRITERION	to determine the compatibility of the strategy with the value system(s), i.e., purposes and goals, of the implementing agency	to determine the impact of the strategy on other components (sub-systems) of the system and on the weights and interrelationships of those system elements	to determine the practicability of the solution strategy in terms of achieving its stated purposes (end-products)	to determine the relative desirability of the solution strategy (in comparison with alternative solution strategies) in terms of the ratio of necessary inputs (costs) to expected outputs (effectiveness)
ILLUSTRATIVE TYPES OF QUESTIONS	What are the stated purposes and/or program foci of the implementing agency? Do these purposes and/or program foci actually operate on the implementation level? Is the focus of the strategy in harmony with the purposes and program foci on both the policy level and the operational level?	What are the implications (both short and long-range) of implementing the strategy in terms of the purpose and/or program foci of the agency? Assuming that the agency's total program involves several components, e.g., research, service and instruction, which are in some kind of balance, what are the implications of the strategy for the balance of the total program?	Are the actions called for reasonable in terms of state of the art constraints? Is the solution strategy realistic in terms of resource availability constraints? Does the strategy adequately provide for logistical concerns? Are there internal (organizational) constraints which might function as barriers to success? Are there external (contextual) constraints which might act as impediments to the achievement of the solution strategy purposes?	What outputs can be expected at various resource use levels? What are the <u>real</u> benefits which will be realized from this strategy? What are the costs of the <u>real</u> benefits which will be realized?

with the group charged with implementing that strategy and refers to whether the strategy is within the purview of the implementing group. The application of this criterion points up some interesting and highly crucial types of questions regarding levels and types of responsability and authority, for example; do the group have the necessary authorization to implement this plan?

Compatibility. The criterion refers to the compatibility of the proposed strategy in terms of the purposes of the implementing group or agency, compatibility is directed toward the question "should the group do it?".

Balance. The criterion of balance assumes that the settings in which solution strategies will be applied are multi-objective and are composed of numerous program elements. It helps to determine the impact of the strategy on other components of a determined program.

Practicability. The criterion of practicability refers to how realistic the proposed solution strategy is in terms of achieving its stated purposes. In order to apply this criterion and to assess the strategies and plans in terms of it, the developing team must be aware of six major types of constraints which might operate as barriers to success (Caldwell, 1968):

1) State of the art constraints. It is conceivable that a solution strategy might require some piece of hardware, for example, an individual computer and accompanying programs. Do such items of hardware and software exist? Have techniques necessary to implement some crucial phase of the strategy been developed? These questions are indicative of information which the

developing team needs in order to make sound judgments regarding potential constraints.

2) Resource availability constraints. The developing team must have information regarding the resources, e.g. men and material, necessary to implement the proposed solution strategy and must assess the availability of such resources. If such resources are not on hand, can they be procured and, if so, are founds available to do so?

3) Logistical constraints. Logistical constraints are those associated with the transportation and scheduling of resources, for example, having the right resources, in right amounts,at the right time. In order for the planner to assess the programming of a solution strategy, he must have information which gives him a picture of the total activity.

4) Organizational constraints. A particular solution strategy may require close coordination and cooperation among several unities or individuals with the development team. If no communication bridges have been built or, if for some reason the organizational structure of the school inhibits such communication, a number of problems might arise.

5) External constraints. In all contexts there are forces or specific organizations which might place constraints on a given solution strategy. These forces or organizations must be taken into account in the implementation of the strategy.

Cost/Effectiveness. This criterion grows out of juxtaposing the strategy with alternative strategies. This criterion should provide

decision-makers with the indications of which of several alternatives will result more efficient. Decision-makers need to compare alternative strategies in terms of what benefits are expected and what these benefits will cost.

B. Instructional Materials Selection

Selecting the best instructional materials for a course is not an easy job. The choice is complex and difficult because it is based upon a combination of interrelated factors. Braby (1973) exemplified this difficulty with the next questions:

-What instructional material(s) would be most practical for packaging, implementing, and updating a program or course?

- Is equipment needed to use the medium selected, and if so, is it readily available? Is special equipment justified to implement the course?
- To what extent must student achievements after the course be exactly according to prescription?
- Does the value of the course, the amount of behavior change, the number of students to be trained, or the life of the course - justify the cost of the medium or media considered?

These questions, and many others that must be answered in the process of course development, indicate the complexity of the problem. Yet, achievement of course objectives is an integral part of the entire course planning process. The course developer must demonstrate ability to evaluate and balance priorities among the many choices available. He must develop courses that not only meet students requirements, but

at the same time respect organizational realities, as has been pointed out in the previous section.

There are no simple formulas or reference tables that match any specific instructional material with any particular course objectives.

The problem of instructional materials selection has been complicated by a tendency to consider this activity as an isolated and independent function that is undertaken at some point well along in the instructional development process.

The Instructional Materials Selection Process

Sometimes the problem of selecting the optimum instructional material for a course is eliminated by the phrasing of the training request. "We need a TV program on..." or "a slide-tape unit should be made for the course on..." are statements that replace one problem with another. The statement implies that several essential steps of the process of course development have been completed: the need for the course has been clearly established; student population has been analized and defined; content and objectives have been specified, and the medium has been selected on the basis of some judgmental decisions. Implicit, for example, is that this recommendation for a specific medium is based upon careful thought about the most efficient and effective instructional vehicle available and best suited for production, distribution, and utilization within the educational organization. Unfortunately, such considerations are not always the bases for selection of media at the time of the initial request, and so the course developer's or instructor's efforts are often seriously impeded.

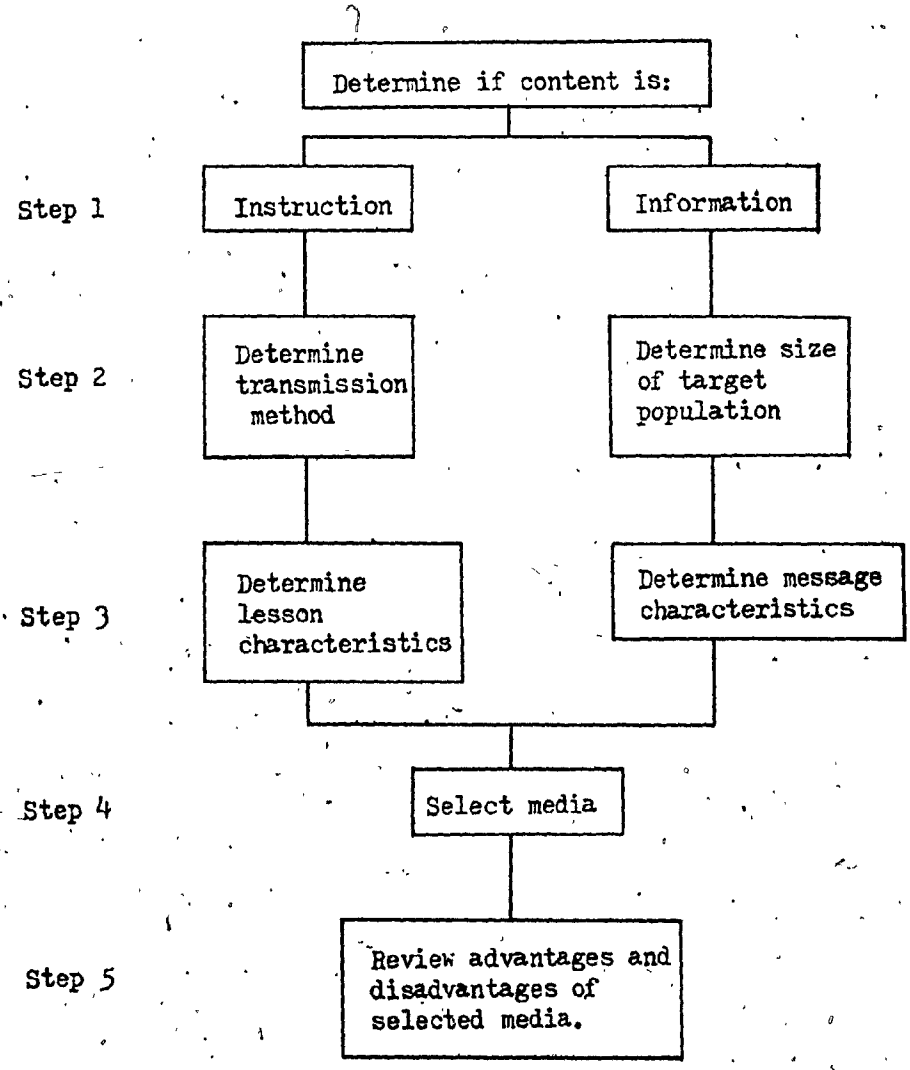
The next part presents a procedure adapted from Anderson (1976), that includes a logical series of steps for the selection process. The procedures for the process are based upon a series of questions that relate course objectives and content with alternative materials characteristics. The questions posed are on a series of material choices in order to narrow the number of alternatives by reasonable decisions.

Overview of the Selection Process

Step 1. Information or Instruction? This step determines whether the purpose of the material is information or instruction. For practical purposes Anderson placed all communication projects in one of two categories: information (in which is included entertainment) or instruction. Both of these categories have sufficiently different characteristics to suggest that different media may be justified for each. The principal differences between information and instruction materials are these:

> Information. The receivers of the information are not held responsible for measurable, specific actions or performance. Often the presentation is general in content, and is to give an overview of ideas or subject matter. Purpose maybe to generate interest, to give background information, or to promote and idea (Anderson, 1974).
>
> Instruction. The receivers of the instruction are to give demonstrable proof that they have learned. Course writers and/or instructors and students are held

Figure 3 Process of Selection of Instructional Materials. (Anderson, 1976)



responsible for the success of the instructional program; and all have evidence of the results (Anderson, 1974).

Step 2. Determining transmission method. This decision is directed by institutional practices or policy. That is, for example, do teachers lecture, give demonstrations, lead class discussions, set up student activities with continual and close instructor supervision? If this is the case then, likely, the selection process is focused to seek instructional aids, that is, materials that are designed and produced to be used by the instructor in teaching. The overhead projector would be an example of a piece of equipment typically used as an instructional aid; some other examples are slides, maps, graphs, and chalkboards. Instructional aids are then, media and equipment used to help an instructor to produce learning. But if the educational organization provides instructional materials for many students scattered over an area and the materials contain a complete message, it is most likely that the selection process be focussed in the second form of resources, instructional media.

Instructional media are those media that provide a direct link between the course developer and the student. Generally, with this use of instructional media, the role of an instructor is different from that of stand-up teacher. When using instructional media, the role of the teacher is usually that of a course monitor, administrator counselor, and supervisor (Anderson, 1976). Most student work is undertaken by selfdirection and by the guidance provided within the instructional media themselves.

With these evident differences in the types of instructional materials, it could be understood the use of the terms instructional aids and instructional media. Thus, it is important that the decision on the method of instruction should be made early in the instructional planning.

Step 3. Determining lesson characteristics. An assumption is made here, that the need has been analyzed, the goals of the instruction determined and the instructional objectives defined. The analysis of the lesson characteristics narrow the choices of media and lead to selecting a specific class of instructional materials appropriate for the instructional purposes. This is achieved through a series of questions i.e. Is motion necessary? Sound? Color? (see Appendix A for the complete guideline of instructional materials selection).

Step 4. Select an initial class of media. For this step Anderson (1976) developed an arbitrary classification of the instructional materials, based on their characteristics (see Table 3). As he pointed out, each medium has a number of characteristics, including special capabilities and limitations. Some media are specially capable of presenting information in graphic form; others are also able to communicate with graphics, but can also add an element of motion. There are media that are specially made to provide sound. Some media are especially economical in presenting verbal information, and some, non verbal graphic information, and some are specially convenient to produce modify, and revise, or to package and distribute.

Step 5. Analysis of media characteristics. This final step refers to the analysis of the instructional capabilities and limitations of

Table 3 Media Classifications (Anderson, 1976)

	MEDIA CLASS	INSTRUCTIONAL MEDIA	INSTRUCTIONAL AID
I.	AUDIO	- Audio Tape (reel to reel, cassette)	- Telephone ("Live" person)
		- Audio disc	- Radio (used in "two-way" dialogue)
		- Radio (generally recorded "one-way" transmission)	
II.	PRINTED MATERIAL	- Programmed texts	- Hand-outs
		- Manuals	- Easels
		- Job-aids	- Chalkboards
			- Charts, graphs, maps, etc. used by instructor
III.	AUDIO-PRINT	- Student Workbook and audio tape or disc	
		- Forms, charts, reference materials, etc. used with audio tape or disc	
IV.	PROJECTED STILL-VISUAL	- Slides and Film strips (when supported by verbal message)	- Slides
			- Trasparencies
			- Film strips
			- Holograms

(Table 3 continued)

MEDIA CLASS	INSTRUCTIONAL MEDIA	INSTRUCTIONAL AID
V. AUDIO-PROJECTED STILL-VISUAL	- Sound filmstrip (Audio tape or disc and film strip)	
	- Sound slide set (Slides of all types with audio tape or disc)	
VI. MOTION-VISUAL	- Silent - Motion Film (with captions)	- Silent movie film
VII. AUDIO-MOTION VISUAL	- Motion Picture Film Video	- Picturephone
VIII. PHYSICAL OBJECTS	- Actual objects	- Actual objects
	- Mock-ups or models of the real things	- Mock-ups or models of the real things
IX. HUMAN AND SITUATIONAL RESOURCES		- Role play situations
		- Case studies using group members
		- Group participation in decision making
		- Field trips
X. COMPUTERS	CAI-CMI	
	- Computers and various terminal display equipment	

the particular medium in order to reconsider the decisions made in Step 4. The selected medium must be one that seems to suit the local production capacities, facilities, polices and budget. Information about the specific characteristics of particular media can be found in the books from Dwyer (1972); Kempt, (1975) and Anderson (1976).

Figure 3 is the graphical representation of the instructional materials selection process. Each one of the steps have been identified as well as the differentiation between the type of message as instruction or information.

In making decisions on the instructional material, after the selection process, three options are basically available: to adopt, to adapt or to create them. First, in the adoption stage, an effort is made to adopt established materials. If they are adopted, the next step is to develop an intrinsic review of them. However, if the decision is to reject outright adoption, the alternative is to adapt ongoing materials which would contain, nearly appropriate elements of the new course. If the decision is to reject any adapatation, it may be necessary to create entirely new materials.

CHAPTER VI

PROCESS EVALUATION

A. General Principles

Once a course of action has been selected and implementation has begun, process evaluation is necessary to provide periodic feedback to the persons responsible for implementing plans and procedures. Stufflebeam et al. (1971) identifies three main objectives of process evaluation: 1) to detect or predict defects in the procedural design or its implementation during the developmental stages, 2) to provide information for programmed decisions, and 3) to maintain a record of the procedure as it occurs.

Among the strategies to be followed in process evaluation there are three of paramount importance. The first is to identify and monitor continuously the potential sources of failure in the project. These include: interpersonal relationships among development team and technical personnel, communication channels, logistics, understanding of and agreement with the goals of the program by persons involved in and affected by it, physical facilities, staff, and time schedule.

The second involves projecting and servicing pre-programmed decisions to be made by project managers during the implementation. An example of these pre-programmed decisions in instructional materials development is found when it is necessary to choose a specific sample of students for participation in the tryout of newly written programmed

materials. Such a decision can be projected well in advance of the actual time when the choice is to be made. Such decision situation are numerous in most projects and denote explicit process evaluation data requirements. Further, they can materially affect the success of a project, as they often pertain to the development, tryout, and selection of resources needed to operationalize the project design. If the input evaluation phase has been implemented effectively, the evaluator should have little difficulty in delineating pre-programmed decisions and their associated information requirements.

The third process evaluation strategy is the recording of events through regular data collection. In this way project outcomes can be interpreted with a better understanding of what occurred during the program development. This information will be especially useful later in determining why objectives were or were not achieved.

Process evaluation is closely related to the concept of formative evaluation, since formative evaluation information is provided by interim data produced by the activities in process evaluation.

An operational definition of formative evaluation was given by Cunningham (1972) as follows:

"Formative evaluation is the gathering of information which would be of use to the developers of instructional materials, those persons who are trying to choose or produce the parts, the elements which will combine to form the successful whole. Information of concern to developers is usually that which will help them determine the success of their initial efforts so that modifications can be made." (p.111).

Formative evaluation incorporates the collection and use of

feedback data to developers of educational processes and products. The purpose of these activities is to facilitate further assessment, revision, and improvement of instructional materials and other educational components during the formative period of development. Baker and Alkin (1973) mention that substantial amounts of funds are wasted on the purchase and installation of educational products that later prove to be inappropriate or ineffective, and that this can be prevented, if the developers are engaged in formative evaluation of all products.

B. Formative Evaluation of Instructional Materials.

At this point in the instructional materials development process pieces of the intended final product are beginning to emerge. For instance, the development of a film often begins by constructing verbal descriptions of the images that are intended to be filmed. A frequent next step is the construction of a story-board and the simulation of the visual and oral stimuli. Some film producers use video-taped versions of their film for debugging purposes. Written materials also follow the same sort of pattern, construct of some sort of topical outline, rough drafts of chapters or units, revisions based upon feedback and small-scale tryouts of each chapter. The point is that in the development of nearly any product many opportunities exist prior to the completion of the initially satisfying version of the complete product for evaluative information to be collected. The particular techniques useful for process and formative evaluation of various instructional materials will differ somewhat from medium to

medium, but many general principles can be noted.

Formative evaluation information involve collecting internal information such as descriptive information, processing critical appraisals (Sanders & Cunninghan, 1974), and student tryouts. Descriptive information refers to the objective information that can be generated by inspecting the pieces of preliminary versions of the product. Critical appraisals are judgments made concerning the pieces by samples of concerned populations, as content experts, teachers and students.

a) Descriptive information. The purpose of collecting descriptive information is to determine what is the state of development of the material, and to revise once some deficit has been identified.

One type of descriptive information, physical specifications is simply a description of the tangible characteristics of the product consisting in large part of media characteristics. This type of information is best collected by means of a checklist that includes the majority of the characteristics upon which products can vary. These characteristics are usually media specific in that any general purpose checklist would be impossible to construct. Some illustrative characteristics to be included in a checklist are showed in Table 4 including some example questions. (See next page).

One method offering promise for describing product content is content analysis. The content analysis is defined by Berelson (1954) as a "research technique for the objective systematic and quantitative descriptions of the manifiest content of communication." Content analysis, however, does not lend itself easily to a consideration of

Table 4 Illustrative Characteristics Used as Criteria in Formative Evaluation of Instructional Materials

1.0 Content

1.1 Authenticity

Is the content accurate?

Is the content up-to-date?

Are translations and retellings faithful to the original?

1.2 Believility

Does the presentation give the impression of authenticity?

2.0 Instructional Design

2.1 Objectives

Are specific objectives stated for teacher use?

Does the content of the material relate closely to the objectives?

2.2 Organization (scope and sequence)

What is the scope of content covered in the materials?

How is the scope of the material organized?

Is there a specific sequence in the material?

(Table 4 continued)

What is the basis of the suggested sequence?

2.3 Methods of instruction

What method or methods of instruction are suggested?

What role is emphasized in the method, teacher or pupil?

What function is expected from the teacher?

2.4 Evaluation

What test materials are included for student's and teacher's use?

What do the test measure?

3.0 Resources

What are the criteria for the use of instructional material in specific situations?

Are the teachers adept in the use of the materials?

Are additional support materials needed?

Does material usage require special training of teacher?

Is some all or part of the material consumable?

4.0 Provisions for use

How much preparation time is required to use materials?

Is a manual included with the material?

Can materials be used with other than the stated population?

5.0 Formats

(Table 4 continued)

Are the formats adequate for the achievement of objectives?

6.0 Student needs

Are prerequisites needed to use the materials?

Are there suggested follow up materials?

7.0 Quality of materials

7.1 Visuals

Are the photography clear and artistic?

Are printed items adequate in size?

Is color used adequately?

7.2 Sound

Is the sound clear, intelligible, and realistic?

Is the speed adequate?

7.3 Durability

Can nonconsumable components be replaced?

Are materials durable?

the relationship among concepts in the subject matter. In this regard, the learning structure analysis of Gagné (1970) is very useful. This technique is a type of task analysis procedure (Davis, 1973) which includes the nexts steps: 1) specification of the main task; 2) identification of subtasks, 3) identification of sub-tasks until the subtasks are equivalent to the subjects entry behavior.

The usefulness of the collection of descriptive information reside in the fact that when a product fails to perform as expected explanations must be found. An adequate inventory of descriptive information will assist greatly in locating the points at which the product needs revisions.

b) Critical appraisal. Critical appraisal consist of the collection of information from the people who are using the product and people other than those directly involved in its use (Sanders & Cunningham, 1974). Collection of opinionsfrom content experts, teachers, students administrators, authors, etc. can be accomplished by means of questionaries, checklists, interviews, panels, etc. The criteria against which each of these populations can appraise the materials would vary; for instance, teachers would be concerned with such factors as congruence of content, practicability of the format, mode, and requirements of the instruction and so on. Any or all of this information can bear upon the subsequent revision of instructional materials especially when external information supports the critical appraisal.

One important judgmental data source that deserves attention is the author. The term "author" differs somewhat from medium to medium.

In the case of print media, text or audiotapes, the definition is fairly easy, but with film or video-tape the term "author" is probably closest in meaning to director. The author is often not recognized as a source of revision information but he is in fact a major, especially at the early stages of the product. The author makes literally thousands of decisions when he embodies the content he intends to teach in suitable form-decisions concerning sequence, phrasing, orientation, value, difficulty level, and so on.

Authors, however, are often dimly aware of the decision process. Explicit standards are rare and probably as a consequence, consistency in decision making is less frequent than would be desirable. Some profit might occur, therefore, by increasing author awareness of his decisions. For instance, Lawson (1973) has constructed questionnaires and checklists that should prove useful in this regard. In one of his questionnaires, authors are queried on whether specific learner objectives are provided, whether entry behaviors are specific, whether provision is made for learners to enter the product at points other than the beginning, whether the format and display are appropriate for the intended population, whether examples and illustrations used are likely to be of interest to the intended population, etc.

The effects of such procedures upon authors is unknown at this time.

It should be noted that the descriptive and critical appraisal techniques described thus far on formative evaluation can be used at the product evaluation stage as well. The difference is primarily one of the closeness of an interim format that is being evaluated to the

final product, but the principles involved are generally comparable from stage to stage. Much useful external information at the formative evaluation stage can be gathered by using the same criterion measures which will be used at the product evaluation stage.

c) Student tryout. At the interim stage of product development one should not limit his information-gathering activities to highly structured procedures. Much useful information can be gathered in informal types of activities. One that has received increasing attention is variously called developmental testing (Markle, 1967), individual student tryout (Scott & Yelon, 1969), oral problem solving (Cunningham, 1973) and developmental testing (Anderson, 1976). Essentially this technique consist of placing the evaluator with one or more students as they use the materials. Ideally the students will help the author locate ambiguities, errors of sequence, and the like, and allow the author to test assumptions concerning the mental operations that will be employed by students using the material (Sanders & Cunningham, 1974). The information gathered from students will allow the developer discard or heavily revise the material in its early development stages, this permits avoid wasting time and money if completely finished materials are developed before any revision with the intended population.

CHAPTER VII

PRODUCT EVALUATION

A. General Principles

Upon completion of a given cycle of an instructional materials development, a summative product evaluation should be developed.

Product evaluation is used to determine the effectiveness of a project after it has run full cycle (Stufflebeam, 1968). Its objective is to relate outcomes to objectives and to context, input, and process evaluations. The general method of product evaluation includes devising operational definitions of objectives, measuring criteria associated with objectives of the activity, comparing these measurements with predetermined absolute or relative standards, and making rational interpretations of the outcomes using the recorded context, input, and process information (Stufflebeam et al. 1971).

Criteria may be either instrumental or consequential, a distinction proposed by Scriven (1967). Instrumental criteria are related to accomplishments that are at an intermediate level and that contribute to the achievement of ultimate objectives. Consequential criteria are those pertaining to the fundamental conditions being sought. The specification of consequential criteria obviously depends upon the particular intents of a particular group of decision makers.

Traditionally, the evaluation of instructional materials has meant only what has been referred as product evaluation. However, the other three types of evaluation (context, input, and process evaluations)

must be part of the instructional materials evaluation process, as has been pointed out in the previous sections.

Both context and product evaluation assess the extent to which ends are being attained. Context evaluation does this systematically with respect to a total system, and product evaluation does so with respect to change efforts within the system (Stufflebeam et al. 1971). Thus context evaluation provides the specifications in terms of which product evaluation is later carried out.

Product evaluation investigates the extent to which objectives have been or are being attained; process evaluation assesses the extent to which procedures are operant as intended. Both types of evaluation provide feedback for controlling and evolving change procedures in process. Process evaluation makes it possible to determine if actual procedure is discrepant from design, and product evaluation assists in determining whether objectives are being attained.

Product evaluation reports that objectives were or were not achieved, process evaluation provides a basis for interpreting the reason for the outcome. If the objectives were not achieved, it would be important to know whether the intended procedure was actually implemented; process evaluation provides information for this determination. If objectives were achieved, it would be important to have a description of the actual procedure that produced the outcome.

On the other hand, the term summative evaluation, often used as a point of contrast with formative evaluation, refers to the collection of information of use to the consumer of the instructional materials.

those who seek to determine what they are getting for their money, what they can expect from the product, whether the product is better than others, and so forth. Thus summative evaluation is the assessment of the overall effectiveness of an already developed program. The purpose of summative evaluation is to help make decisions regarding the program's future, its continuance, termination, replication and/or dissemination.

The result of a summative product evaluation, typically is a report. If all has gone well, the report indicates areas of program success -intended outcomes attained-, and areas of program failure- intended outcomes not attained- and negative unintended outcomes noted.

The activities and procedures in this phase require the tryout and evaluation of instructional materials in school classrooms. Product evaluation depends on strong experimental design.

The product evaluation of instructional materials is encompassed by the set of activities denominated media research. Salomon and Clark (1977) identify three major objectives of media research. The first refers to obtaining knowledge about the instructional effectiveness of a chosen medium. The second is to increase the understanding of how media function and what psychological effects they have on learners. The third objective of media research is to enhance the practice of education through providing and evaluating improved media, materials, procedures, and technologies.

The field of the instructional materials evaluation is mainly focussed on the achievement of the third objective of media research in that the purpose is to determine the instructional effectiveness of the

selected and produced instructional materials. Instructional effectiveness in this case, is concerned with the contribution of some media attribute to learning in light of the demands of a specific task (Salomon & Clark, 1977). It becomes clear, then, that the instructional materials product evaluation is a restricted area of the research on media; however, the research methodologies are, or should be, the same.

B. Evaluation Research

Evaluation may be viewed as a part in systematic program development. Summative product evaluation follows program implementation and provides a basis for further planning and program refinement.

There are considerable differences of opinion as to what constitutes evaluation research. From one stand point, activities that provide information that facilitates decisions on programs, treatments, and interventions can be considered to be evaluation research . A second type of activities often referred to as evaluation research under a broad definition are judgments on whether or not certain activities, treatments, and interventions occur in conformity with generally accepted standards.

From the point of view of the behavioral sciences, evaluation research represents an application of the scientific method that is quite different from basic research. Wrighstone (1969) suggests that research is more concerned with the basic theory and design of a program over an appropriate period of time, with flexible deadlines, and with sophisticated treatment of data that have been carefully obtained. Evaluation on the other hand, may be concerned with basic

theory and design but its primary function is to appraise comprensively a practical activity to meet a deadline.

Suchman (1967) argues that the distinction between basic research and evaluative research is one of purpose rather than method, and that evaluative research applies the scientific method to problems that have administrative consequences, whereas basic research is concerned with problems of theoretical significance.

Weiss (1972) suggests that the purpose of evaluation research is "to measure the effects of a given program against the goals it set out to accomplish as a means of contributing to subsequent decision-making about the program and improving future programming." Within such a definition, she suggests there are four key features. First, there is the idea that a research methodology be used to measure the effects. Secondly, there is emphasis on outcomes in relation to specific goals as evidence of the effects of the program. Third, Weiss (1972) suggests, that the comparison of effects with desired goals stresses the use of criteria for judging the effectiveness of a program. And finally, she emphasizes the subsequent decision-making and future improvement of programming as the social purpose of the program, and therefore of the evaluator.

Evaluation research represents a form of applied research; it attempts to insure the effectiveness or adequacy of a program, project or treatment under operating conditions. Evaluative research pays attention not only to the desirability of a particular project or treatment but also to its feasibility, its expenses compared to resources available and to other alternatives, the case or difficulty

of implementation, and its ability to withstand subversion by routinization (Archibald, 1976).

Evaluation research can be distinguished from basic or nonevaluative research in several aspects. First, the evaluative researcher typically has his subject matter given to him. He does not formulate his own hypothesis; they come from program goals (Guttentag, 1971). Another distinction is that there are very few variables over which the evaluator can exert control. However, the most significant difference is one of purpose (intent) and not of method. Both types of studies attempt to utilize research designs for data collection and analysis based upon the logic of the scientific method, just as the validity of both types of studies rest equally on the degree to which they satisfy the principles of scientific methodology (Suchman, 1969).

The four models of evaluation described in chapter II contain useful organizationalframeworks whithin which the details of evaluation planning may be placed. There are not systematic taxonomies of evaluation designs analogous to those of Campbell and Stanley (1966) for experimental or nonevaluative research, rather, there are some general guidelines which have been suggested on how one can go about planning evaluation studies.

C. Methods of Investigation

A review of the methods of investigation presented by Gilbert et al. (1975) serves to illustrate the main ways of gathering information and assisting the evaluator of instructional materials in the understanding of his task in this phase of the model.

a) Introspection, theory, analysis, and simulation. These methods according to Gilbert et al. (1975), form an important group of parallel methods for finding out how a program might work. Although these methods can all have substantial inputs from empirical information, they often do not have a strong empirical base; indeed, sometimes they have none, except for casual observation and analogy.

b) Anecdotes, casual observation, and case studies. These methods are very commonly used in medicine, anthropology, law, sociology and education. Like the first set of methods, they are likely to suggest theories and discover difficulties that could not other wise be detected. They provide a firm record of a special event. They are specially weak, however, in giving a bridge from the case at hand to the wider realm of situations that the evaluator wishes to influence in the future.

c) Quantitative observational studies. These studies including sample surveys and censuses, are widely used. They are specially good at telling the current state of the world. This kind of study is ordinarily not designed to include the administration of new treatments, Usually, sample surveys involve a stage of random sampling, but this randomness is designed to reduce certain kinds of bias in collecting information, and for deciding which items should be observed.

Censuses are a form of observational study. Again, they are ordinarily designed to aid the appretiation of matters as they stand, rather than to administer treatments and see what their effects might be.

Among quantitative observational studies, Gilbert et al. (1975)

include searches of records and studies to see how various people fare under various treatments that have been imposed through the natural processes of society, without the investigator's interfering with those processes.

From the evaluation point of view, the general feature of an observational study is that some individuals, groups, or institutions have in the course of events experienced one regimen, and others have experienced other regimens, and these regimens have led to outcomes on variables of interest (Gilbert et al., 1975).

d) Experimentation and innovation. The concept of "experiment" has been utilized in both senses: 1) as the introduction of an innovation, such as the establishment of social security, a new child assistance program, or a new curriculum in a school. The word "experiment" is then being used in the sense of a new and sometimes tentative treatment. And 2) as the scientifically designed experimental trials capable of ascertaining with some precision the outcomes produced by such innovations or of identifying the factors responsible for the outcomes observed.

Because the term "experiment" is frequently used in both of these ways, Gilbert et al. (1975) have referred to the controlled scientific experimental studies as "controlled field trials", and when these trials employ the special precaution of randomization, they prefix the word "randomized."

e) Non randomized field trials. In nonrandomized field trials the investigator initiates new treatments but without the use of randomization. This non randomized method is a step forward from the

observational study because the investigator rather than the social process chooses the treatment, and so the treatment, and the outcome can be related. The difficulties of this procedure have to do with a variety of matters, but the key problem is that the effect of the treatment is not distinctive in all cases, so that the treatment, cannot be proved to have caused the effect of interest. Selection effects and variability of previous experience have often led to biases and misinterpretations (Campbell & Stanley, 1966)

f) Randomized controlled field trials. The randomized controlled field trials are currently the best device for appraising new programs. The word "trial" suggests the direct comparison of the effects of treatments. These trials are also named "studies" because the investigators usually go beyond just the initial comparison of treatments (Gilbert et al., 1975). In these studies, the experimental units -individuals, families, classrooms, school districs, whatever the unit- are randomly assigned to treatments or regimens and carefully followed to find out what the effect of the regimen might be. The randomization helps in several ways. First, it avoids the dangers of self-selection or of biased selection that has been mentioned earlier. It provides objectivity to outsiders. The use of a random method to select the cases assures that the effects of selectivity have been controlled. Second, it helps to control for variables not otherwise able to be controlled.

The expression "field trial" implied that the treatment or treatments are being studied in the field rather than in the laboratory, and that they are being tried out in practice rather than through simulation or theory. A field trial might consist of one

treatment only, but it could refer to several. For instance, a field trial of a new instructional material might be for the purpose of determining the effectivity of such instructional material as a learning device and to determine side effects. "Controlled" refers to two matters: 1) that the choice of treatment for a site or an individual is primary that of the investigator rather than the individual, and 2) that two treatments are being compared or at least one treatment against a criterion referenced measure. "Randomized" refers to the use of chance at some stage to choose which units get which treatments. The purpose of randomized controlled field trials like that of the other kinds of evaluative investigations discussed, is marshalling information to aid decisions.

It is common to think of this device as especially important for the scientist rather than for the decision maker. This leaves us with the idea that a scientific needs a very fine tool to get information about a process, but the decision maker should be content with uncertain information.

D. Evaluation Designs

Design relates to the conditions and procedures that guide data collection. Evaluation is concerned with judgments of merit or worth. The purpose of an evaluation design is to facilitate gathering data, thereby making possible valid statements about the effects or outcomes of the program under study. In most cases, evaluation designs have been borrowed from research.

Experimental design is essentially a problem of organizing the

observation of various alternatives and of specifying criteria and instruments of measurement, resulting in a plan which includes: deciding how to assign experimental units (persons, classrooms, schools districts, etc.) to treatments (alternative types of instruction or to experimental and control conditions); describing treatments; and deciding what measures to apply to the behavior of the units to assess their responses to the treatment. The ability to drawn valid and useful information from experiments rests on the care and insight employed during the design stage.

There are times when evaluators must face the problem of compromising tenets of "good" experimental design to accommodate the political, economic, and social realities to implementation (Rapp et al., 1970). The design must be related to the type of program or service being evaluated; that is, the selection of a particular design is guided by the decisions that will have to made as a consequence of the data. In turn the adequacy of a particular design can be determined by the extent to which the results may be interpreted and the questions answered.

Campbell and Stanley (1966) distinguish between three types of research designs commonly used in evaluation -pre-experimental, quasi-experimental and experimental- evaluating a number of specific designs in each category according to their ability to withstand threats to their validity. That is, the criterion differentiating the three groups of designs, as well as the quality of the designs in each group, is the extent to which the design protects against the effects of extraneous variables. More specifically, the criterion is

the extent to which the design protects against nine threats to internal validity, nine kinds of variables that if not controlled, will affect the outcomes of the program and thus the accuracy of the interpretations that can be made of the data.

The nine threats to internal validity are:

1. History. Events external to the programs, such as changes in the economy or television programming, that can affect the subjects of a program and thus the program results. Outside events are likely to occur when the program being evaluated extends over a long period of time.

2. Maturation. Processes within respondents, such as fatigue or growth, produce change as a function of the passage of time.

3. Testing. The effect of a test on the scores of a second test.

4. Instrumentation. Changes in the instruments themselves, in calibration or difficulty level, or changes in the observer or scorers.

5. Selection. Biases resulting from differences between types of individuals recruited for comparison groups.

6. Statistical Regression. Non-program effects can appear during statistical manipulations. When groups are selected for a study on the basis of extremely high or low scores, their scores on subsequent tests will tend to regress statistically, that,is move back toward the mean of the group.

7. Instability. Unreliability of measures, which causes fluctuation in scores independent of the program under investigation.

8. Experimental mortality. Differential loss of subjects from comparison groups.

9. Selection-Maturation Interaction. Selection biases result in differential rates of maturation or changes as a function of time.

In addition to concern about internal validity, the designer of evaluation studies may also have to contend with problems of external validity. External validity relates to the problem of the generabizability of the results of the study: to what groups, settings, schools, and so forth can the observed effect be generalized (Campbell & Stanley, 1966).

The emphasis of evaluation studies at a local level typically will be upon pragmatic, situation-bound results. Rarely the results of the evaluation studies will be generalized to other schools. Evaluation studies at a local level will most often emphasize utility and action and not study relationships and nuances related to substantive contributions to an area of knowledge (Airasian, 1974). Results are much more likely to be applied to the school itself than they are to be exported to other schools. Hence while external validity may have some relevance for the design of evaluation studies, it is secondary in relation to internal validity.

True experimental designs protect against all of these possible threats to internal validity; quasi-experimental designs generally protect against most of them. Quasi-experimental designs require almost the same rigor, but they are more practical than the true experimental model in many real-world situations. Pre-experimental designs totally lack control and according to Campbell and Stanley (1966) are "of almost not scientific value". Examples of pre-experimental designs are: 1) the one-group pretest postest design

in which a single group is pretested, exposed to a program, and post-tested; depending upon the length of time between the pretest and posttest, the design is open to the threats of history and maturation; 2) the static-group comparison, in which a group that has received a program or service is compared with a group that has not -a comparison that is suspect since the original equivalence of the two groups is unknown; and 3) the one-shot case study in which a single group is studied once. In spite of their limitations the pre-experimental designs has been used widely in the research with and on media. A survey to determine the variety of designs employed in instructional technology experiments in studies published by AV Communication Review between 1970 and 1975, conducted by Clark and Snow (1975) shows that the most common design problem in a majority AVCR studies (33 of 49) was the reliance on pre-experimental designs, i.e. one shot case studies, static group comparisons, etc. "In all of the 33 studies using these designs, none included random assignment of subjects (or units such as classrooms) and few used control groups for comparisons. In some instances, two treatments were compared without random assignment of subjects or control groups. In a few studies this problem was complicated by non random subject or unit attrition between pretest and treatment or during successive treatment application and posttest" (Clark & Snow, 1975, p. 377).

The pre-experimental strategy is exploratory in nature not confirmatory, and should probably not be reported in isolation. The results of such studies are difficult to interpret. While they may

provide important sources of descriptive data for process evaluation purposes, researchers and evaluators should consider alternatives to formal experimental designs only when control procedures such as random assignment are impossible to accomplish. In such instances, many quasi-experimental designs are available (for instance, Campbell & Stanley, 1966; Riecken & Boruch, 1974).

Although quasi-experiments are preferable to pre-experimental designs, they should be considered only when all posible routes to fully controlled experimentation have been explored and rejected as either too expensive, impossible because of the logistical problems, or unrepresentativeness of the environment in which the material is to be used.

Campbell and Boruch (1975) recently have strengthened the argument for the use of true randomized experiments. They say: "It may be that Campbell and Stanley (1966) should feel guilty for having contributed to giving quasi-experimental designs a good name. There are program evaluations in which the authors say proudly 'we used a quasi-experimental design'. If responsible Campbell and Stanley should do penance, because in most social settings there are many equally or more plausible rival hypothesis than the hypothesis that the puny treatment indeed produced an effect." (p. 202).

Although there has been much discussion concerning the utility of applying controlled experiments to the field of educational evaluation (see for instance Guba, 1969a, b; Stanley, 1972), experimental design is to many educators the cornestone of evaluation. Campbell and Stanley (1966) state unequivocally that they are "...committed to the

experiment: as the only means for setting disputes regarding educational practices, as the only way of verifying educational improvements, and as the only way of establishing a cummulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favor of inferior novelties."

Stufflebeam (1971) asserts that if the assumptions required by experimental design can be met, the evaluator has a powerful and efficient tool for answering certain evaluation questions, having a major role in product evaluation.

Classic experimental design incorporates two important techniques that together rule out the possibility that something other than the program caused the observed results, and thus, they confirm the legitimacy of the interpretations made from the data. These techniques are the use of control or comparison groups and randomization. This means that samples of the target population are randomly selected and assigned to either the experimental group receiving the treatment or the control group, which receives a different treatment or not treatment.

True experimental designs used in evaluation. The conventions utilized by Campbell and Stanley (1966) will be adopted to represent symbolically the designs: O represents an observation or measurement; R , the random assignment of subjects (or units) to groups; X , the presence of the treatment under investigation. The following designs has been described and analyzed by Campbell and Stanley (1966), here is presented a brief description.

1. Posttest only control group design:

R X O

R O

The posttest only control group design is perhaps, the potentially most useful of the experimental designs. This design utilizes two groups, one of which experiences the treatment while the other does not, thus controlling for history and maturation. Furthermore, group assignment is made on a random basis, which controls for selection and mortality. In addition, no pretest is given to either group in order to control for simple testing effects and the interactions between testing and treatment. This design is quite ideal, in that it controls all threats to validity or sources of bias. The posttest only control group design is both simple and efficient. Selection variables can often be adequately controlled for by randomization, thus minimizing the need for pretest data.

2. Pretest posttest control group design.

R O X O

R O O

Two groups are employed in this design, one group, the experimental group, receives a treatment while the second group, the control group does not. Both groups are given a pretest and a posttest, the use of a pretest is the only difference between this design and the previous one. By the utilization of a control group, which has the same experiences as the experimental group other than the experience

of the treatment itself, this design controls for history, maturation, and regression. By randomizing the subjects across experimental and control conditions, both selection and mortality are controlled. However, the use of a pretest introduces additional slight design difficulties to those encountered in the posttest only control group design. There is no control for a testing effect, which may reduce internal validity; nor is there any control for the possible sensitization to the treatment that a subject might gain by having the pretest experiences.

The designs described are the most basic ones that include control of all relevant internal validity threats. If the evaluator of instructional materials is able to implement either of these designs, he can make strong cause-effect conclusions from his data with little concern about the existence or reasonability of alternative explanations for the observed outcomes.

A final point in this section refers to the selection of a specific research design in order to evaluate instructional materials. The design provides procedural guidelines establishing the conditions and procedures for collecting the data required to answer the questions of concern. The design must be related to the type of program being evaluated; that is, the selection of a particular design is guided by the decisions that will have to be made as a consequence of the data. The point is, evaluation designs must accomodate the characteristics and informational needs of the program, not the other way around.

E. Implementation of Experimental Procedures

In discussing the implementation of summative product evaluation the experimental approach is considered as an ideal. At this stage it is necessary to emphasize, as Campbell (1969) did in his article "Reforms as experiments" that social action programs and particularly educational evaluation can be conceived of in experimental paradigms.

The implementation of the evaluation of instructional materials within an experimental approach includes a number of topics to be discussed in this section.

a) Definition of Criteria. A very important issue to be considered under the heading of criteria is the dependent variable(s) identification. The previously identified goals of the instructional materials in the phase of context evaluation permit to characterize the dependent variables. Clearly, the more specifically the goals of an instructional material are stated, the more likely that appropriate dependent variables can be identified and agreed upon. These variables are achievement-related variables which are measurable and relatively convenient to observe. However, it is also necessary to avoid the trap of measuring only that which is readily quantificable. There is always a tendency to evaluate the instructional materials solely in terms of achievement gain because the test (pretest and/or posttest) measures are available in the form of standardized achievement tests. If there are other goals, they must be evaluated even if measures need to be constructed for the specific purpose. If another material goal, for example, is to increase the ability of students to work together in

the solution of a problem, this must be evaluated along with achievement gain.

Another important aspect in the evaluation of instructional materials is the affective component, that is, the learner's affective reaction to the materials. Bronfenbrenner (1976) has pointed out that the simplest way to find out if learners like the materials is to ask them. As learners run through the instructional materials and are tested, it is not difficult to ask them a few questions, such as: "Did you find this interesting or boring?" "Was it too easy or too difficult?" "Would you like to be taught more lessons this way?". Further refinements to these techniques can be made. Information can be obtained by having each learner, after working through the materials, rate statements on Likert type scales. Typical statements would be: "I found the materials interesting" rated from "strongly agree" to "strongly disagree". Information about the learner's likes and dislikes can, in this way be quantified.

It is also important to test the affective reactions of the teacher, or whoever it is who will be present when the materials are used. The student may learn what they are supposed to, and they may like the materials very much. But if the teachers do not approve the way the learners are to be taught, then it is likely that success will be difficult to achieve.

b) Development of Instrumentation. The measuring instruments should reflect the instructional content of the materials being evaluated. The problems connected with the previous statement are not small. One of the dilemmas facing most evaluators is the question of

whether or not to use well-standardized, well-validated traditional measures or measures that are tailor-made for their particular needs. This is illustrated by the next quotation by Severy (1975): "For each article that suggests that instrumentation be standardized and traditional one finds another suggestion that new instruments be tailored to hoped for responses." However, it seems that a compromise between the two positions would bear the greatest potential for most evaluation purposes. Such a compromise would mainly involve incorporating several standardized and traditional measures to allow as much comparability as possible, while at the same time including new instrumentation to reflect as precisely as possible the nuances and differences in each unique program.

c) Scoring modes. In interpreting measurements the evaluator must discern which scoring mode is appropriate for which purpose. Raw scores and grade equivalents are essentially absolute scores; percentiles and standard scores reflect achievement levels relative to those of all test participants. The scoring mode depends on the purpose of the evaluator.

e) Appraisal of the Evaluation. It is important to ask if there has been a good job of evaluation done. Tallmadge and Horst (1974) have developed a procedural guide that involves 23 steps towards validating the effectiveness of educational programs using existing evaluation data. The guide is constructed to allow "branching" and a particular answer to any one question leads to another question in the guide. The complete procedure is not presented here, but perusal of the list of questions should indicate what would be classified as a

good evaluation. The steps in the procedure involve the following questions:

1. Are the test instruments adequately reliable and valid for the population being considered?

2. Are pre-or post-test score distributions of any groups curtailed by ceiling and floor effects?

3. Is there reason to believe that the pre-testing experience may have been at least partially responsible for the observed experimental outcome?

4. Is there reason to believe that knowledge of group membership may have been at least partially responsible for the observed experimental outcomes?

5. Is there reason to believe that student turnover may have been partially responsible for the observer experimental outcome?

6. Does the evaluation employs a control group?

7. Were pre-test scores used to select the treatment group?

8. Are normative data available for testing dates which can be meaningfully related to the pre and post testing of the program pupils?

9. Do the norms provide a valid baseline against which to assess the program of the treatment group?

10. Is the comparison between the treatment group and the control based on pre and posttest scores or on gain scores?

11. Have appropriate statistical test been employed to assess the significance of the gain in treatment group performance relative to the control group?

12. Are pre and/or posttest scores available?

13. Can appropriate statistical tests be employed to assess the significance of gain in treatment group performance relative to the control group?

14. Were the students, either matched or unmatched, randomly assigned to the experimental and control groups.

15. Is there evidence that members of the experimental and control groups both belong to the same population?

16. Are post-treatment comparisons made in terms of posttest or gain scores?

17. Can data be obtained which enable application of analysis of covariance techniques; would such analysis be appropriate; and is there reasonable expectation that they would produce significant results?

18. Is the control group superior to the experimental group on the balance of educationally relevant variables?

19. Have covariance analysis techniques been employed to adjust for initial differences between groups?

20. Have appropriate statistical tests been employed to compare posttest and gain scores?

21. Can data be obtained which would enable appropriate test to be made?

22. Do analysis results favor the treatment group at the preselective level of statistical significance?

F. Dissemination

Tyler and Klein (1973) provide ample support for the need to develop a practical system for the collection and dissemination of instructional

materials evaluation information. Alkin and Fink (1974) mention that the preparation of reports oriented to needs, requirements and instructional contexts of school administrators, curriculum and instruction personnel, and teachers, has been neglected.

There are several immediately obvious reasons for the failure to provide adequate user oriented reports on instructional materials. Consumers do not demand contextual summative evaluation information. Potential users have not become sufficiently sophisticated to call for and expect comprehensive and understandable information before making their decision of purchase a determined material. User decisions are most commonly based on the reputation of the commercial publisher or the presumed academic qualifications of the product's author (Tyler, Klein, & Michael, 1971).

The problem of establishing standards for product development reports is not new (Lumsdaine, 1965). However, with the increasing number of available products, the problem has become more obvious. One document concerned with the user's needs (Tyler, Klein, & Michael, 1971) contains a set of recommendations for curriculum and instructional materials to guide producers, users, and funding agencies. The recommedations are intended to guide producers in the development of instructional materials; consumers in the selection of instructional materials; and funding agencies in the evaluation of instructional materials. This part is concerned with the description of what minimum information is needed to be included in an evaluation report, mainly directed to the users of the instructional materials.

The minimum information that a report should provide is.

(Alkin and Fink, 1974):

a) The material's description

b) The material's purpose

c) The intended population

d) The material's development and testing

e) The material's effectiveness

f) The material's efficiency

a) Material description. The information described in this category is concerned with the name, developer, distributor (if any) andphysical attributes. From these descriptive characteristics, the user is provided with indicators of how the product is likely to fit into his instructional context. For example, the title of the product might provide the user with a basis for deciding whether or not the subject matter content is relevant to his needs. The material developer could be known for a particular style of instruction that may or may not be of interest (e.g., self-paced or group instruction). The physical attributes of the material might help the user determine the equipment and whether the product include additional or complementary materials as workbooks, guides, etc.

b) Material's purpose. One of the first questions to be asked by potential users about the instructional material is "what can my students learn from the materials?" Stated another way, the user wants to know what the product's criterion goals and objectives are so that he can determine whether they are congruent with those previously identified as important in his particular instructional context. Thus

in education, a product that contains a set of precisely stated objectives is more valuable to the potential user than one that consists only of broad statements of instructional intent. Precise statements of instructional objectives not only permit users to assess whether a product is appropriate, but also enable them to judge the extent to which it can fulfill the requirements of a given instructional situation.

c) Intended learners. An adequate description of the products' intended learners should specify the background characteristics of learners who might be expected to achieve the goals and objectives of instruction, based on the product's success in field tryouts with similar learners. These specifications should include such factors as age, sex, and pre requisite knowledge and skills.

d) Materials development and testing. A least two types of information must be provided for the user to help him make a sensible choice: 1) An explanation of the procedures actually employed to develop and test the materials to make it "work", and 2) a review of the empirical evidence that demonstrates the extent to which the product effectively promotes learning.

Explanations of materials development and testing provided for users should be convincingly organized to demonstrate that the product has been systematically designed to achieve its purposes in actual instructional situations. Such explorations should not be made equivalent to technical reports or case studies reporting each of the product's historical problems, since this would probably distract, or even confuse. Instead, a brief summary of the techniques employed to

prepare the materials and to assure consistency among its objectives, instruction, and assessments, should be given.

e) The material's effectiveness. Once the user has determined whether the product aims at achieving significant goals for students similar to those for whom he is responsible, he wants to know how effectively the product performs. Basically, he needs data about the success of the material. Such data should be displayed and summarized so that it is readily interpretable. The type of data will depend of the type of design utilized in the field trial. The data to be provided could include the students' average score for each objective; the number (or percentage) of students who achieve a given objective, and the number of students who achieved all or a given percentage of objectives.

f) The material's efficiency. Efficiency considerations require that a developed product should either demonstrably increase the number of students who are successful with the product within a given time period or drastically reduce the cost of maintaining the current success rate. An additional consideration related to an instructional material's efficiency is its exportability. Potential users can benefit significantly from information about the product's ability to yield reliable student outcomes given specified instructional procedures, but also assuming some inevitable deviations from these procedures in many school situations. The latter circumstance is particularly important because of the rarity with which instructional programs are implemented exactly as planned. Thus it would be an advantage for the user to have some idea of the product's flexibility prior to its

acquisition and installation.

CHAPTER VIII

EVALUATION OF THE MODEL

The validity of the proposed model will be determined through two strategies. First, an analysis of the rationale of the model will be discussed, using the systems approach and problem solving methodologies and secondly a plan for its empirical validation through a field testing of the model in an operational environment will be presented.

A. The Systems Approach

Systems approach can be defined as a process by which needs are identified, solutions are chosen from alternatives, methods and means are obtained and implemented, results are evaluated, and required revisions to all or part of the system are made so that the needs are eliminated (Kaufman, 1972). Systems approach is a type of logical problem solving (i.e. Checkland, 1976; Kaufman, 1972; Robertshaw et al. 1978).

Applied to educational problems it helps to identify and resolve important problems, used both as a process tool for more effectively and efficiently achieving required educational outcomes, and as a mode of thinking that emphasizes problem identification and problem solution. It utilizes a formulation of logical problem solving steps generally presented as a model.

The word model in a general sense, is a description or prescription of a thing in existence or being planned. Ackoff et al. (1962) have pointed out that in everyday language the term model has at least three different usages, as a noun, as an adjective, and as a verb. As a noun model implies a representation, as an adjective implies an ideal, as a

verb to model means to demonstrate. Ackoff et al. (1962) suggests that part of each three meanings of the word model has been incorporated in model building. A model, then, is created as an idealized representation of a real or planned system.

The Input-Output System Model. The best known system conceptualization, and thus far possible most useful in this analysis, is that of the input-output system model. In fact, the fundamental premise of all systems thought is rooted in the input-output processing relationship (Immegart and Pilecki, 1973). The basic notion of system, as revealed by all forms of systems theories, is built on the input-output relationship. This conception of the system deal with the transformation of inputs through processing subsystems into outputs, and the resulting effects of output on subsequent input and action through feedback and output evaluation (Ashby, 1956). The "black box" model of an open system illustrates this idea graphically (see Figure 4).

According to process conceptualization of systems inputs in the generic forms of operands (those inputs to be processed) and operators (those inputs which are to do the processing) are transformed through functional subsystems into outputs (outcomes or products of system action), (Kopstein, 1956). Results of system action are invariably evaluated in open systems and such information, as feedback, is channeled back into the system and affects future system activity.

In the case of the model for evaluating instructional materials, more than one subsystem is used in transforming input into output. Figure 5 illustrates how several subsystems are linked in the model. Each of the processing subsystems is in fact, an input-output system in

Figure 4 The "Black Box" model of the Input-Output System.

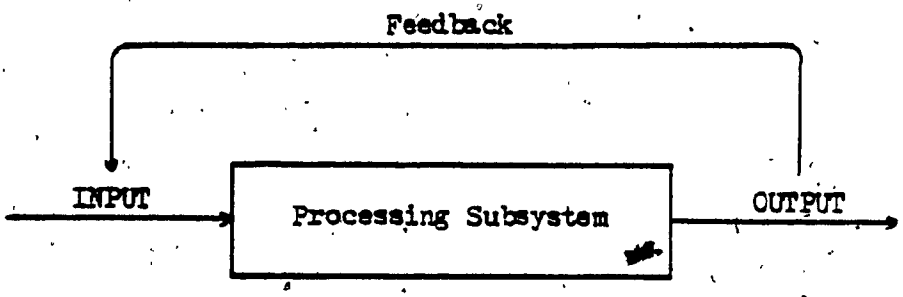
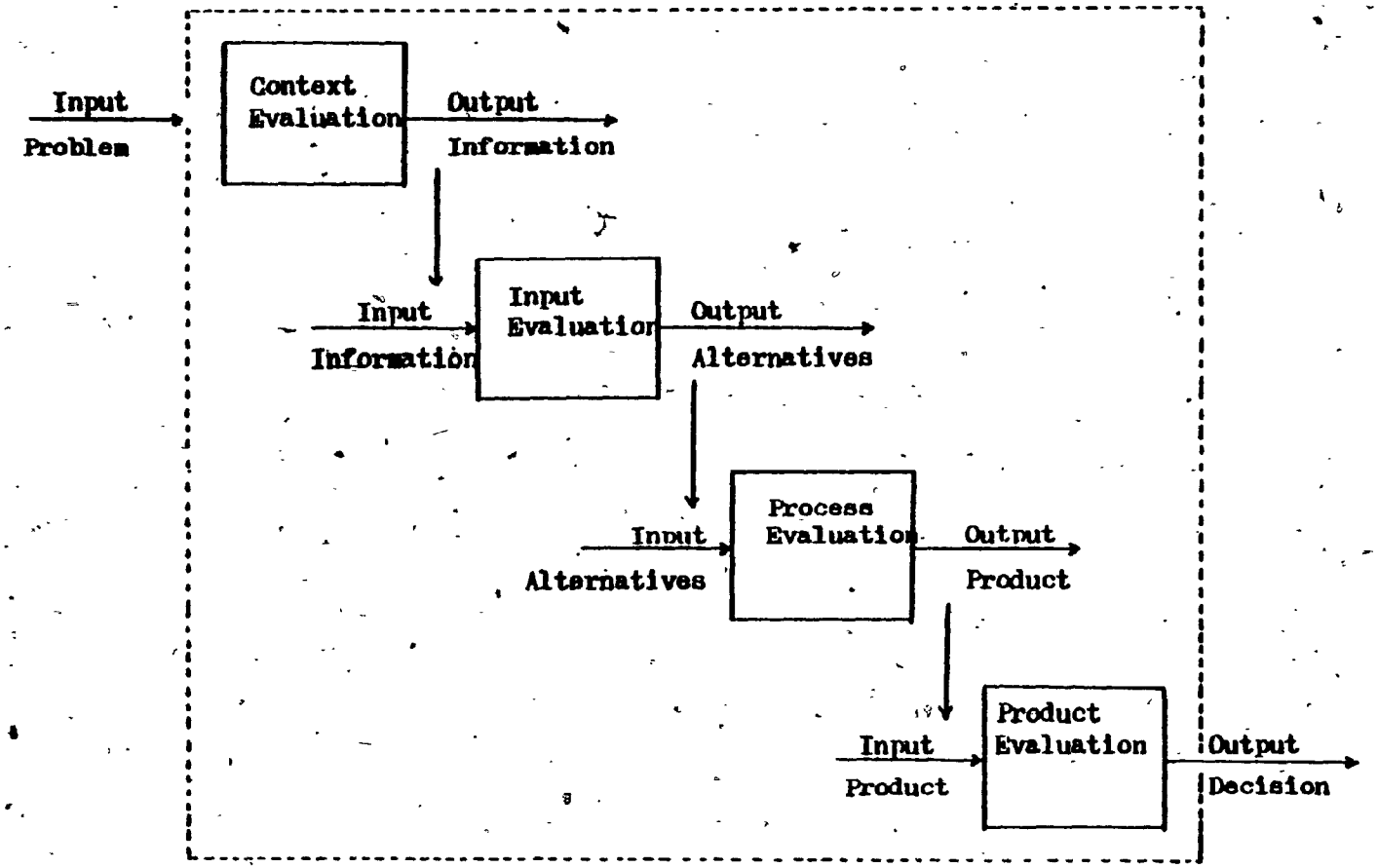


Figure 5 Linkage of Subsystem Input and Output



its own right. This is in turn presented in Table 5. All but the most restricted system activity can, therefore, be subdivided into definable and linked input-output processing subsystems. Further, as seen in Table 5, all subsystem output is linked directly to, or becomes, input for new subsystem functioning.

Also, as noted above, there are both operand and operator categories of input. In an evaluation situation, as is the case of the model input consist of a problem, data, and alternatives which must be operated on, as well as people, analysis and evaluation procedures. Likewise, output is also multidimensional and can be categorized generically in terms of productivity and affectivity. In other words, decisions are productive outcomes of decision-making activities, just as work done is a resultant of work activities. In each case, system activity affects the people in the system (or in its subsystems) and those in the system's environment. That is, decision-making or work is succesful or unsuccesful and people from the system or the environment feel good or bad about the outcomes of such activities. This is illustrated in Figure 6.

Feedback, which can logically be considered an output of system activity, can also be viewed in greater detail in terms of its overall effect on system functioning. Not only is there internal system feedback or evaluative information occurring within the system but also, since they exist and function within an environment, open systems characteristically receive external feedback resulting from environmental evaluation of system action. Further, both internal and external feedback affect future system action in terms of effects on

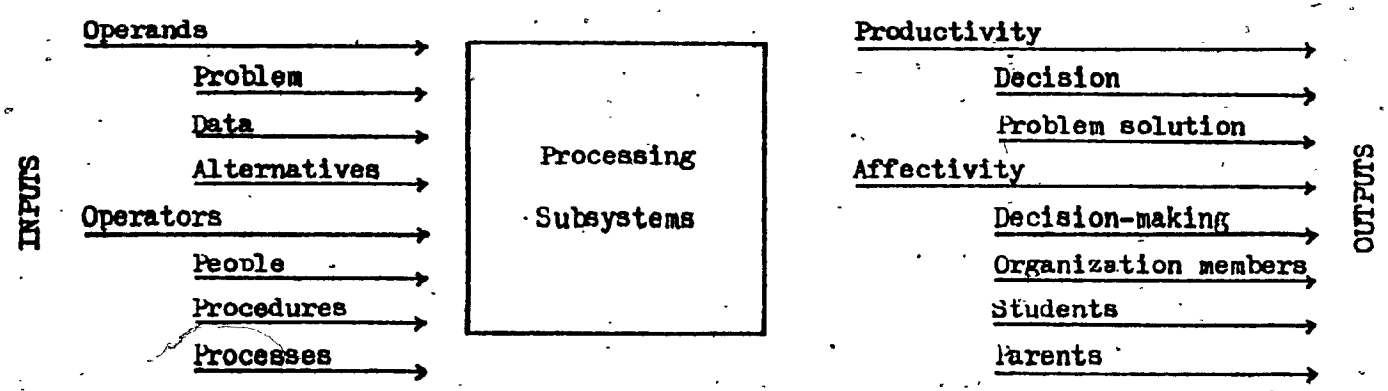
Table 5 Input, Process, Output Subsystems of the Instructional Materials Evaluation Model

	Input	Process	Output
Context Evaluation	Information about school conditions	Analyzing processing information	Listed needs
Needs Assessment	Listed needs	Identifying priority needs and goals.	Priority goals
Input Evaluation and Materials Seleccion	Priority goals and information about instructional materials	Selecting among several instructional materials	Identification of instructional materials according to the goals
Determination of Resources	List of selected instructional materials	Deciding which instructional materials will be developed according to disponible resources	A decision: to continue or to turnback to select other materials
Instructional Materials Development	Information about the decision and type of instructional materials to be developed	Application of the technical processes in the development of instructional materials	Adoption, adaptation or production of instructional materials
Process Evaluation	Information about the selected alternatives	Monitoring the the development of instructional materials	Final product

Table 5 (Continue)

	Input	Process	Output
Product Evaluation	A produced instructional material	Determining the worth of the product	Identification of the strenghts and weaknesses of the instructional material
Fate of Program	Pros and cons of the product	Deciding whether continue discontinue, expand or revise the program of development of instructional materials	Decision

Figure 6 The Multidimensionality of input and output in the Problem Solving System.



(Immegart & Pilecki, 1973)

input, and system or subsystem structure and processes. Figure 7 illustrates these kinds and functions of feedback.

B. The Systems Approach and Problem Solving

The systems approach is a model for problem solving (Robertshaw, Mecca, & Rerick, 1978). It consists of a non-algorithmic iterative series of overlapping processes: defining the problem. generating alternatives, and evaluating alternatives; this is illustrated in Figure 8.

In the definition of the problem, we consider what it is we want to do, how we will judge a solution (or part of a solution), and what constraints we must accommodate. The generation of alternative solutions produces descriptions of actions which may solve the problem. Finally, the possible solutions are evaluated. The three basic processes involve modeling. The models are generated by the questions associated with the basic steps of the systems approach. For example, questions associated with these processes in the context of evaluation of instructional materials include:

Definition of the Problem:

What is to be evaluated?

What must be accomplished?

Who is the decision maker?

What is his value system?

How will he pick among alternatives?

What are the constraints?

Figure 7 The Input-Output System Model, Feedback Processes.

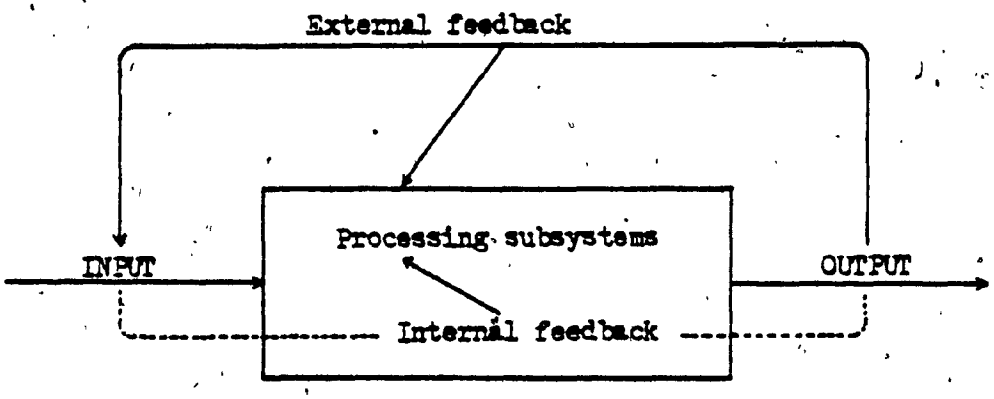
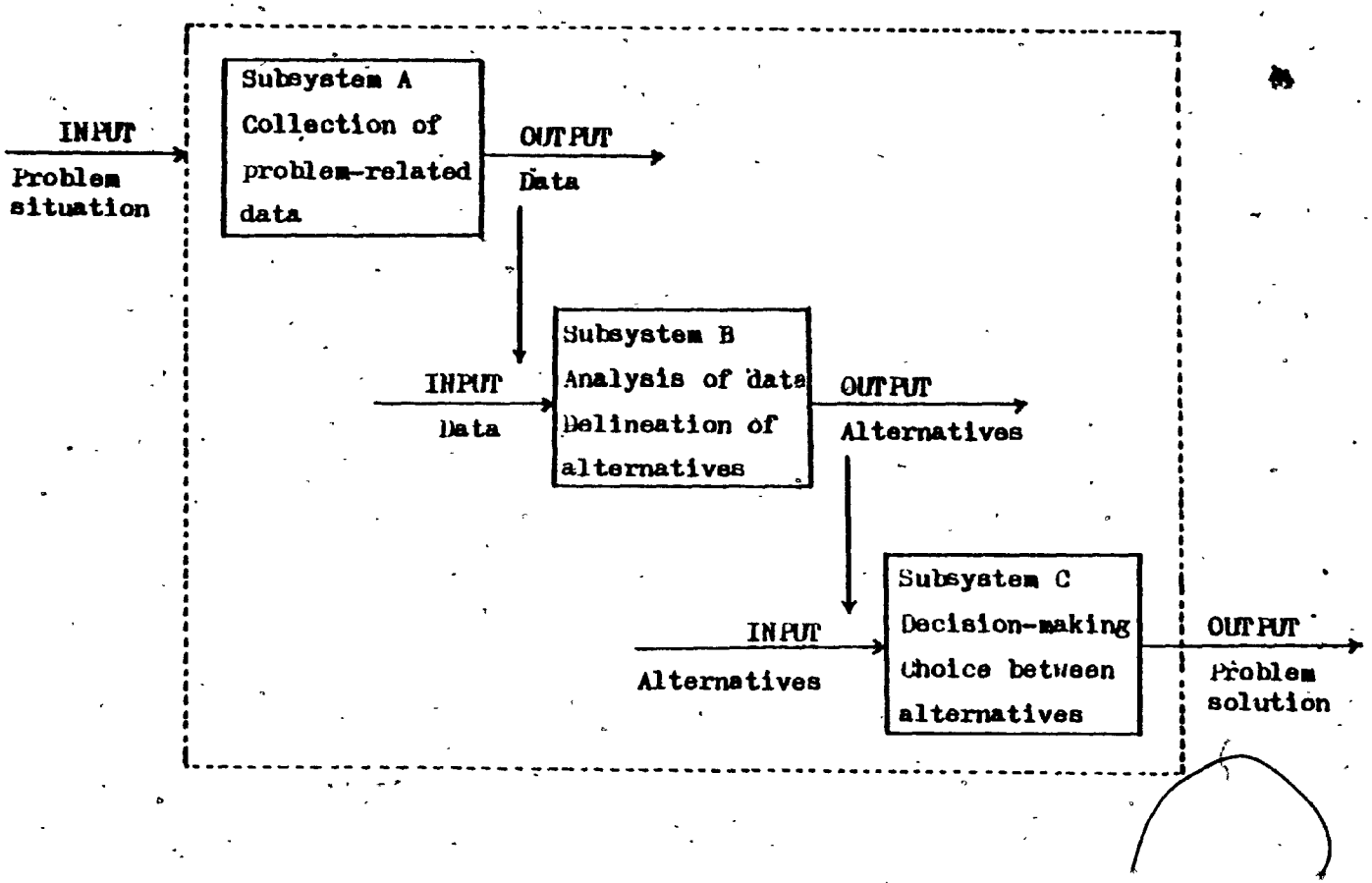


Figure 8 The Problem Solving System



Generation of Alternatives:

- What are the alternatives?

  How these alternatives will operate under the conditions of the problem?

  How much do they cost?

  What will they produce?

Evaluation of the Alternatives:

  Which alternative will be picked?

  What are the factors affecting the worth of each alternative?

The proposed model for evaluation of instructional materials follows the same conceptual rationale of the problem solving process under the systems approach, but its structure is based in the Stufflebeam's CIPP Model. Each one of the elements (Context evaluation, input evaluation, process evaluation, and product evaluation) is a subsystem which produces specific outcomes.

C. Plan for the Field Testing of the Model

The suggested model for the evaluation of instructional materials has been presented in the spirit that its effectiveness can and should be documented. To test its effectiveness and to record discrepancies between the model and events as they occur in the real world a plan for the field testing is suggested; however actual field testing was not carried out.

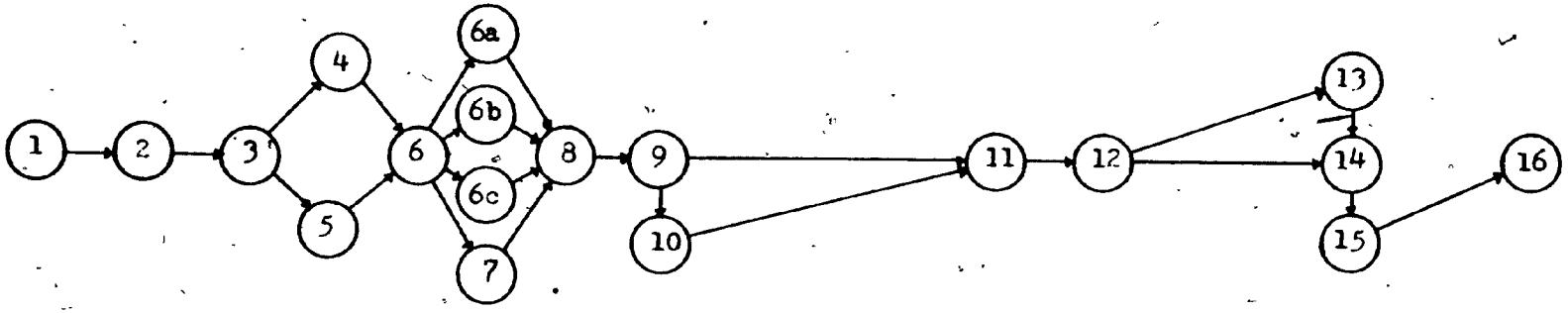
The context for the field testing of the model is given by the program of "Human Resources Development" currently being implemented in the "Centro Latino Americano de Tecnologia Educacional para la Salud", located in México City. This program has as a major goal to design and develop training courses and workshops directed to the health sciences personnel currently involved in educational activities in health institutions and universities throughout the country.

The types of contents comprised in the instructional activities include topics such as instructional design, learning evaluation, design and production of instructional materials, curriculum design, and educational administration.

The model will be tested using the curriculum design and development of a 50-hour course focussed to the training of nursery instructors in the use of group dynamics techniques in the classroom.

Figure 9 shows the summary network of the activities included in the plan for the evaluation of the model. Figure 9 also indicates a time scale that represents the estimated time required for the accomplishment of each activity and the total project.

Figure 9 Summary network for the field testing of the model



Days

1------ 5----- 10----- 16---- 22---- 23---- 26----- 28---------- 50------ 51---------- 75---------- 85

Event Identification

1. Project Start - Organization of Team.
2. Start Context Evaluation.
3. Start Needs Assessment.
4. Start Definition of Goals and Objectives.
5. Start Input Evaluation.
6. Start Instructional Materials Selection.
   6a. Adoption
   6b. Adaptation
   6c. Creation
7. Start Determination of Resources.
8. Start Process Evaluation.
9. Start Instructional Materials Development.
10. Start Formative Evaluation.
11. Complete Formative Evaluation.
12. Start Product Evaluation.
13. Start Determination of Fate of Program.
14. Start Analysis of Results.
15. Start Report of Results.
16. Project Complete.

A team of 2 curriculum developers, 2 substantive advisors, and 1 evaluation advisor will be organized for participation in the curriculum development effort. Participants will follow each one of the phases proposed by the model. Table 6 presents a description of the tasks, some illustrative questions, and expected results for each one of the activities summarized in Figure 9.

The final product will be an evaluative report focussed mainly to the analizis and comparison of the actual processes and results in order to determine the discrepancies with those proposed by the model. The report will describe and examine the discrepancies between anticipated and actual accomplishment, determining whether each one of the phases and the total model adjust to the reality of the curriculum development effort.

Indicators of the effectiveness of the model on which measures may be obtained during the project are: (1) reports of the development team; (2) the attitude of the organization's administration; (3) the use of information by the development team and organizational administration; (4) the degree to which evaluation was used in decision making; (5) the degree to which the materials meets the performance standards; and (6) the general acceptance of the model.

On the basis of the first application of the model further cycles of improvement and refinement will be undertaken.

Table 6

Tasks, Illustrative Questions, and Expected Results of the Field Testing of the Model

1. Project start - Organization of the team

Tasks:	Illustrative questions:	Result:
Organization of the team.	What is the purpose of the organization of such a team?	A five-member team including: 2 curriculum developers 2 substantive advisors 1 evaluation advisor

2. Start context evaluation

Tasks:	Illustrative questions:	Result:
Provision of baseline information. Determination of target population.	What are the general goals of the project?	Identification of constraints and determination of a plan of activities. Definition of the goals of the team, responsabilities and work schedule.

3. Start needs assessment

Tasks:	Illustrative questions:	Result:
Identification and prioritization of needs, via literature review and consultation with subject matter specialists.	What are the needs to be met? Is this problem a high priority issue?	A hierarchized list of instructional needs.

(Table 6 continued)

8. Start process evaluation

Tasks:	Illustrative questions:	Result:
Definition of preprogrammed decisions.	Should the project be reoriented prior to completion?	List of preprogrammed decisions.
Establishment of procedures for the collection of data.	Is the project on schedule?	Procedures for the collection of data.
Establishment of a methodology for identifying deviations from specified procedures.		Methodology for identifying deviations.
Comparison of actual and scheduled completion dates for project events already completed.		

9. Start instructional materials development

Tasks:	Illustrative questions:	Result:
Development of the instructional materials.	Does the materials being developed adjust to the planned procedures?	Instructional materials (or part of them) ready for tryout with students.
Monitoring of materials being produced.	Is the technical quality adequqted?	

(Table 6 continued)

4. Start definition of goals and objectives		
Tasks:	Illustrative questions:	Result:
Specification of goals and behavioral objectives of the course.	Are the objectives clearly stated?	List of goals of the course.
Design of the learning evaluation instruments.	Which of the stated objectives are most feasible to achieve through the use of instructional materials?	A hierarchized list of objectives of the course, stated in precise terms related to observable phenomena.
Development of the instructional design of the course.	Is there a sound rationale for the measures selected?	Evidence that measurement techniques have desired psychometric properties.
5. Start input evaluation		
Tasks:	Illustrative questions:	Result:
Identification and analysis of strategies for selecting and developing instructional materials that are already operating in similar institutions.	Does the strategies satisfy the criteria for implementation of plans?	Plan for the selection and development of the required instructional materials.
Design of a plan for the selection and development of instructional materials for the course.	What specific strategies and time schedule will be needed to implement the plan?	

(Table 6 continued)

6. Start instructional materials selection

Tasks:	Illustrative questions:	Result:
Identification of the required instructional materials.	What type of materials are necessary to fill the needs, goals and objectives?	List of the selected instructional materials.
Determination of instructional materials to be adopted, adapted and/or created.	Are there materials available which might adopted and/or adapted to the purposes of the course?	Acceptance by substantive advisors of proposed instructional materials.
		List of instructional materials to be adopted, adapted and/or created.

7. Start determination of resources

Tasks:	Illustrative questions:	Result:
Determination of budget to acquire, adapt and/or produce the required instructional materials.	Are there sufficient resources (money, technical and time)?	List of material unit costs for first version.
	Are alternative, less costly versions of the materials possible?	Ordering of priorities in the development of the materials.
	Are there opportunity costs?	

(Table 6 continued)

10. Start formative evaluation

Tasks:	Illustrative questions:	Result:
Revision of parts or completed instructional materials. Application of procedures for acquire feedback from students and instructors.	Is the instructional design of the materials appropriate to the target population? Are there unanticipated consequences? Can the instructors use the materials in the intended manner? Are the objectives being met?	Evidence that the materials are generally usable. Evidence that the materials have been revised to attend to discovered logistical difficulties.

11. Complete formative evaluation

Tasks:	Illustrative questions:	Result:
Modification of the instructional design and procedures of the materials based in the feedback provided by students and instructors.	Have all the findings in the revision and tryout of the materials been taken into account for further modifications?	Final versions of the instructional materials including guidelines for its use. Evidence that target population could use the materials.

(Table 6 continued)

12. Start product evaluation

Tasks:	Illustrative questions:	Result:
Selection of the experimental design and related instruments and procedures for the testing of the instructional materials. Determination of the instructional value of the materials.	Are the experimental design and the procedures adequate? Are the sampling procedures and the sample size adequate?	A report which will incude information concerning to the results of the experimental tryout of the materials. An evaluation report directed to the users, identifying characteristics, virtues and of the materials, specifying the data and analyses on which conclusions were based.

13. Start determination of fate of the model

Tasks:	Illustrative questions:	Result:
Determination whether the model for the evaluation of instructional materials will remain within the curriculum design and development effort or whether it is necessary to change the strategies.	Are there discrepancies between the activities of the proposed model and the events in the reality?	To continue, revise or discontinue the activities proposed by the model.

(Table 6 continued)

14. Start analysis of the functioning of the model and 15. Start report of results

Tasks:	Illustrative questions:	Result:
Determination of the usefulness, applicability and validity of the model in real-world settings.	How useful was the model, both as a framework and as a guide, within a real-world setting curriculum design and development effort? How many and which were the discrepancies between the model and the events in the reality?	A report describing the results of the application of the model, its pros and cons, and suggested modifications.

16. Project complete

CHAPTER IX

IMPLICATIONS AND LIMITATIONS OF THE MODEL

A. Consideration of Costs

Educators are not econometricians and should not be expected to be skilled in identifying all the financial, human, or time costs associated with programs they operate; that bit of leniency ideally cannot be extended to the evaluator whose job it is to bring these factors to the attention of developers, teachers and administrators who are responsible for their products or programs. Educators are often faulted for choosing the most expensive program from two that are equally effective, just because the expensive one is packaged more attactively or has been widely advertised. The real fault lies with the evaluation of those programs which fail to consider cost factors along with the other variables. As any insighful administrator knows, costs are not irrelevant, and it is important to know how much "Program X" will accomplish and at what cost. Otherwise, administrators have no way to know what they are gaining or giving up in looking at other alternatives which range in both cost and effectiveness.

In the determination of the cost evaluation two main considerations must be taken into account. Costs data must be comprehensive and verified (Borich, 1974).

Cost Data Comprehesiveness. This means covering maintenance as well as capital costs, psychic as well as dollar costs, and costs of in-service updating of needed helpers as well as direct costs. There

should be some consideration of opportunity costs other than those covered previously under critical competitors, and a pass should be made at cost-effectiveness analysis where possible.

Cost Data Verification. Cost estimates and real costs should be verified independently. It is really not satisfactory to treat cost data as if they are immune to bias. The cost data require verification for reasons that have not thus far been generally recognized. Costing is an extremely difficult business, requiring technical skills that at the moment are not part of the training of evaluators (Borich, 1974). As a matter of fact, there are plenty of educators who are quite incompetent at estimation of costs in evaluation studies. However, the advice of one or more CPA's is certainly required in costing large projects.

Cost evaluation of instructional materials according to the proposed model has two main components:

- the investment in the development of the course itself, and
- the basic cost of running the course

The cost evaluation needs to be looked at differently in relation to the two different costs listed.

Inevitably, different instructional materials have widely differing development costs. The more expensive the medium, the more it is necessary to know how the medium works and how effective it is. Heavy development costs of particular media can only be justified either if the educational gain is overwhelmingly greater than an alternative method, or if the material is likely to reach such a large audience or used repeated so often that the set-up cost is spread over such large

number, so as to be justified over time. If a course is to be run once only, then the development cost will be great in relation to its other costs. Theoretically, it would made sense to research into it at this stage. On the other hand, if the course is not to be re-run then no action is likely to take place as a result of the evaluation and there is therefore no point in carrying it out. Additionally, heavy set-up costs may not be justified for subject matter which is likely to become out of date soon.

The second cost, that of running the course, will be a recurring cost and money spent on researching the operation of the course to reduce this recurring cost may well be cost-effective, particularly if the course is flexible enough to allow amendments to be made before it is re-run.

The variable cost associated with the amount or personel involved is more complex. It is helpful to look at the variable cost associated with the number of persons supporting the development and evaluation activities of the course. The amount that is worth spending on evaluation of the instructional materials depends on the proportion of the cost of the course that is "variable" cost, i.e. The more complex and sophisticated the materials, the more this cost is likely to be.

The "value" of the evaluation is difficult to quantify since, as has been described , for the purpose of the evaluation of instructional materials, evaluation is a component of the instructional design rather than a separate activity.

B. The Role of the Evaluator

Evaluation is not usually seen as an explicit and systematic part of the process of curriculum development. This means that the use of a more formal system of evaluation rather than the informal daily evaluation based on no more than opinions, could promote action, rather than reaction.

The role of the evaluator, based on the previous suggestion and according to the model for the evaluation of instructional materials is defined by the range of activities that he must develop within the curriculum development team.

Evaluation can play many roles in curriculum development and implementation. One role for the evaluator emphasized here concerns what has traditionally been denied by curriculum developers. This involves the identification of information about demands, assumptions, values, and beliefs of the population to which the course and the materials are directed, and to present this information in a manner useful to the curriculum developer and materials designers in an attempt to build interesting and useful instructional materials. The evaluator as a part of the curriculum development team, provides throughout the developmental process the incentive and capability to act, rather than only react.

There have been many arguments for the evaluator to take on specific roles. Guba and Stufflebeam (1970), for instance, while noting several different activities in which evaluators are engaged, argued that the primary role of the evaluator should be to provide information to the decision makers. The evaluator, thus, should be a specialist

who identifies, collects, analyzes, and reports information. The role of making summary judgments is left to program administrators.

Stake (1967) implicitly suggested that the evaluator is the one who actually should make the judgments. Following his model, a set of conditional (that is, "If these conditions exist then this is best") statements should be offered by the evaluation specialist. Ideally, from this viewpoint, the output from an evaluation study would include: 1) a complete description of all important variables operating in and on the program and 2) a set of judgmental statements. If this is done completely, then any rational person should be able to reach the same conclusions about which course of action is best.

Other writers, influenced by the political nature or non-rationality of much administrative decision-making have argued that an evaluator should be given the responsability of making decisions based on the results of his evaluation study. The rationale for this argument is that the evaluator is the one person who has become knowledgeable about all the important aspects of the program.

There are a number of roles the evaluator can play in the performance of his task. He may be a raiser of questions; he may collect and interpret data; he may serve as a judge. In the model the evaluator collects, throughout the developmental process data about the instructional needs of the population to which the course is directed; he participates in the establishment of the goals and objectives; he analyzes the criteria for the selection of the instructional materials collecting information about its characteristics and aiding the final decision; he applies methodologies and techniques for obtaining information for formative purposes during the stage of

design and production of the materials and finally he develops evaluative research of the instructional effectivity and psychological effects of the products.

Perhaps the most difficult task facing the curriculum development team is the integration of accountabilities to create materials within the curriculum that are most responsive to the audiences that are affected by them.

There are two disparate viewpoints referred to who should be charged with the responsability for conducting the evaluative activities. One of them is the popular view that external evaluation is important. Scriven (1972) suggests that the formative evaluator come in without knowledge of the program goals and provide his or her assessment of the actual effects of the project. He calls his approach "goal-free evaluation". Scriven's approach is that of an auditor who comes to see what the effects of a program are without influence generated by the rhetoric supplied by program plans, specifications, and the like. Stake (1972) portrays the "responsive evaluator" as one who observes and reacts, with fewer quasi-scientific trappings than typical evaluators carry. The responsive evaluator should , according to Stake, serve to stimulate project staff rather than to provide only the usual sorts of "data for decisions". Stake and Scriven both see externally based formative evaluators making serious contributions to program improvement. Scriven sees the evaluator as a dispassionate outsider providing a circumspect preview of summative findings. Stake, while also rejecting the all-constraining influence of program objectives, expects the evaluator to get to know his clients well, to design his reports in attention-arresting ways, to become closely interactive

with the project itself. As "external" evaluators, they presumably avoid pitfalls, such as serious financial dependency on the project and inordinate predisposition to see good things emerging in any case.

On the other hand, because of convenience, intimate knowledge of program capabilities, and location, evaluation may often be conducted internally by trained development personnel; that is the approach supported in this thesis. This approach represents the view that the evaluator and developer cannot function without interrelating the needs and requirements of each other's tasks (Butman & Fletcher, 1974). To do this it is suggested an approach in which the developer and evaluator work as a team members for a common cause, that of developing the best possible instructional materials.

The assumption to support this approach comes from the fact that the developer/evaluator team has much more power than the evaluator who is located in a division separate from development. Evaluators who betray excessive concern about not getting co-opted misunderstand their role, particularly the formative evaluation role, and misunderstand the nature of the development process. When evaluators are separated from the development process, they cannot maintain contact with the shifts and changes so characteristic of development work. The work of the isolated evaluator is usually ignored in the natural process of development because the developers have not been part of the process of deciding what they needed to know and how to get it.

C. Relation of the Model with the Instructional Design

Instructional design models are composed of sequential steps which lead to the production of instructional materials (Dick, 1974). The steps of instructional development most often include these phases: identification of goals and objectives, design of instructional materials, development, production, and evaluation (Briggs, 1970). The rationale often used for placing evaluation last in the development process is that it provides a method for revising the various curriculum and instructional components prior to dissemination. Usually, though, the repetitive re-cycling loop at that point is too late and cumbersome for cost-effective revisions. Several authors recognize the need for activities that provide for evaluation of the feasibility of the instructional materials, evaluation during the formative design and development phases, and continued evaluation after implementation (i.e. Grobman, 1968; Guba & Stufflebeam, 1970; Saylor & Alexander, 1974; Stake, 1969).

Evaluation and development involve concurrent activity in an iterative process controlled by feedback of information gathered at each step. This information determines the portion of the output of a previous step that should be altered in light of new information.

As the model has been presented the resemblance with the general steps of several models for the design and development of instructional systems is obvious (Briggs, 1970; Dick & Carey, 1978).

A characteristic fundamental to these models is the use of the basic approach of the scientific method; this usually involves a statement of a hypothesis or problem, experimentation with solutions,

and correction of hypothesis of solution if needed. These models generally are composed of a set of interrelated steps that might be represented in a diagram such as Figure 10 . The diagram represents an application of the systems approach to the development of instructional systems that used an iterative process controlled by feedback at each step.

In the systems paradigm, evaluation is an integral part of instructional design. The systems approach demands that evaluation be carried out.

The design and development of instructional materials within an instructional design effort must be one of the components to be evaluated in an ongoing process in each one of the steps.

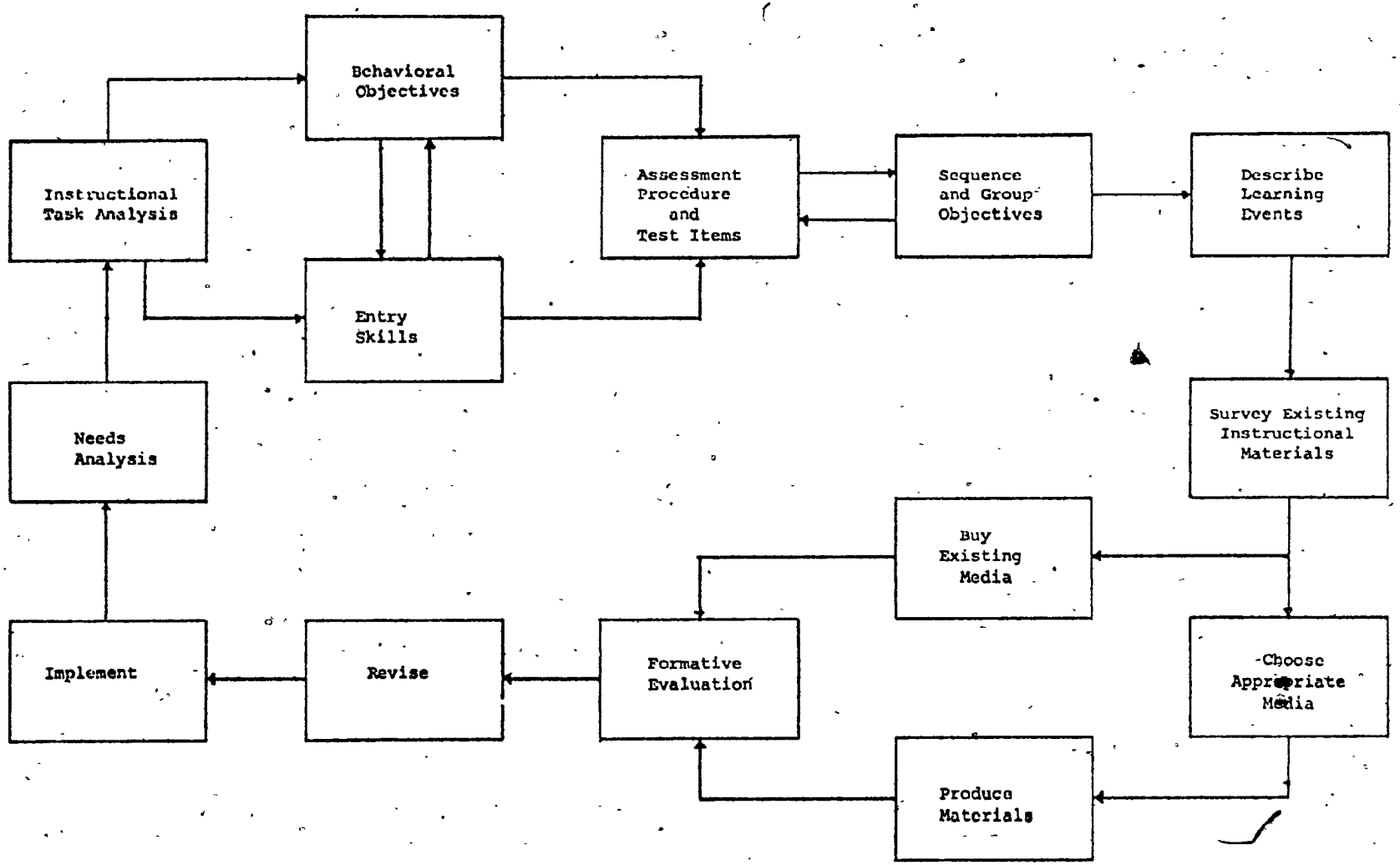


Figure 10 Designing and developing instructional systems (Brien & Nelson, 1977)

CONCLUSIONS

Assumptions About the Applicability of the Model

Several assumptions underlie the instructional materials evaluation model presented in this thesis. These assumptions should be considered before attempting to apply the suggested model to any given effort of instructional development. They are:

1) The development of instructional materials implies a iterative and long process of coordinated efforts. This process can be shortened by efficient planning and coordination among its components. However to skip some of the stages of the developing process may lower the quality of the product. The educational system must realize that the development of instructional materials should be allowed to benefit from long-term investments and adequate planning.

2) The development and evaluation of instructional materials must be a cooperative task of a group of persons with a wide range of expertise. Generally the development and evaluation of instructional materials is a task of individuals or small groups allowing the production of biased materials, that is, when a single author and producer or a very reduced group of people design and develop a specific product one is likely to find pitfalls in the subject matter, teaching methods, evaluation procedures, etc. An integrated team of specialists should be involved in the development of instructional materials. Such a team should replace the traditional single-author approach if significant and reliable instructional materials are to be developed.

3) Formative evaluation should be an integral part of the development process. This assumption has two important implications. First, evaluation has to be conducted from the beginning of the developmental process rather than at the end of this process. Secondly, formative evaluation should serve the needs of the developmental team rather than assess its effectiveness for external audiences. Efforts should be made to reduce the threat of evaluation by demonstrating its constructive role in the developmental process.

4) The evaluation team should not be the only one involved in evaluation activities within the developmental process, although they should bear the main responsability for the conduct of evaluation. It is not feasible to assume that the evaluation team can serve all evaluation needs , nor is it desirable that evaluation be perceived as something that belongs only to the evaluation specialists. This assumption implies that curriculum developers have to be trained in evaluation theory and techniques to gain basic skills and positive attitudes toward evaluation.

5) Decision makers throughout the all levels of the development organization or project should be commited to the utilization of evaluative information in their decision making process. This does not imply that all the decisions are based only on rational considerations, nor does it imply that all recommendations resulting from the evaluation must be adopted. However, decision makers should strive to increase the rationality of their decisions and be ready to justify those that contradict the evaluative information. In such cases the confrontation

of the decision maker with the evaluation information might result in a significant contribution to the developmental process, even though the evaluation recommendations are rejected.

General Comments

Traditionally, the expectations and the demands of most school purchasers regarding instructional materials have centered on factors such as the product's physical construction, its cost, and its content, rather than on its direct effects and discernible affects on learners and teachers. A survey of materials selection practices conducted by Educational Products Information Exchange Institute (EPIE, 1969), after examining construction and cost factors, concludes that selection groups invested most of their prepurchase time and energy in checking the content (i.e., validity for and compability with the local curriculum, educational philosophy, and social values) of materials under consideration. The study also indicated, however, that purchasers neither demanded nor expected to find available evidence regarding the direct instructional effects of materials under consideration. A follow up to this study (EPIE, 1974) concluded that most local and state systems were still concentrating their efforts in judging the content of the materials and deciding whether they would hold up under year-long daily use by students (i.e., How good is its coverage? How good is its cover?).

Once those matters have been decided and a material has been chosen, then is simply assumed that any teacher should be able to see to it that students learn the content of the material in question. In other words, it appears as if nobody seems to feel it necessary to measure either the

effects or the acceptance resulting from the actual use of the materials.

There is a long tradition which assumes that teachers will be able to teach effectively, not matter what materials they may required to use . Until very recently no teacher has been ever been held accountable for achieving a specified level of learning for each and every student within a class. Similarly, no instructional materials publisher has ever been actually required to provide evidence that his materials have measurably helped teachers and learners achieve such a specified learning.

Today, however, the interrelated demands for equal educational opportunity, educational accountability and individualized instruction, among other educational movements, are in the process of making such demands and expectations a part of the shifting realities which are affecting all aspects of education and no least the instructional materials field.

What appears to constitute a major shift in the way materials are selected and produced for student use by many schools is that teachers have begun to maintain that they cannot be held accountable for specific levels of achievement with particular learners if they are required to teach with materials which are not appropriate for the learners and targeted to specified learning objectives. Implicit in such a stand is the fact that teachers are demanding that producers of the materials used in their classrooms be willing to bear a fair share of the total instructional responsability.

Activities for gathering "learner-verification" evidence that can be used by educators to judge the extent to which instructional materials contribute positively to student achievement are more and

more requested to the producers of instructional materials. This request asks for empirically obtained descriptions of the performance of specific materials with students, under specific teaching conditions, and within a variety of educational contexts. The producers also must make it clear that they use the empirical findings to improve or maintain at a high level the future performance of their materials.

To date, the verification activities such as field tests, tryouts, or other forms of formative evaluation to measure the results of product use have been used only during the product's development final stages. A more comprehensive evaluation, however, calls for a continued gathering of data since the material need is sensed.

Such a kind of verification and revision is suggested within the discussed instructional materials evaluation model. It is a practical framework for developing and refining materials within educational or training institutions.

Lacking a hard science of learning and motivation and throughly scientific technology for instructional material development, the practical empiricism implicit in the model seems a reasonable way to proceed. It is more than possible that, in time, increasingly scientific principles of product development will emerge from observable trends in the sorts of internal product changes and external product adaptations resulting from the continuous verification of products. If such general principles are forthcoming, the process of instructional materials evaluation, which is essentially an empirically-based corrective feedback device, will also become the basis of a much

needed scientifically-based collective feed-forward mechanism.

The aim of this thesis has been to move one step further in the direction of an adequate framework for the evaluation of instructional materials. It is clear that taking this step involves considerable complication of the activities and procedures of adequate instructional materials evaluation, by comparison with what has frequently appeared under this heading in the past.

References

Ackoff, R. L., Gupta, S. K., & Minas, J. S. Scientific method : optimizing applied research decisions. New York: Wiley, 1962.

Airasian, P. W. Designing summative evaluation studies at a local level. In W. J. Popham (Ed.), Evaluation in education. Berkeley, Calif.: McCutchan, 1974.

Alkin, M. C. The use of behavioral objectives in evaluation : relevant or irrelevant? Paper presented at the 18th Annual ETS Western Regional Conference on Testing Problems. San Francisco, 1969.

Alkin, M. C., & Fink, A. Evaluation within the context of curriculum development: a user orientation. In G. D. Borich(Ed.), Evaluating educational programs and products. Englewood Cliffs, N.J.: Educational Technology Publications, 1974.

Anderson, R. H. Selection of media: another perspective. Improving Human Performance, 1974, 3, 81-107.

Anderson, R. H. Selecting and developing media for instruction. New York: Van Nostrand, 1976.

Anderson, S. A., Ball, S., & Murphy, R. T. Encyclopedia of educational evaluation, San Francisco, Calif.: Jossey-Bass, 1975.

Antonoplos, D. A consumer's guide to training materials. San Francisco, Calif.: West Laboratory for Educational Research and Development, 1977.

Archibald, K. A. A classification and comparison of evaluative activities, Berkeley, Calif.: Institute of Urban and Regional Development, University of California, 1976.

Armstrong, J. R. A generalized model for the evaluation of instructional materials and media. In J. R. Armstrong (Ed.), A sourcebook for the evaluation of instructional materials and media, Washington, D.C.: Office of Education (DHEW), 1973. ED 107 050

Ashby, W. R. An introduction to cybernetics. London: Chapman and Hall. 1956.

Baker, E. L., & Alkin, M. C. Formative evaluation of instructional developments. AV Communication Review, 1973. 21, 389-418.

Berelson, B. Content analysis. In Handbook of social psychology, Cambridge, Mass.: Addison Wesley, 1954.

Braby, R. An evaluation of ten techniques for choosing instructional media. Orlando, Florida.: Naval Training Equipment Center, 1973. ED 088 425

Brien, R. L., & Towle, N. J. Instructional design and development: accelerating the process. Educational Technology, 1977, 27, 12-17

Briggs, L. J. Handbook of procedures for the design of instruction. Pittsburg, Pennsylvania: American Institute for Research, 1970.

Brofenbrenner, U. The experimental ecology of education. Educational Researcher, 1976, 5, 5-15.,

Bunda, M. A. Evaluation bibliography, Aid 1. Instructional Aid Series. The Evaluation Center, College of Education, Western Michigan University, 1976. ED 166 244

Butman, J. W., & Fletcher, J. L. The role of the evaluator and developer in educational research and development. In G. D. Borich (Ed.), Evaluating educational programs and products. Englewood Cliffs N. J.: Educational Technology Publications, 1974.

Caldwell, M. S. Input evaluation and educational planning. Columbus, Ohio: Ohio State University, College of Education, Evaluation Center, 1968. ED 025 043.

Campbell, D. T. Reforms as experiments. American Psychologist, 1969, 24, 409-429.

Campbell, D. T., & Boruch, R. F. Making the case for randomized assignment to treatment by considering the alternatives: six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Benett & A. A. Lumsdaine (Eds.), Evaluation and experiment. New York: Academic Press, 1975.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Carter, W. E. A taxonomy of evaluation models: use of evaluation models in program evaluation. Paper presented to the Annual Meeting of the American Educational Research Association, Washington, D. C., 1975. ED 109 244

Checkland, P. B. Towards a systems-based methodology for real-world problem solving. In J. Beishon & G. Peters (Eds.), Systems behavior. London: Harper and Row, 1976.

Clark, R. E., & Snow, R. E. Alternative designs for instructional technology research. AV Communication Review, 1975, 23, 373-394.

Cowles, H. M. Textual materials evaluation: a comprehensive checklist. Foreign Language Annuals, 1976, 9, 300-303.

Cunningham, D. J. Comments on the case studies of formative evaluation: the source of information. Viewpoints, 1972, 48, 111-117.

Cunningham, D. J. Evaluation of replicable forms of instruction. AV Communication Review, 1973, 21, 351-367. ED 051 263

Davies, I. Task analysis: some process and content concerns. AV Communication Review, 1973, 21, 73-85.

Dick, W. An overview of the systems approach. In R. N. Singer and W. Dick (Eds.), Teaching physical education: a systems approach. Boston: Houghton-Mifflin, 1974.

Dick, W. & Carey, L. The systematic design of instruction. Glenview, Illinois: Scott Foresman, 1978.

Dwyer, F. M. A guide for improving visualized instruction. Pennsylvania: State College Learning Services, 1972.

Educational Products Information Exchange Institute. EPIE Report No. 1. New York: 1969.

Educational Products Information Exchange Institute. EPIE Report No. 2 New York: 1974.

Felter, W. R. An evaluation instrument for instructional materials. Educational Technology, 1978, 8, 55-56.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart, and Winston, 1970.

Gagné, R. M., & Kneller, G. P. Behavioral objectives, yes or no? Educational Leadership, 1972, 19, 394-400.

Gardner, D. E. Five evaluation frameworks. Implications for decision making in higher education. Journal of Higher Education. 1977, 48, 571-593.

Gilbert, J., Light, R. J., & Mosteller, F. Assessing social innovations: an empirical base for policy. In C. A. Benett & A. A. Lumsdaine (Eds.), Evaluation and experiment. New York: Academic Press, 1975.

Glaser, R. Theory of evaluation of instruction. Paper presented at the Symposium on Problems in the Evaluation of Instruction. University of California, Los Angeles, 1967.

Grobman, H. Evaluation activities of curriculum projects: a starting point. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNally, 1968.

Guba, E. G. The failure of educational evaluation. Educational Technology, 1969a, 9, 29-38.

Guba, E. G. Significant differences. Educational Researcher, 1969b, 20, 4-5.

Guba, E. G., & Stuflebeam, D. L. Evaluation: the process of stimulating aiding, and abetting insightful action. Monograph Series in Reading Education No. 1. Bloomington: Indiana University, 1970.

Guttentag, M. Models and methods in evaluation research. Journal of the Theory of Social Behavior, 1971, 1, 75-95.

House, E. R. Assumptions underlying evaluation models. Journal of Educational Research, 1978, 71, 4-12.

Hug, W. E. Evaluation of educational media. Improving College and University Teaching, 1973, 21, 33-34.

Immegart, G. L. , & Pilecky, F. J. An introduction to systems for the educational administrator. Menlo Park, Calif.: Addison-Wesley, 1973.

Kaufman, R. A. Educational system planning. Englewood Cliffs, N. J.: Pretice Hall, 1972.

Kempt, J. Planning and producing audiovisual materials. New York: Thomas Y. Crowell, 1975.

Kopstein, F. General system theory as the basis for a theory of instruction, Paper presented at the Annual Meeting of the Society for General Systems Research, December 1956.

Lathan, G. Measuring teacher response to instructional materials. In J. Armstrong (Ed.) A sourcebook for the evaluation of instructional materials and media. Washington, D.C.: Office of Education (DHEW), 1973. ED 107 050

Lawson, T. E. Formative instructional product evaluation instruments. Urbana, Ill.: Center for Instructional Research and Curriculum Evaluation, 1972.

Lindvall, C. M., & Cox, R. C. The role of evaluation in programs for individualized instruction. In R. W. Tyler (Ed.) Educational evaluation: new roles, new means. The 68th Yearbook of the NSSE, part 2. Chicago: University of Chicago Press, 1969.

Lumsdaine, A. A. Assesing the effectiveness of instructional programs. In R. Glaser (Ed.), Teaching machines and programmed learning II. National Education Association of the U.S., 1965.

Mager, R. F. Preparing objectives for programmed instruction. Belmont, Calif.: Fearon, 1962.

Markle, S. M. Empirical testing of programs. In P. C. Lange (Ed.), Programmed instruction. The 66th Yearbook of the NSSE, part 2. Chicago: University of Chicago Press, 1967.

Martin, P. Guidelines for evaluating training tapes. Educational and Industrial Television, 1973, 5, 40-43.

McLaughlin, J. A., & Trlica, J. S. Teacher evaluation of instructional materials. Educational Technology, 1976, 16, 51-54.

Naegle, V. J. The evaluation of instructional materials: Bibliography. Washington, D. C.: Mid Atlantic Region, Special Education Instructional Materials Center, 1970. ED 044 447

Nystrom, L. J. Metro 1977 multimedia evaluation report. Windsor, Conn.: Metro Media Evaluation Center, 1977. ED 148 391.

Paulson, C. F. A strategy for evaluation design. Mammouth, Ore.: Teaching Research, 1970.

Penta, M. Q., & Cavero, C. The appraisal of mediated materials for use in modular curriculum. Journal of Biocommunication, 1976, 3, 8-14.

Popham, W. J. Establishing instructional goals. Englewood Cliffs, N. J.: Prentice Hall, 1970.

Popham, W. J. Educational needs assessment. In J. Weiss (Ed.), Curriculum evaluation. Curriculum Theory Network, monograph supplement. Toronto: The Ontario Institute for Studies in Education, 1972.

Popham, W. J. Educational evaluation. Englewood Cliffs, N.J.: Pretice Hall, 1975.

Provus, M. M. Discrepancy evaluation for educational programs improvement and assessment. Berkeley, Calif.: McCutchan, 1971.

Rapp, M. L., Root, J. G., & Summer, G. Some considerations in the experimental design and evaluation of educational innovations. Santa Monica, Calif.: The RAND Corporation, 1970. ED 052 530

Riecken, H. W., & Boruch, R. F. Social experimentation: a method for planning and evaluating social intervention. New York: Academic Press, 1974.

Robertshaw, J. E., Mecca, S. J., & Rerick, M. N. Problem solving: a systems approach. New York: Petrocelli, 1978.

Salomon, G., & Clark, R. E. Reexamining the methodology of research on media and technology in education. Review of Educational Research, 1977, 47, 99-120

Sanders, J. R., & Cunningham, D. J. Formative evaluation: selecting techniques and procedures. In G. D. Borich (Ed.), Evaluating educational programs and products. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Saylor, J., & Alexander, W. Planning curriculum for schools, New York: Holt, Rinehart, and Winston, 1974.

Scott, R. O., & Yelon, S. L. The student as co-author. The first step in formative evaluation. Educational Technology, 1969, 9, 76-78.

Scriven, M. The methodology of evaluation. In R. W. Tyler, R. Gagné, and M. Scriven (Eds.), AERA Monograph Series on Curriculum Evaluation No. 1, Chicago: Rand McNally, 1967.

Scriven, M. Goal-free evaluation. In E. House (Ed.), School Evaluation. Berkeley, Calif.: McCutchan, 1972.

Severy, L. J. Application of the experimental method to program evaluation; problems and prospects. Princeton, N.J.: ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1975. ED 117 190

Stake, R. E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stake, R. E., & Denny, T. Needs, concepts and techniques for utilizing more fully the potential of evaluation. In R. W. Tyler (Ed.), Educational evaluation: new roles, new means. The 68th Yearbook of the NSSE, part 2. Chicago: university of Chicago Press, 1969.

Stake, R. E. Responsive evaluation. ERIC ED 075 487, 1972.

Stanley, J. C. Controlled field experiments as a model for evaluation. In P. H. Rossi & W. Williams (Eds.), Evaluating social programs. New York: Seminar Press, 1972.

Steele, S. R. Contemporary apprpaches to program evaluation and their implications for evaluating programs for disadvantaged adults. Syracuse, N. Y.: ERIC Clearinghouse on Adult Education, 1973. ED 075 715.

Stufflebeam, D. L. Toward a science of educational evaluation. Educational Technology, 1968, 8, 5-12.

Stufflebeam, D. L. Evaluation as enlighment for decision making. Improving educational assessment and an inventory of measures of effective behavior. Washington, D. C.: Association for Supervision and Curriculum Development, National Education Association, 1969.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. Educational evaluation and decision making in education. Itasca, Ill.: Peacock, 1971.

Suchman, E. Evaluative research, New York: Rusell Sage Foundation, 1967.

Tallmadge, G. K., & Horst, D. P. A procedural guide for validating achievement gains in educational projects. Los Altos, Calif.: RMC Corporation, 1974. ED 096 344.

Teague, F. A. A form for all reasons. Audiovisual Instruction. 1978, 23, 40-41.

Thompson, J. E. Media evaluations 1974. Indiana Association for Educational Communications and Technology, 1974.

Tyler, L. L., Klein, M. F., & Michael, W. B. Recommendations for curriculum and instructional materials. Los angeles: Tyl Press, 1971.

Tyler, L. L., & Klein, M. F. Caveat emtor: let the buyer beware Educational Technology, 1973, 13, 52-54.

Webster, W. J. The evaluation of instructional materials. ERIC Clearinghouse on Informational Resources. Stanford Center for Research and Development in Teaching, 1976. ED 131 861.

Weiss, C. H. Evaluating action programs: readings in social action and organization. Boston: Allyn and Bacon, 1972.

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

Wrighstone, J. W. Evaluating educational programs. The Urban Review. 1969, 3, 19-26.

Wrigth, W. J., & Hess, R. J. A criteria acquisition model for educational product evaluation. In G. D. Borich (Ed.), Evaluating educational programs and products. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

APPENDIX

Instructional Materials Selection Guide*

Determine if you need to provide instruction or information.

1. Is your purpose to produce a measurable behavior change?

   Yes. Go to item 2. Instruction.

   No. Go to item 41. Information.

Determine transmission method.

2. Does the lesson need to be reproducible? (Is it intended for large target population or distributed over a large area?)

   Yes. Use an instructional medium. Go to item 22.

   No. Go to item 3.

3. Is lesson content intended to be self-paced?

   Yes. Use an instructional medium. Go to item 22.

   No. Go to item 4.

4. Is it critical that the content be unchangeable?

   Yes. Use an instructional medium. Go to item 22.

   No. Go to item 5.

5. Is face to face communication impractical? (Not enough qualified instructors, insufficient number of students per class, etc.)

   Yes. Use an instructional medium, go to item 22.

   No. Use an instructional aid to support instructor-led presentation. Go to item 6.

Selecting instructional aids.

6. Does the lesson involve objects or things unfamiliar to the student?

   Yes. Go to item 13.

   No. Go to item 7

*Adapted from Anderson, 1976.

7. Does the lesson involve interpersonal skills?

   Yes. Go to item 8.
   No. Go to item 10.

8. Is display of motion necessary to model performance or to provide feedback?

   Yes. Consider: Class VII
   Class IX

   No. Go to item 9.

9. Is display of sound necessary?

   Yes. Consider: Class I
   Class III
   Class IX

   No. Consider: Class II
   Class IV
   Class VIII

10. Is display of motion necessary?

    Yes. Go to item 11.
    No. Go to item 12.

11. Is it practical to demonstrate the real thing in class?

    Yes. Consider: Class VIII support with Class II

    No. Consider: Class VII
    Class VIII
    Class IX

12. Is display of sound necessary?

    Yes. Consider: Class I
    Class II
    Class V
    Class VIII

    No. Consider: Class II
    Class IV
    Class VIII

13. Is it practical to demonstrate the real thing in class?

    Yes. Go to item 17.
    No. Go to item 14.

14. Is display of motion necessary?

    Yes. Go to item 18.
    No. Go to item 15.

15. Is color necessary?

    Yes. Go to item 21.
    No. Go to item 16.

16. Is sound necessary?

    Yes. Consider: Class III (Black and white)
                   Class V
                   Class VIII
    No. Consider: Class II (Black and white)
                  Class IV

17. Is it desirable to display exaggerated views of objects?

    Yes. Consider: Class II
                   Class IV
                   Class VIII
    No. Consider: Class II

18. Is color necessary?

    Yes. Go to item 20.
    No. Go to item 19.

19. Is sound necessary?

    Yes. Consider: Class VII
                   Class VIII
    No. Consider: Class VI (Black and white)
                  Class VIII

20. Is sound necessary?

Yes. Consider: Class VII
No. Consider: Class VI
Class VIII

21. Is sound necessary?

Yes. Consider: Class III
Class V
No. Consider: Class IV
Class VIII

Selecting instructional media.

22. Does the lesson involve objects or things unfamiliar to students?

Yes. Go to item 30.
No. Go to item 23.

23. Does the lesson involve interpersonal skills?

Yes. Go to item 26.
No. Go to item 24.

24. Is display of motion necessary?

Yes. Go to item 28.
No. Go to item 25.

25. Is display of sound necessary?

Yes. Consider: Class I
Class III
No. Consider: Class II
Class III

26. Is display of motion necessary?

Yes. Go to item 29.
No. Go to item 27.

27. Is display of sound necessary?

Yes. Consider: Class I
Class III
No. Consider: ClassII

28. Is display of sound necessary?

Yes. Consider: Class II and Class VII
No. Consider: Class VI with Class II
Class VII with Class II

29. Is display of sound necessary?

Yes. Consider: Class VII and Class II
No. Consider: Class VI with Class II
Class VII and Class II

30. Is display of motion (either real or exaggerated) necessary?

Yes. Go to item 38.
No. Go to item 31.

31. Is color necessary?

Yes. Go to item 35.
No. Go to item 32.

32. Is sound necessary?

Yes. Go to item 34.
No. Go to item 33.

33. Is it practical to have student interact with the real thing?

Yes. Consider: Class VIII with Class II (Black and white)
ClassVIII and Class I
Class III (Black and white)
No. Consider: Class II (Black and white)
Class III (Black and white)

34. Is it practical to have student interact with the real thing?

Yes. Consider: Class VIII with Class I
Class III (Black and white)
Class V with Class II (Black and white)
No. Consider: Class III (Black and white)
Class V and Class II

35. Is sound necessary?

Yes. Go to item 37.
No. Go to item 36.

36. Is it practical to have student interact with the real thing?

Yes. Consider: Class VIII with Class II (or I)
Class V and Class II
No. Consider: Class V with Class II (Black and white)
Class II (Color)

37. Is it practical to have student interact with the real thing?

Yes. Consider: Class VIII with Class II (or I)
Class V and Class II
No. Consider: Class V and Class II
Class I and Class II

38. Is color necessary?

Yes. Go to item 40.
No. Go to item 39.

39. Is it practical to have student interact with the real thing?

Yes. Consider: Class VIII and Class II
Class VII and Class II
No. Consider: Class VII and Class II

40. Is it practical to have student interact with the real thing?

Yes. Consider: Class VIII and Class II
Class VII and Class II
No. Consider: Class VII and Class II (Color)

Selecting information aids and media.

41. Is message intended for large target population? (Difficult to reach everyone at one time)

Yes. Go to item 49.
No. Go to item 42.

42. Is exact message content critical?

Yes. Go to item 46.
No. Go to item 43.

43. Is face to face communication feasible?

Yes. Go to item 45.
No. Go to item 44.

44. Does the message deal with objects not familiar to target population?

Yes. Consider distribution of:
Printed materials support with simple drawings.
Audio tapes support with drawings or photos.
No. Consider: Conference or individual calls.
Distribution of memos letters, etc.

45. Does the message deal with objects not familiar to target population?

Yes. Consider:
Small group meetings support with real things.
Printed materials support with drawings or photos.
No. Consider:
Small group meetings.
Individual meetings.
Conference calls.
Distribution of printed materials.

46. Is face to face communication feasible?

Yes. Go to item 48.

No. Go to item 47.

47. Does the message deal with objects not familiar to target population?

Yes. Consider distribution of:

Printed materials support with drawings, photos.

Audio tapes support with drawings, photos.

No. Consider:

Printed materials.

Conference calls.

Audio tapes.

48. Does the message deal with objects not familiar to target population?

Yes. Consider:

Group meeting (entire group at one time) support with real-things, handouts, photos, slides.

Printed materials support with photos, drawings.

No. Consider:

Group meeting (entire group at one time).

Distribution of printed material or audio tapes.

49. Is exact message content critical?

Yes. Go to item 53.

No. Go to item 50.

50. Does the message deal with objects not familiar to target population?

Yes. Go to item 52.

No. Go to item 51.

51. Is attitude change critical?

Yes. Consider distribution of:

Audio tapes.

Slide tapes.

Video tapes.

Printed materials.

No. Consider distribution of:

Printed material.

Audio tapes.

52. Is attitude change critical?

Yes. Consider:

Video tape.

Slide tape.

Audio tape with print.

Printed material.

Film.

No. Consider distribution of:

Printed material support with photos or drawings.

53. Does the message deal with objects not familiar to target population?

Yes. Go to item 55.

No. Go to item 54.

54. Is attitude change critical?

Yes. Consider:

Video.

Slide tapes.

Audio tapes.

Film.

No. Consider:

Printed material.

Audio tapes with print.

55. Is attitude change critical?

Yes. Consider:

Video.

Slide tape.

Filmstrip tape.

Audio tape support with drawings or photos.

No. Consider:

Printed material support with drawings or photos.

Audio tape support with drawings or photos.