Basic Concepts of Evaluation Research

Dessalegn Chalchisa*

Introduction

Evaluation research is important to policy makers, program managers, and curriculum developers. This topic discusses:

- The steps in conducting evaluation research,
- Criteria for judging the quality of an evaluation study,
- Major quantitative and qualitative approaches to evaluation research, and
- The role of an evaluation research in educational research and development (R & D).

Definitions of Educational Evaluation and Evaluation Research

There are several definitions of the term *evaluation.* They differ in the level of abstraction and the specific concerns of the person who formulates them.

A comprehensive definition of evaluation was provided by Beeby (cited in Wolf, 1984) as *the systematic collection and interpretation of evidence, leading, as part of the process, to a judgment of value with a view to action.* There are four key elements in this definition. First, the use of the term *systematic* implies that what information is needed will be defined with some degree of precision and that efforts to secure such information will be based on plans. This does not mean that only information which can be gathered through the use of standard tests and other related measures will be useful. Information gathered by means of observational procedures, questionnaires, and

* Assistant Professor, Institute of Educational Research, Addis Ababa University.

Basic Concepts of Evaluation Research

Dessalegn Chalchisa*

Introduction

Evaluation research is important to policy makers, program managers, and curriculum developers. This topic discusses:

- The steps in conducting evaluation research,
- Criteria for judging the quality of an evaluation study,
- Major quantitative and qualitative approaches to evaluation research, and
- The role of an evaluation research in educational research and development (R & D).

Definitions of Educational Evaluation and Evaluation Research

There are several definitions of the term *evaluation*. They differ in the level of abstraction and the specific concerns of the person who formulates them.

A comprehensive definition of evaluation was provided by Beeby (cited in Wolf, 1984) as *the systematic collection and interpretation of evidence, leading, as part of the process, to a judgment of value with a view to action*. There are four key elements in this definition. First, the use of the term *systematic* implies that what information is needed will be defined with some degree of precision and that efforts to secure such information will be based on plans. This does not mean that only information which can be gathered through the use of standard tests and other related measures will be useful. Information gathered by means of observational procedures, questionnaires, and

* Assistant Professor, Institute of Educational Research, Addis Ababa University.

interviews can also contribute to an evaluation enterprise. The important point is that whatever kind of information is gathered should be acquired in a systematic way.

The second element in Beeby's definition, *interpretation of evidence*, introduces a critical consideration sometimes overlooked in evaluation. The mere collection of evidence does not, by itself, contribute to evaluation work. Yet uninterpreted evidence is often presented to indicate the presence (or absence) of quality in an educational venture. High dropout rates, for example, are frequently mentioned as indications of the failure of educational programs. Doubtless, high dropout rates are indicators of failure in some cases, but not always. There may be very good reasons why students dropout of educational programs. Personal problems, acceptance into higher level educational programs, and landing on a good job are some reasons for dropping out and these may not reflect the weakness of the educational program which may in no way reflect on the program.

Clearly, information gathered in connection with the evaluation of an educational program must be interpreted with great care. In the above example, the problem of interpretation was rather simple. Dropout statistics are easily gathered and one can usually have confidence in the numbers. More complex situations arise when one uses various tests, scales or observational and self–report devices such as questionnaires and opinionnaries. In this situation the interpretation of evaluation information can be extremely difficult.

The third element of the Beeby's definition –*judgment of value* –takes evaluation far beyond the level of mere description of what is happening in an educational enterprise. It requires the evaluation worker, or the group of persons responsible for conducting the evaluation to make judgments about the worth of an educational endeavor.

The last element of Beeby's definition–*with a view to action* –emphasizes the distinction between an undertaking that results in a judgment of value with no specific reference to action and that which is deliberately undertaken for the sake of future action. Educational evaluation is clearly decision oriented. It is intended to lead to better policies and practices in education.

Why has evaluation research attracted so much interest? The main reason is that administrators have come to view evaluation as an important tool in policy analysis, in political decision making–process, and in the program management.

With respect to policy analysis, evaluation research yields important data about the costs, benefits and problems of various program alternatives. Policy analysts can use this data to prepare position papers, which are then reviewed by persons with decision–making authority. With respect to political process, evaluation findings are used to create advocacy for particular legislation and budget appropriations. Opponents of such legislation may sponsor their own evaluations to generate evidence favoring their causes.

Finally, evaluation research is becoming an increasingly important component of program management. For example, the *cost-benefit* evaluations (also called efficiency evaluations) are done to determine whether programs are producing benefits that justify their costs. Another use of evaluations is to hold managers *accountable* for producing results. Evaluation also are done to help managers make *sound decisions* related to program design, personnel, and budget.

Educational Evaluation is the process of making judgments about the merit, value, or worth of educational programs (Gall, Borg, & Gall, 1996). The term program is used as a generic label for the various phenomena (e.g., methods, materials, organizations, individuals) that are the focus of educational evaluation.

The Relationship between Evaluation and Research

Are evaluation studies the same as research? Is an educational researcher qualified to fill a position involving program evaluation? Our answer to the first question would be: they are both similar and different. To the second question: yes, but an educational researcher probably would need to acquire additional training.

The greatly accepted view is that educational research and educational evaluation overlap to a great extent. In practice, evaluators make substantial use of the research designs, measurement tools, and data analysis techniques that constitute the methodology of educational research. For this reason, evaluation studies are referred as *evaluation research*. Yet there are important differences between evaluation and other types of research.

There is usually a marked difference in content, presentation, and often method between research inspired by scholarly interest or an academic requirement, and an evaluation undertaken with a definite practical problem in mind. Research typically aims at producing new knowledge which may have no specific reference to any practical decision, while evaluation is deliberately undertaken as a guide to action (Gall, Borg, & Gall, 1996).

An evaluation study is usually initiated for policy, management, or political strategy decisions. The purpose of the evaluation study is to collect data that will facilitate decision making. In contrast, the purpose of research study, broadly stated, is to develop an understanding of a particular phenomenon. Of course, the findings of a research study also can be used to guide decision making; and evaluation data can be relevant to developing an understanding of a particular phenomenon (Gall, Borg, & Gall, 1996).

A more basic distinction between evaluation and research lies in the generalizability of results that each type of activity produces. Research is concerned with the production of knowledge that is *generalizable* as possible. For example, a research worker may undertake an investigation to determine the relationship between student aspiration and achievement. The study will be designed and carried out in such way as to ensure results that are generalizable possible. They will be obtained over a wide geographical area, apply to a broad range of ages, and be as true in several years as now. Generalizabilty of results is critical in research (Gall, Borg & Gall, 1996).

Evaluation, in contrast, seeks to produce knowledge *specific to a particular setting.* In contrast, evaluation is done for a very specific purpose. Decision makers might be interested in how well their particular program works, and thus they commission a site-specific evaluation study to collect data relevant to their special concern. For instance, evaluators who are concerned with the evaluation of a reading improvement program for third graders in a school or region, will direct their efforts toward ascertaining the effectiveness of the program in that locality. The resulting evaluative information should have high local relevance for teachers and administrators in that region. The results may have no scientific relevance for any other school in any other region.

Another important distinction between evaluation and research involves judgments of value. Evaluation studies are designed to yield data concerning the worth, merit, or value of educational phenomena. Their findings tend to be stated in such phrases as "this reading program is superior to the other program with respect to ...." or "the teachers in this woreda thought that this new approach to inservice training is superior to the existing approach because ...." Researchers, however, design their studies to discover the essential characteristics of educational phenomena. Their findings tend to be

couched in such phrases as "It appears that variable X has an influence on variable Y," or "Using a grounded theory approach, we discovered the counselors attribute at-risk students' behavior to two types of motivation." Educators may make value judgments and decisions based on such research findings , but this is a secondary use of the findings.

Steps in Conducting Program Evaluation

An evaluation study follows essentially the same steps as those involved in doing a research study. A few additional factors must be considered, however, depending upon the evaluation model that is used. Gall, Borg, & Gall (1996) identified seven steps in conducting evaluation research.

Clarifying Reasons for doing an evaluation

An evaluation study can be initiated because of the evaluator's personal interest in doing it or because some person or agency required it. Both reasons can be involved in initiating the study, as when the evaluator's personal interest and an agency's need for evaluation happen to coincide.

If the evaluation study is done to answer questions primarily of interest to you, you will need only to clarify for yourself why the study is being done. Such study is different from the one which may be initiated by, for example, a school as in the following case: *The program evaluation was undertaken by the evaluator at the request of the director and the professional staff of the school. The decision makers need to know how effective their school was in terms of individual students and in terms of educational programs.*
When an evaluation is requested in this way, the evaluator should consider to determine all the reasons for the evaluation request. Evaluations can be requested because they are required by a funding

agency. Such evaluations are usually legitimate. Evaluations can be also requested for more skeptical reasons. If someone wants to use evaluation to shape the behavior of program staff, the evaluation serves a watchdog function. If someone wants the evaluator to gather evidence that can be used to justify an already made decision to terminate the program or reduce its funding, the evaluator becomes a "hired-gun."

An evaluation request can be made for various reasons, some of them covert and some overt. Therefore, the evaluator needs to spend time interviewing key individuals to determine whether the request is reasonable and ethical. Evaluation experts should refuse to conduct an evaluation if it leads to any breach of ethics.

Clarifying reasons for an evaluation request is also useful in selecting an appropriate evaluation model, which will be discussed later.

Identifying Stakeholders

A *stakeholder* is anyone who is involved in the program being evaluated or who might be affected by or interested in the findings of the evaluation. It is important to identify the stakeholders at the outset of an evaluation study. They can help you clarify the reasons why the study was requested, the question that could guide the evaluation, the choice of research design, the interpretation of results, and how the findings should be reported and to whom.

Ignoring the stakeholders can have a serious political consequences. Stakeholders can sabotage the evaluation process or discredit the results if they think that the evaluator has not responded appropriately to their need for involvement. The threat of sabotage does not mean, however, that you must involve all stakeholders on the same level. Some stakeholders may wish simply to be kept

informed, whereas others may want to influence the questions that guide the study and the evaluation design.

Deciding what is to be Evaluated

One of the first tasks that confront the evaluator is *program delineation*, which is the process of identifying the most important characteristics of the program to be evaluated. Careful delineation is important even in local evaluation research. It is common for persons working in a program to know only those aspects that affect them directly. Unless all program components are delineated, an important component might be overlooked in the evaluation process.

Following program delineation, the program should be analyzed to determine which of its components are to be included in the evaluation study. Program components can be grouped into the following categories: goals, resources, procedures, and management.

Goals: Judgments about the merit of program goals are central to most evaluation studies. A *goal* is the purpose, effect, or end point that the program developer is attempting to achieve. If the program does not have goals, or if the goals are not perceived as worthwhile, it is difficult to imagine how the program itself can have merit.

Some programs have carefully specified goals. In the other programs the evaluator must infer the goals that the developer has in mind. Once the program goals have been identified, the evaluator may be asked by the stakeholders to determine the extent to which the program actually achieves its intended goals.

Resources: Resources are the personnel, equipment, space, and other cost items needed to implement program procedures. Stakeholders might want to know the answers to such questions as: are our present resources sufficient to execute the program as

intended by its developers? Is the program very expensive? Are there hidden costs in the programs? Will the program take away resources needed by other programs? Each of these questions requires the evaluator to focus on the program resources.

Procedures: Procedures are the techniques, strategies, and other processes used in conjunction with resources to achieve program goals. Examples of evaluation questions that concern program procedures are "How long did teachers need to use the materials before students mastered the content?" "Did teachers have difficulty in using the inquiry approach to science teaching?" "To what extent did teachers actually use the inquiry approach?" Answers to these questions usually require close and repeated observation of the program in operation.

Evaluation of program resources and procedures is especially helpful for understanding the observed effects of a program. Suppose a new instructional program is observed to have negligible effects on student achievement. Decision makers might choose to discontinue the program because the evaluation results were negative. Yet the program may have been ineffective because needed materials did not arrive on time, or because teachers experienced many interruptions that reduced the total time allotted to program implementation. If the evaluator had collected data on these resources and procedural problems, the decision makers might have chosen another alternative course of action, for example, to remove the "defects" from the program and try it again. In fact, collection of data on all three aspects of the program – resources, procedures and goal attainment – is important in any type of formative evaluation. Decisions about program revision can be made more effectively if developers know how well the current version of the program is working and why.

Program Management: Most programs have program management systems to monitor resources and procedures so that they are used effectively to achieve program goals. So, evaluation studies could focus on management systems in response to such questions as "Is the management system ensuring the effective use of program resources?" "Is the management system as efficient as it can be?" "Are the management procedures being used as intended by the program developers?" Each of these questions requires the evaluator to design research that delineates the management systems and to examine its operation in practice.

Identifying Evaluation Questions

Cronbach (cited in Gall, Borg & Gall, 1996) distinguished two phases in selecting questions for an evaluation study. The *divergent phase* involves generating a comprehensive list of questions, issues, concerns and information that might be addressed in the evaluation study. As the evaluator, you should invite all stakeholders to contribute to this list. The second phase is the *convergent phase*. It involves reducing the initial list of evaluation questions to manageable numbers. This phase is necessary because of the expense involved in answering each evaluation question. The evaluator, in collaboration with the significant stakeholders, must screen the list to the most important questions that can be answered with available resources.

Developing an Evaluation Design and Timelines

An *evaluation design* is a complete description of the evaluation study in such a way that the reader has a clear, complete picture of what will be done in the process of conducting the evaluation. Many evaluation studies are similar to research studies in design, execution and reporting. Thus, any research design can be incorporated into the design of an evaluation study.

One issue in an evaluation design is whether the evaluation is to be done by an internal or external evaluator. Most types of evaluation can be done by an internal evaluator, especially when the evaluation findings are to be used to guide the program management and decision making. Summative evaluation is best done by an external evaluator. The purpose of summative evaluation is to determine the merits of a fully operational program and possibly to compare it with a competing program. The evaluator is obliged to represent the interests of the consumer to whom the evaluation study will be reported, or of the external agency that is sponsoring the evaluation. An external evaluator is in a much better position than an internal evaluator to satisfy the interests of the stakeholders. Even so, both an internal evaluator and an external evaluator could be pressurized to bias an evaluation design to produce particular results. If such pressure became too intense, the evaluator's only recourse would be to terminate the evaluation on ethical grounds.

In designing an evaluation study you should be aware that evaluation activities can both beneficial and harmful. On the one hand, persons involved in a program appear to do best when they feel they are valued unconditionally and do not have an evaluator watching over their shoulder. On the other hand, it appears that people move truer and more certainly toward excellence to the extent that they clarify their purposes, measure the impact of their action, judge it, and move on-in short, evaluate their progress. The perils of evaluation involve not only people, but the program itself. A program might be good, but a poor evaluation can cause others to misjudge it and contribute to its downfall. A program might have the potential to be good, but a negative evaluation while it is under development can lead administrators to withdraw funding. Furthermore, evaluation activities use up resources that could be allocated to support further program development.

The benefits and harms of evaluation are difficult to reconcile. Some evaluators recommend that you weigh all possible consequences of a planned evaluation activity. You should determine that the potential benefit outweigh the potential harm before you make a decision to proceed with evaluation. Also you should design the study to minimize potentially harmful effects. One way to accomplish this goal is to involve significant stakeholders in the design of the study. For example, they can assist in selecting or developing measures that reflect the outcomes most likely to be achieved by the program. Including the measures suggested by the stakeholders will make the evaluation less threatening to them. Also, the measures may reveal effects that cast light on the program.

In evaluation studies stakeholders usually want the final report by a certain time. In this situation, the evaluator will need to create a *timeline* as part of the evaluation design to ensure that the study is completed by the requested date. One advantage of creating a detailed timeline is that it can be used to identify and document the resources needed to complete a study by the requested date.

Collecting and Analyzing Data

Data collection and analysis in both evaluation studies and research studies are similar. This involves the collection of evaluation information using various methods and the summarization and interpretation of this evaluation information.

Reporting Evaluation Results

A research study will yield a single report, for example a technical report. The reporting of an evaluation study involves the dissemination of evaluation results to various stakeholders. Since various audiences are involved, the reporting of an evaluation study sometimes is more complicated.

Criteria for Good Evaluation Study

Good evaluation study satisfies four important criteria: utility, feasibility, propriety, and accuracy (Gall, Borg, & Gall, 1996).

Utility: An evaluation has utility if it is informative, timely and useful to the affected persons. Utility is operationalized in terms of:

- *Stakeholder identification*: All the groups affected by the evaluation should be identified.
- *Evaluator credibility*: the evaluator should be competent and trustworthy.
- *Information scope and selection*: The information to be collected should pertain directly to the evaluation questions and stakeholders concerns.
- *Value identification*: the evaluator's basis for making value judgments from the obtained results should be made clear.
- *Report clarity*: the evaluators' report should be comprehensive and easily understood.
- *Report timeliness and distribution*: evaluation reports, including interim reports, should be distributed to users in timely manner.
- *Evaluation impact*: the evaluation should be concluded so as to encourage appropriate action by the stakeholders.

Feasibility: Feasibility means, first, that the evaluation design is appropriate to the setting in which the study is to be conducted, and

second, that design is cost effective. It can be operationalized in terms of:

- *Practical procedures*: the evaluation procedures should be practical and minimize disruption to participants.

- *Political viability*: The evaluator should obtain the cooperation of affected interest groups and keep any of them from subverting the evaluation process.

- *Cost effectiveness*: The benefits produced from the evaluation should justify the resources expended on it.

Propriety: An evaluation has propriety if it is conducted legally and ethically. It can be operationalized in terms of:

- *Service orientation: The* evaluation should help stakeholders meet the needs of all their clients and the larger society as well.

- *Formal agreements*: The formal parties to evaluation should state their obligations and agreements in a written contract.

- *Rights of human subjects*: The rights and welfare of persons involved in the evaluation should be protected.

- *Human interactions*: Evaluators should show respect in their interactions with persons involved in the study.

- *Complete and fair assessment*: The strengths and the weakness of the entity being evaluated should be explored completely and fairly.

- *Disclosure of findings*: Individuals with a legal right to know and those affected by the results should be informed about the evaluation results.

- *Conflict of Interest*: If a conflict of interest should arise, it should be treated openly and honestly.

- *Fiscal responsibility*: expenditures of resources for the evaluation should be prudent and ethically responsible.

Accuracy: Accuracy refers to the extent to which an evaluation study has produced valid, reliable, and comprehensive information for making judgments of a program's worth. It can be operationalized in terms of:

- *Program documentation*: All pertinent aspects of the program being evaluated should be described in detail.

- *Context analysis:* Aspects of the program's context that affect the evaluation should be described in detail.

- *Described purposes and procedures*: The evaluation's purposes and procedures should be described in detail.

- *Defensible information resources*: Sources of data should be described in sufficient detail that their adequacy can be judged.

- *Valid information*: The data collection procedures should yield valid interpretations.

- *Reliable information*: The data collection procedures should yield reliable findings.

- *Systematic information*: The evaluation data should be reviewed and corrected, if necessary.

- *Analysis of quantitative information*: Analysis of quantitative data in an evaluation study should be thorough and should yield clear interpretations.

- *Analysis of qualitative information*: Analysis of qualitative data in an evaluation study should be thorough and should yield clear interpretations.

- *Justified conclusions*: Evaluators should provide an explicit justification for their conclusions.

- *Impartial reporting*: Evaluation reports should be free from bias and the personal feelings of any of those connected to the evaluation.

Quantitative Models to Evaluation

Like educational research, educational evaluation takes diverse forms. This is because evaluators over time have developed different purposes for doing evaluation, different philosophies, and different methodologies. These differences gradually led to the development of various approaches to evaluation research. An evaluator should review the various approaches to evaluation to determine which best suits his or her philosophical orientation to research. The following are the major approaches.

Evaluation of the Individual

Evaluation of the individual can be traced back at least to the early 1900s, when the testing movement began. Binet's intelligence test was published in 1904, and group ability testing began during World War I. Evaluation primarily involved the assessment of individual differences in intelligence and academic achievement. Test results were used for assigning grades and for selecting students into different ability tracks and special services.

This model of evaluation still is widely followed in education. In fact, it has been extended to the evaluation of teachers, administrators, and other school personnel. . Like assessment of students, personnel evaluation focuses on measurement of individual differences, and judgments are made by comparing the individual with a criterion or a set of norms.

Objectives-Based Evaluation

Ralph Tyler's work on curriculum evaluation in the 1940s brought about a major change in educational evaluation. Tyler's view was that the curriculum should be organized around explicit objectives and that its success should be judged on the basis of how well students achieve these objectives. This model marked a shift from a concern with evaluating individual students to a concern with evaluating the curriculum. In doing so, the model implied that students might perform poorly not because of lack of innate ability, but because of weaknesses in the curriculum.

Tyler's model has had an important influence on subsequent developments in educational evaluation. The practice of collecting data on academic achievement of students, and competence testing of students are developed based on the Tyler model.

Educational evaluators have developed other evaluation models that support Tyler's emphasis on the measurement of explicit objectives as the basis for determining an educational program's merit. For example, Provus (cited in Stufflebeam & Webster, 1983) developed the *discrepancy evaluation*, which focuses on the search for discrepancies between the objectives of a program and students' actual achievements of the objectives. The resulting information about the discrepancies can be used to guide program management decisions.

Another objectives based approach is cost analysis. Evaluators use *cost analysis* to determine either:

- the relationship between the costs of a program and the benefits of a program when both costs and benefits are calculated in monetary terms (called the *cost benefit ratio*); or

- the relationship between the costs of various interventions relative to their measured effectiveness in achieving a desired outcome (called *cost effectiveness*).

Different programs can be compared to determine which is most cost effective, that is, which promotes the greatest benefits for each unit of resource expenditure.

If you are planning a study of students' achievement of instructional objectives, one of your major concerns will be the measurement of these objectives. To facilitate measurement, it is helpful to state objectives in behavioral terms, meaning that the program outcomes are stated as behaviors that anyone, including the evaluators can observe in a program participant. This type of objectives commonly called a behavioral objective, usually has three components:

- statement of the program objective as an observable, behavioral outcome;
- criteria for successful performance of behavior; and
- the situational context in which the behavior is to be performed.

Here is an example of a behavioral objective: Given a set of 20 single digit multiplication problems, the learner will be able to solve them by writing the correct answer beneath each problem in less than five minutes with no more than two errors.

Behavioral objectives have been criticized on the ground that they reduce education to a matter of teaching only that can be stated and measured in the language of behavioral objectives. Of course, behavioral objectives, like any other techniques, can be misused. Used appropriately, however, they simplify the task of developing suitable instruments especially domain-referenced instruments to measure learner's achievement of objectives.

Another issue in evaluating program objective involves which objectives to measure. Evaluators often rely on the program developers or experts to make this decision. Scriven (1983), however, argued that evaluators should not know the program goals in advance because they might become co-opted by them and thus overlook the effects of the program, especially adverse side effects. Scriven suggested that to avoid this problem , evaluators should conduct research to discover the actual effects of the program in operation, which may differ markedly from the program developer's stated goals. This strategy for evaluation has come to be known as *goal free evaluation.*

Although goal free evaluation has a merit, there are many situations in which an evaluator is expected to collect evaluative data about specific program goals. Even in this situation, however, the evaluator can attend to the stated goals but also remain alert to the possibility that the program may have actual effects (both beneficial and adverse) quite different from those intended by the program developers.

Needs Assessment

A *need* can be defined as a discrepancy between an existing set of conditions and a desired set of conditions. For example, suppose an educator makes the assertion, "We need to place more emphasis on science education in our elementary school curriculum." The educator is saying in effect that there is a discrepancy between the existing curriculum and the desired curriculum. This statement of need reflects a judgment about the present merit of the curriculum. Also, note that the assessment of needs provides a basis for setting objectives for curriculum or program development. Because needs assessment is closely related to objective based models of evaluation, it is treated as a quantitative approach. Quantitative research methods enable researchers to measure the precise extent of discrepancy between an existing state and a desired state. Nevertheless, qualitative needs assessments are also conducted in education.

Several problems can arise in doing needs assessment. One of them is the definition of *needs*. Exactly what is a *desired* set of conditions? Roth (cited in Gall, Borg and Gall, 1996) identified five types of desired states: ideals, norms, minimums, desires (wants), and expectations. A need can be a discrepancy between an actual state and any of these five desired states. The goal of college education for all citizens who desire one (an ideal desired state) is certainly a

different kind of desired state from the goal of a basic skill in reading for all children (a minimum desired state)

Another problem with needs assessment is that the values underlying needs often are not clearly articulated. It is helpful to determine quantitatively the extent to which certain groups view particular elements of education (e.g., large class size, self contained classroom, computer assisted instruction) as needs. Personal values and standards are important determinants of needs, and they too, should be assessed to develop a thorough understanding of needs among the groups being studied.

The CIPP Model

The CIPP model was formulated by Stufflebeam (cited in Stufflebeam, 1983) to show how evaluation contributes to the decision making process in program management. CIPP is an acronym for the four types of educational evaluation included in the model: **C**ontext evaluation, **I**nput evacuation, **P**rocess evaluation, and **P**roduct evaluation. Each type of evaluation is tied to a different set of decisions that must be made in planning and operating a program (Context evaluation to inform planning decisions, input evaluation to serve structuring decisions, process evaluation to guide implementing decisions, and product evaluation to serve recycling decisions).

Context evaluation. The primary orientation of a context evaluation is to identify the strengths and weaknesses of some object, such as institution, a program, a target population, or a person, and to provide direction for improvement. The main objectives of this type of study are to assess the overall status of the objectives., to identify its deficiencies, and to diagnose problems whose solution would improve the well being of the objectives. A context evaluation is also aimed at examining whether existing goals and priorities are attuned to the

needs of whoever is being served. Whatever the focal object, the results of a context evaluation should provide a sound basis for adjusting the existing goals and priorities, and targeting the needed changes. A variety of measures such as interviews, observation, document analysis, diagnostic tests may be used in context evaluation. The results of context evaluation may provide information about the strengths, weaknesses, needs and opportunities and priority problems. Context evaluation could assist individuals and groups to set priorities for improvement efforts. Context evaluation records are an excellent means by which to defend the efficacy of one's goals and priorities (Stafflebeam, 1983).

Input evaluation concerns judgments about the resources and strategies needed to accomplish program goals and objectives. Information collected during this stage of evaluation should help decision makers to choose the best possible resources and strategies within certain constraints. Input evaluation deals with such issues as whether certain resources are unavailable or are too expensive, how well a particular strategy is likely to achieve program goals, whether certain strategies are legally or morally acceptable, and how best personnel could be utilized as resources. Input evaluation requires the evaluator to have a wide range of knowledge about possible resources and strategies, as well as knowledge about research on their effectiveness in achieving different types of program outcomes (Stafflebeam, 1983).

Process evaluation involves an ongoing check on the implementation of a plan. The objectives of process evaluation are to:

- provide feedback to managers and staff about the extent to which the program activities are on schedule, are being carried out as planned, and are using the available resources in an efficient manner;

- provide guidance for modifying or explicating the plan as needed, since not all aspects of a plan can be determined in advance and since some of the initial decisions may later prove to be flawed;

- assess periodically the extent to which program participants accept and are able to carry out their roles, and

- provide an extensive record of the program that was actually implemented and how it compared to what was intended, and a full account of the various costs incurred in carrying it out and how observers and participants judged the quality of the overall effort (Stafflebeam, 1983).

A variety of measures including observations, staff–kept diaries, interviews, or questjonnaires may be used to collect data during process evaluation. The main use of process evaluation is to obtain feedback that can aid staff to carry out a program as it was planned, or, if the plan is found to be seriously flawed, to modify it as needed (Stafflebeam, 1983).

Product evaluation involves the determination of the extent to which the goals of the program have been achieved. In this type of evaluation, measures of the goals are developed and administered. The resulting data can be used by program administrators to make decisions about continuing and modifying the program. Feedback about what is being achieved is important both during the program cycle and at its conclusion. The basic use of a product evaluation is to determine whether a given program is worth continuing, repeating, and/or extending into other settings (Stafflebeam, 1983).

Each of the types of evaluation described above requires that three broad tasks be performed: delineating the kinds of information

needed for decision making, obtaining the information, and synthesizing the information so that it is useful in making decisions. The first and the second steps (delineation and synthesis) should be done as a collaborative effort between the evaluator and decision maker. The second step, obtaining the information, is a technical activity that can be delegated primarily to the evaluator.

You may have noted that the CIPP model incorporates elements of the other evaluation models described above - objectives-based evaluation and needs assessment. The CIPP model is distinguished by its comprehensiveness, by the fact that it is an ongoing process, and by its purpose, which is to guide the decision making function in the program management. Although the CIPP model has been used primarily in quantitative evaluation research, there is no reason why it cannot be adopted for evaluation research from a qualitative perspective.

Qualitative Approaches to Evaluation

The quantitative approaches to evaluation, while useful, do not satisfactorily address several important aspects of evaluation. The objectives based approach, for example, tends to take a program's objectives or observed effects as givens. It does not offer much guidance if you wish to understand why particular objectives are considered worthwhile, or why certain stakeholders agree or do not agree on the worth of certain objectives.

The politics of evaluation are also not given serious attention in most quantitative approaches. Various groups have a stake in the outcomes of an evaluation study, and they may try to influence the evaluation process accordingly. Should you resist these political influences or incorporate them into the design of an evaluation study?

Another problem is that evaluations may do more harm than good under certain conditions. People generally do not like being evaluated, so the evaluation process itself might hamper the very performance that is being assessed. How can you work with the client so that the evaluation produces the most benefit and the least harm?

To address these questions and others, researchers have developed approaches to evaluation that rely heavily on the qualitative research methods. These approaches differ clearly from quantitative approaches since they do not assume that there are objective criteria for judging the worth of an educational program. Rather, qualitative approaches take the position that the worth of an educational program depends heavily on the *values and perspectives* of those doing the judging. Therefore, the selection of the individuals and groups to be involved in the evaluation is critical.

Responsive Evaluation

Stake (1983) pioneered the qualitative approach to educational evaluation. His approach, called *responsive evaluation*, focuses on addressing the concerns and issues of stakeholders. A concern is any matter about the stakeholders feel threatened, or any claim that they want to substantiate. An issue is any point of contention among stakeholders. Concerns and issues provide a much wider focus for evaluation than do the behavioral objectives that are the primary focus of some quantitative approaches to evaluation.

Four major phases occur in a responsive evaluation. These are:

- *Initiating and organizing the evaluation*: which may include the identification of stakeholders such as school board, school administrators, school teachers, students and their parents, city council staff in which the school system is

located, and influential members of the community, for an evaluation designed to formulate school policy.

- *Identifying key issues and concerns*: concerned with the identification of concerns and issues such as perceptions of citizen lockout from decisions, the abrogation of power by a small elite, centralized or decentralized school policy formulation, whether school policy is formulated by professionals or by lay groups, whether the school board should be elected or appointed, etc.

- *Gathering useful information*: concerned with the collection of information about the concerns, issues and values identified by the stakeholders. The data may include descriptive information about the entity being evaluated and about the standards that would be used in making judgments about this entity. Such information can be collected through various methods, including naturalistic observation, interviews, questionnaires, and standardized tests.

- *Reporting results effectively and making recommendations*: Frequently a case study format is used in such reports, but when appropriate, a traditional research format can be used. The report will contain extensive descriptions of the concerns and issues identified by stakeholders, and judgments and recommendations of the evaluator in negotiation with the stakeholders based on the gathered information.

Quasi Legal Models of Evaluation

Adversary evaluation and judicial evaluation rare two approaches to evaluation modeled on the procedures derived from the field of law.

Adversary Evaluation: Adversary evaluation is distinguished by the use of a wide array of data; the hearing of testimony; and, most importantly, an adversarial approach, meaning that the two sides present positive and negative judgments, respectively, about the program being evaluated.

The following are the four stages of adversary evaluation:

- generate a broad range of issues concerning various stakeholders;
- reduce the issues to manageable numbers, e.g., through priority ranking by a group of volunteers;
- formation of two opposing evaluation teams, each of which prepares an argument either in favor of or in opposition to the program on each issue; and
- conduct of prehearing sessions and a formal hearing, in which the adversarial teams present their cases before those who must make a decision about the program.

Adversary evaluation is useful in exposing strengths and weaknesses of programs and in raising questions that need to be answered. However, evaluators have also discovered that adversary evaluation has shortcomings. Its results can be biased if one of the evaluation teams is more skilled in argumentation than the other. Some evaluators modify the elements of the model to deal with these problems. However, other problems with this approach are built into the evaluation design, and hence are not easily modified. First of all, by its very nature adversary evaluation promotes a combative, "innocent vs. guilty" approach to program evaluation, which may contribute to further alienation among different types of stakeholders. Second, adversary evaluation requires a great deal of time and a

large number of people, and thus is very expensive. These problems most likely explain why very few adversary evaluations can be found in the educational research literature.

Judicial Evaluation: The judicial evaluation model (Wolf, 1983) stimulates the use of legal procedures for the purpose of promoting broad understanding of the program, clarifying the subtle and complex nature of the educational issues it raises, and producing recommendations and policy guidelines that lead to the institutional growth and /or improved practice. Unlike adversary evaluation, the judicial evaluation model does not involve a debate between two evaluation teams with victory or persuasion as the desired outcomes.

In judicial evaluation, a public presentation of the data is made, following the format of hearings in a court of law. A panel comprised of policy makers, citizens, and other interested stakeholders is convened to hear the evidence. Case presenters call witnesses who present their views in order to make a case relative to a given issue. All witnesses may be subjected to two phases of direct examination by the two case presenters. As in a court of law, opening and closing arguments are presented. After all evidence are presented, the panel deliberates and makes its recommendations.

Judicial evaluation has proved useful in both formative and summative evaluations of educational programs. While a full fledged judicial evaluation tends to be expensive and time consuming, the procedure can be scaled down without scarifying its essential quasi legal nature.

Expertise Based evaluation

The use of experts to make judgments about the worth of an educational program is a time honored and widely used method of evaluation. For example, most institutional programs in America are

reviewed periodically by accreditation boards composed of experts. If you do a thesis or dissertation project, its quality will be judged by a panel of professors because of their presumed expertise.

Educational Research and Development

Evaluation plays a key role in educational research and development (R & D). *Educational R & D* is an industry based development model in which the findings of research are used to design new products and procedures, which then are systematically field tested, evaluated and refined until they meet specified criteria of effectiveness, quality, or similar standards. It has a great promise for improving education because it involves a close connection between systematic program evaluation and program development.

One of the most widely used models of educational research and development is the systems approach model designed by Dick and Carey (cited in Gall, Borg, & Gall, 1996) which includes ten steps. *Step 1* involves the definition of goals for the instructional program which often includes need assessment. In *step 2*, an instructional analysis is undertaken to identify the specific skills, procedures and learning tasks that are involved in reaching the goals of instruction. *Step 3* is designed to identify the level of entry behaviors, or enabling skills, that learners are likely to bring to the learning task, as well as other learning characteristics (e.g., specific personality traits such as test anxiety) that might affect their learning. *Step 4* involves translating the needs and goals of instruction into specific performance objectives. Performance objectives (described earlier in this paper by the label "behavioral objectives") provide a means for communicating about the goals of instructional program at different levels with different types of stakeholders. They also provide the basis for precise planning of test items, instructional materials, and the instructional delivery system.

During *step 5* criterion referenced test items are developed. These test items can be used to diagnose and place learners, to check the individual learners' progress, and to evaluate the overall effectiveness of the instructional program in helping learners achieve the intended learning objectives. In *step 6* a specific instructional strategy is developed for assessing learners' efforts to achieve each performance objective. *Step 7* involves the development of instructional materials, which may include print materials such as textbooks and teacher training manuals, or other media such as audiocassettes or interactive video systems. If the instructional plan specifies a teacher, the teacher would develop lesson plans or guidelines for instruction this person also would be developed as part of step 7.

Step 8 involves the designing and conducting of formative evaluation. Formative evaluation is done by program developers while program is under development, in order to support the process of improving its effectiveness. In some situations formative evaluation findings instead may lead to a decision to abort further development, so that resources are not wasted on a program that has little chance of ultimately being effective. In this model formative evaluation is conducted throughout the development process, and its results are used to revise (*step 9*) any of the work carried out during the first seven steps that is, revise the instructional goals, instructional analysis, entry behaviors, performance objectives, test items, instructional strategy, and/or instructional materials in ways that appear desirable based upon the formative evaluation results.

Dick and Carey (cited in Gall, Borg, & Gall, 1996) recommend a three level process of formative evaluation:

- trying out prototype materials one on one (that is, one evaluator working with one learner);

- a small group tryout with six to eight students; and
- a field trail with a whole class of learners.

This phase of evaluation relies heavily on qualitative methods, for example interviewing and observing by the developer. Based upon the preliminary results the educational program is modified and further developed, and then tried out with the larger (although still somewhat small) number of learners. The evaluation again involves primarily qualitative methods, although quantitative methods (e.g., performance tests or self report ratings) might also be used. Based on these results, the program is refined and expanded further, and subjected to a field trial in a situation fairly close to the context in which it ultimately will be used (e.g., with a regular size class of learners). At this point, the evaluation tends to be quantitative in nature, involving tests and other measures to determine the extent to which the program is achieving its intended objectives.

When the program has completed the development process, it is subjected to summative evaluation (*step 10*). *Summative Evaluation* is conducted to determine how worthwhile the final program is, especially in comparison with the other competing programs. This type of evaluation is done by individuals other than the program developers (an independent evaluator).

Formative and summative evaluation are of great value in improving the programs and the materials that are constantly being developed in education, and in helping educators make adoption decisions about them. Unfortunately, most programs are still developed without systematic evaluation. For example, textbook publishers spend minimal time on formative evaluation of print materials with the exception of expert reviews.

Summary and Conclusion

This paper reviewed available literature on evaluation research. It tried to present the various aspects of evaluation research. Among the major ideas covered in this material are the concepts of evaluation, educational evaluation, steps in evaluation research, quantitative and qualitative models in education research and educational research and development.

Evaluation research is conducted for policy, management, or political strategy decisions. The purpose of the evaluation research is to collect data that will facilitate decision making. Of course, the findings of a research study can also be used to guide decision making; and evaluation data can be relevant to developing an understanding of a particular phenomenon. It is hoped that this paper will assist evaluation researchers to be acquainted with basic ideas of evaluation research.

References

Assessment and Program Review (2nd ed.). New York: Prager Publishers.College Record. 64, 672–683.

Cronbach L. J. (1983). *Course Improvement Through Evaluation*. In Madaus, G. F.; Scriven, M. S. & Stufflebeam, D. L. (eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer–Nijhoff Publishing.

Gall, M; Borg, W; & Gall, J. (1996). **Educational Research**: **An Introduction**. New York: Longman Publishers.

Popham, W. J. (1975). **Educational Evaluation**. Cliffs, N. J.: Prentice-Hall.

Scriven, M. S. (1983). *Evaluation Ideologies.* In Madaus, G. F.; Scriven, M. S. & Stufflebeam, D. L. (eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer–Nijhoff Publishing.

Stake, R. E. (1983). *Program Evaluation, Particularly Responsive Evaluation.* In Madaus, G. F.; Scriven, M. S. & Stufflebeam, D. L. (eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer–Nijhoff Publishing.

Stufflebeam, D. L. (1983). *The CIPP Model for Program Evaluation.* In Madaus, G. F.; Scriven, M. S. & Stufflebeam, D. L. (eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer–Nijhoff Publishing.

Stufflebeam, D. L. & Webster, W. (1983). *An Analysis of Alternative Approaches to Evaluation.* In Madaus, G. F.; Scriven, M. S. & Stufflebeam, D. L. (eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer–Nijhoff Publishing.

Wolf, R. L. (1983). *The Use of Judicial Evaluation Methods in the Formulation of Educational Policy.* In Madaus, G. F.; Scriven, M. S. & Stufflebeam, D. L. (eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer–Nijhoff Publishing.