Principles of Program Evaluation for Arts Development

By
Peter T. Ewell

This paper was presented at a seminar for Community Coordinators and Program Directors from the state arts agencies of Illinois, Indiana, Michigan, Ohio and Wisconsin which was held at Sangamon State University in Springfield, Illinois in July, 1978.

PETER T. EWELL is Senior Staff Associate
at the National Center for Higher
Education Management Systems in Boulder,
Colorado. He was Project Director for
Institutional Research at Governors State
University in Park Forest South, Illinois
and has done extensive consulting work for
arts organizations.

LIST OF TABLES

APPENDIXES

CONTENTS

Principles of Program Evaluation for Arts Development
I.   Program/Decision Making Process

A. Position of Evaluation

Like "planning" and "needs assessment," evaluation has become an important catch-word in public arts management. As programs have grown in both size and complexity, so has the concern among programmers to ascertain the effectiveness of these programs in achieving agency goals. In addition, as public spending on arts has increased, so has the concern for accountability among public officials; funding agencies, both public and private, increasingly require that evaluation be built into proposed programming as a part of the granting requirements. In a period of public sector retrenchment, evaluation processes, once established, are given increased attention. The purpose of this brief overview of evaluation principles is to guide you through the maze of evaluation processes or options available, and to enable you to make effective choices among them depending on the purpose and type of information required. The emphasis is upon <u>principles</u> because in all areas of public policy, evaluation is more of an art than a science. Program evaluations fail less often because they are "wrong" and more often because they do not meet the needs of the agency that commissioned them.

There are several reasons why an agency might want to keep a particular program operational even though the evaluation proved the program ineffective:

1.   An alternative program may not be available

2.   An existing program may be the only way of serving a particular constituency

3.   An existing program may enjoy strong political support and may be of benefit for reasons beyond its direct effect

Because of this, it is important to remember that evaluation is only one factor which influences the <u>program decision-making process</u>, and therefore information obtained through the evaluation process may not always be used, or used effectively. Evaluation research, similar to other types of policy research, should always be kept in proper perspective in relation to the entire program decision-making process.

The diagram below illustrates the place of evaluation in the context of the total decision-making process. Evaluation is represented by two transition phases, labeled appropriately, Transition Phase I and Transition Phase II, which are located at midpoint between the "doing realm" and the "investigating realm." Evaluation at Phase I serves to make the transition of <u>action goals</u> into <u>evaluation questions</u>, and at Phase II, to make the transition from <u>research information</u> into a <u>policy plan</u> for action.
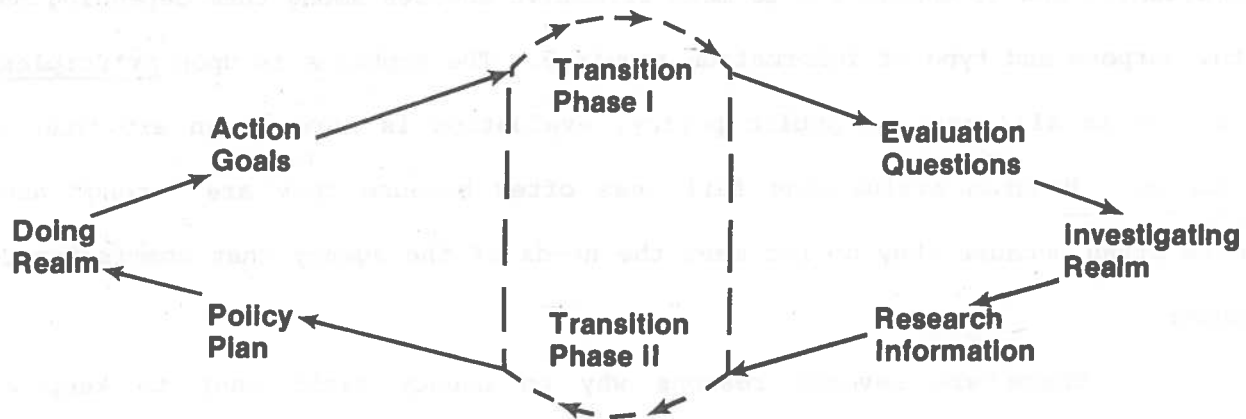
Each stage of the process is dependent upon the stages preceding it, and a failure of one means failure of all. If <u>action goals</u> are not clearly articulated from the outset, the appropriate <u>evaluation questions</u> will never be asked. Evaluation questions will be chosen on the basis of the concerns of

the individuals designing the evaluation research. In other words, questions will not be asked that will obtain information for programmers, information that will assist in program decision-making. If <u>evaluation questions</u> are not properly posed, if they are too broadly or too narrowly focused, <u>research information</u> received in response to them may be misleading. If <u>research information</u> is too detailed or too technical, though valid, it may be difficult to convert it into context or form that policy-makers can use effectively in implementation of a <u>policy plan</u>. Good communication between evaluator and programmer is essential during each stage of this total process. Oddly enough this can be especially difficult if, as in many cases, evaluator and programmer are the same person. It therefore becomes critically important that the person doing the evaluation be constantly reminded of the role he/she plays during each stage of the process. To achieve this distinction in roles, various techniques may be employed:

1. A programmer could organize tasks routinely in order that one day a month is devoted strictly to evaluation and the role of evaluator

2. Separate evaluation files should be established at the outset of the project

3. Memos relating to evaluation should be written in third person insuring the objectivity needed to play the evaluator role

Although these techniques may sound somewhat schizophrenic, they have often proved to be beneficial in alleviating imbalances caused by placing concentrated effort in the "doing realm" at the expense of the "investigating realm." Another major advantage to this conscious shift in perspectives is that it tends to continually bring to attention the question of purpose of evaluation.

B. Purpose of Evaluation

Like all types of policy research, evaluation can be done for a number of reasons. Each reason is valid and each makes its distinct contribution to the entire policy planning, decision-making process.

1. Political Purposes

Much evaluation is initiated for political reasons. Because of this, the goals of the program and the goals of the research will be identical: to gain wider acceptance for the existing program. This implies that all stages of the program decision-making process, as illustrated previously, beyond the establishment and articulation of action goals, are in a sense irrelevant. Research becomes a kind of showmanship, useful to the programming agency to enhance credibility or to demonstrate concern. It is a well known axiom in federal social service programs that commissioning a study is an effective strategy to follow if additional time is needed by an agency to solve a political dilemma. It is also an effective strategy to use as proof of concern for a particular constituency's special interests. What better way to gain confidence of special interest groups than to commission an evaluative study of a program in their area of concern? In addition, at the local level, quality evaluation often presents a positive image of organizational competency and efficiency, and therefore signifies an organization or agency worthy of support.

Given such a situation it is easy to become cynical about all aspects of the evaluation process, but it is important to point out that the political orientation of most evaluation research has significant benefits as well. First, constant awareness of the political context keeps research relevant. Political factors should not be allowed to influence technical aspects or

procedures when answering research questions, but they are appropriate to include initially when formulating the direction of questions and when translating the obtained answers into meaningful action-policy plans. Second, good research can heighten the self-confidence of a programmer facing a political situation; the outcome obtained from well designed evaluation enhances the programmer's knowledge concerning the program, and this can be an advantage in communicating the program's effectiveness or, if necessary, ineffectiveness to others. Finally, effective communication itself is often a result of the political context of evaluation research. Political considerations often lead to clarity and directness in stating evaluation results; they demand that the results of evaluation be documented. A concise documentation of the evaluation and its results, is of course beneficial above and beyond its political context. It is thus completely valid to state that program evaluation is political. But rather than dismiss it for that reason, it is important to recognize the advantages that the political context offers to the evaluation process. At the same time it is equally important to minimize those political considerations when conducting the technical aspects required during the process.

2. Monitoring Purposes

A common reason for evaluation is to see if a program has been implemented correctly and is conforming to program guidelines. This is generally practiced regularly by spot-checking or monitoring a program, and occasiohally entails a limited assessment of a program's effects. Limited assessment might answer questions such as: How many people have been served? How much income has the program generated if it claimed to be partially self-supporting? What are the general attitudes and perceptions of the public toward the program? Questions like these can be answered in a broad variety

of ways, from careful record keeping to extensive field surveys. These techniques will be discussed in detail later. The techniques are similar in that they all produce answers that are straightforward pieces or "bits" of information. In seeking "bits" of information such as audience size, we are not ordinarily interested in researching all possible reasons why people attend performing arts events; we are interested only in an estimate of how the program is doing at that time. Nevertheless, it is always important to carry out this kind of program monitoring in a careful and consistent manner. If adequate records are kept, it may be possible to perceive patterns developing over time. When practiced informally, perceiving such patterns is what we term experience. Formalizing the same process allows us to be a bit more systematic, attempting not to replace, but to supplement experience. A good rule of thumb is to collect each piece of program monitoring information in a form that will be useful in ten years as an indicator of an organization's progress and program performance.

3. Assessment Purposes

The ultimate purpose of all evaluation is to produce better programs. This implies going beyond a simple success/fail assessment of program performance and adopting a broader causal perspective. Why are certain elements of the program producing certain effects? If program monitoring reveals that a particular artist-in-residence program is having greater impact in urban neighborhoods than in rural neighborhoods, an immediate program decision might be to concentrate the program in urban neighborhoods. An alternative would be to find those factors which contribute to this difference in impact, the answer being extremely beneficial in developing future programs that will be effective in rural areas. Questions like these, which pertain to the cause of difference, are more complicated

than those pertaining to program monitoring. Answering them will demand careful planning and sound research design. In large part, the principles involved in approaching such questions constitute the remainder of this paper.

II. Policy/Program Planning Process

A. Program Monitoring & Impact Assessment

Evaluation is a technique which means, quite simply, to assign something a value. If we cannot state the potential of a given program, evaluation will be irrelevant. As stated earlier, this indicates that a primary concern is goal definition. It is often the case, especially with arts programs, that goals are not formally established and articulated until after the fact. To be maximally effective, an evaluation must be related to an explicit planning process where the goals have been established and defined. Once this has been achieved a programmer must answer two questions which provide a starting point for the evaluation plan.

The questions to be answered are:

1.   What should be done?

2.   What results are expected?

The first question provides the foundation for program monitoring. Evaluation questions designed to answer this question will assess, by degree, whether the program is meeting its stated plans or goals. The second question provides the foundation for assessing impact. It attempts to measure, by degree, the comparison between what was expected to happen and what actually was accomplished.

B. Sample Plan: Senior Citizen Program

To consider, for example, the plans of a local arts council to develop a comprehensive program for senior citizens. Assume that the council's expressed goals are:

> 1. To "help children and adults develop a better sense of their own worth through the encouragement of creative skills and talents"
>
> 2. To "provide opportunities for everyone to enjoy and understand the insight and vision of the world's greatest artists"

Based on these goals, the council develops its "Senior Citizen Program" and defines the following strategies and objectives:

STRATEGIES	OBJECTIVES
1. Watercolor Workshop Series	a) to conduct 3 workshops between April–September 1983
	b) to target attendance of 25 people per workshop
	c) to obtain local professional artists to conduct the workshops
2. Bus trips	a) to schedule a trip to Minneapolis to see a performance by the Guthrie Theatre in Fall, 1983
	b) to schedule a trip to Spring Green, Wisconsin to see a performance by American players
	c) to provide lunch or dinner on each trip
	d) to insure that the bus on each trip will be at least 2/3 full

It would be possible to break down these objectives into smaller, more specific components, but the point is clear that the task of constructing a program monitoring scheme for evaluation is straightforward. All that is now required is periodic checks to insure the program is on track and on schedule. A commonly used timetable is monthly and quarterly checks. For more complex programs, a monitoring chart such as that shown in Table 1, page 9a, may prove valuable.

As the Table illustrates, program monitoring is an occasionally tedious, but in principle, extremely straightforward activity which often yields a great deal of information that is immediately valuable in implementing a program. The above discussion has not, however, provided insight into the second evaluation question: Did the program meet the objectives and achieve the goals outlined at the outset? Answering this question requires a somewhat different methodological approach. A conceptual model of this approach shown below exhibits the relationship between the programming agency and the public.

"Program Monitoring"

**Program
Output**

Agency   **Public**

**Feedback
Input**

"Impact Assessment"
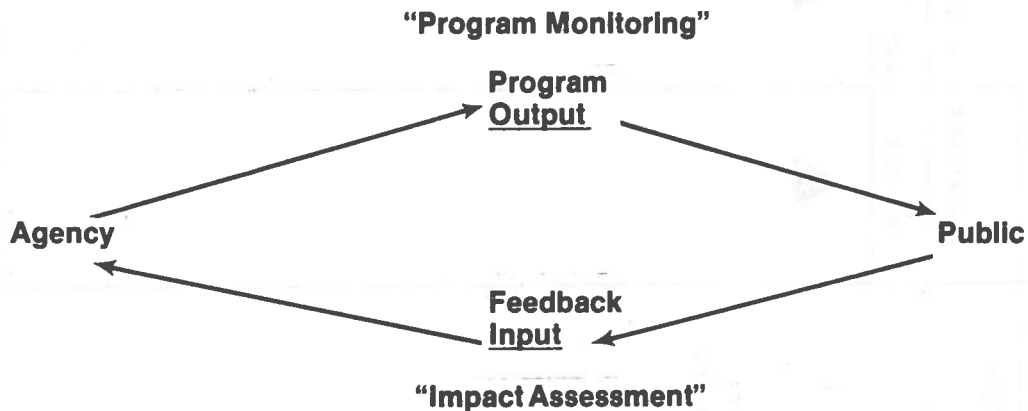
Agencies generally have a great deal of control over their programmatic output, but ordinarily their feedback or informational input from the public is, at best, fragmentary. In our example, the local arts council has control

Table #1 • Sample Program Audit Form

STRATEGY: SENIOR WATERCOLOR WORKSHOP	Target Date (Quarterly)				Period covered: Jan. '83 – Dec. '83
	1 quarter Jan–Mar	2 quarter Apr – June	3 quarter Jy – Sept	4 quarter Oct – Dec	Comments
1. Conduct (3) workshops by October '83	▲	▲	△→▲	▲→	Third site for workshop not available on target date; rescheduled and held 10/14/83
2. Develop list of Artist/ Instructors available for future expansion of existing program			▲		Completed 7/16/83
3. Conduct evaluation of Instructors				△	

Fill in triangles when task is completed. △ → ▲ → △ indicates rescheduling.

9a

in selecting location and in scheduling times for the Senior Citizen Watercolor Workshops, therefore it can determine its specific expected output. Yet the council has not developed channels which provide systematic feedback or informational input from the senior citizens, therefore assessment of the program's impact on the public can not be measured and/or compared with what was expected. Consequently, the program's effectiveness cannot be validly determined.

Over time, an experienced programmer will develop a "feel" for his or her program, especially if constituents are vocal and communication channels are open. This asset is not to be underestimated in formal evaluation. However, its major disadvantage is that the programmer may become too involved with the program on a day-to-day basis to perceive validly the program's broader effects, especially if some of those effects are negative. Ordinarily, the only kinds of effects which emerge from this type situation are gross effects; attendance falls off considerably, or a great deal of explicit praise or criticism is received concerning a particular artist or particular workshop setting. This obvious feedback will not escape the programmer's attention, yet there may be a considerably larger number of effects which contributed to these outcomes that the programmer knows nothing about. Though these obvious indicators, decrease in attendance or explicit praise or criticism, are measures of the success or failure of the "Senior Citizen Program," they fail to indicate why the outcomes occur.

The logic on which evaluation is based is that evaluation acts as a mediator or monitor not only between the output from agency to public, but again during input or feedback from the public to the agency. The interchange between the agency and the public and the position of evaluation as mediator in the policy/program planning process is illustrated on page 12.

In the model, evaluation appears as a pair of boxes, $\boxed{\text{BOX I}}$ and $\boxed{\text{BOX II}}$. Each represents a distinct body of evaluation procedures and techniques which will be described in detail later. For the moment it is important to recognize that though the two are distinct, they are dependent on each other and are an integral part in the policy/program planning process. Because of this, their relationship and integration must be perceived and planned for at the outset of the program in order to be effectively utilized in implementation and evaluation of that program.

III. Program Evaluation Process

To continue, there are two general types of evaluation. $\boxed{\text{BOX I}}$ which in applied research is called "policy analysis," and $\boxed{\text{BOX II}}$ which is called "outcome evaluation." During "policy analysis" the object is to answer the question, What to do? Based on the goals of the agency, the evaluator analyzes a number of possible alternative solutions and makes informed recommendations to policy/program planners regarding the course of actions which should be taken. While "policy analysis" is most often discussed in the context of the policy/program planning process, its relationship to the program evaluation process is its utility in further clarifying a program's expected outcome. $\boxed{\text{BOX II,}}$ "outcome evaluation," is designed to answer the question, What happened? Its object is to assess the impact or feedback of a particular program that has been implemented. Note how the two types of evaluation overlap and depend on each other.

First: It is extremely difficult to answer the What to do? question without knowing something about What happened before?, or the previous impact of the program. In the case of a new program, the policy/program planner will want to know as much as possible about the impact of prior, similar programs.

$\boxed{\text{BOX I}}$

Second: It is impossible to answer the What happened? question effectively without some knowledge of program goals. What was

| BOX II | anticipated or expected to happen? What should have been the results? What were the "trade-offs" among alternatives that had to be made initially when deciding <u>what to do</u>? |

Because of this, it is unwise to undertake either type of evaluation in isolation. Despite the fact that each relies on somewhat different logic and employs a distinct set of techniques, each is dependent on the other. To illustrate, note again the model below which exhibits each type of evaluation, BOX I and BOX II , and its position in the total context of policy/program planning as well as each distinct type of evaluation and its interrelationship in the overall <u>program evaluation process</u>.

Policy/Program Planning Process
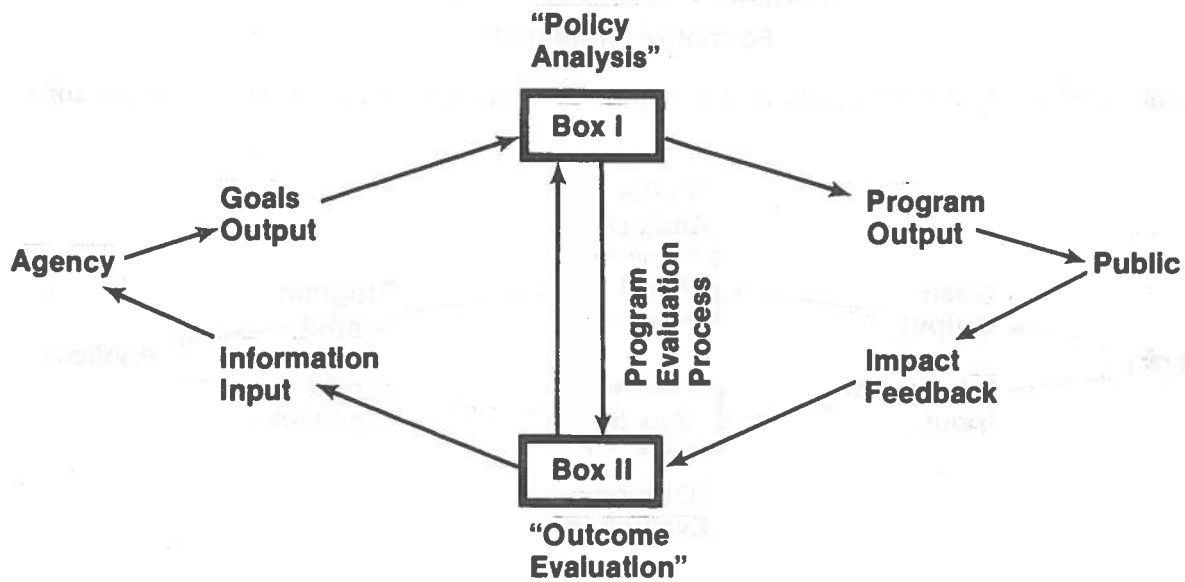
A. Policy Analysis

The two types of evaluation methods and techniques, BOX I and BOX II are distinct and different in degree depending on the purpose of evaluation and what it attempts to accomplish. In our daily lives the two are distinct to a lesser degree. In fact, we tend to merge the two processes completely. The choices we make are intimately bound to prior personal experience and we immediately try to modify our actions bit by bit until we achieve a

satisfactory outcome. The formal counterpart to this example of merging the two types of evaluation is what researchers term, <u>formative evaluation</u>. The primary object is not the assessment of the program's total overall effects, but the development of an effective program strategy. In other words, a relatively low degree of outcome evaluation BOX II , constantly feeds information into policy analysis, BOX I , yielding a program which is slightly but continually being modified. Note the diagram below. This is an excellent strategy when existing programs have proven effective and the intent is to achieve optimum or improved effectiveness.

Program Evaluation Process
Formative Evaluation

(Closer degree of integration between policy analysis and outcome evaluation)
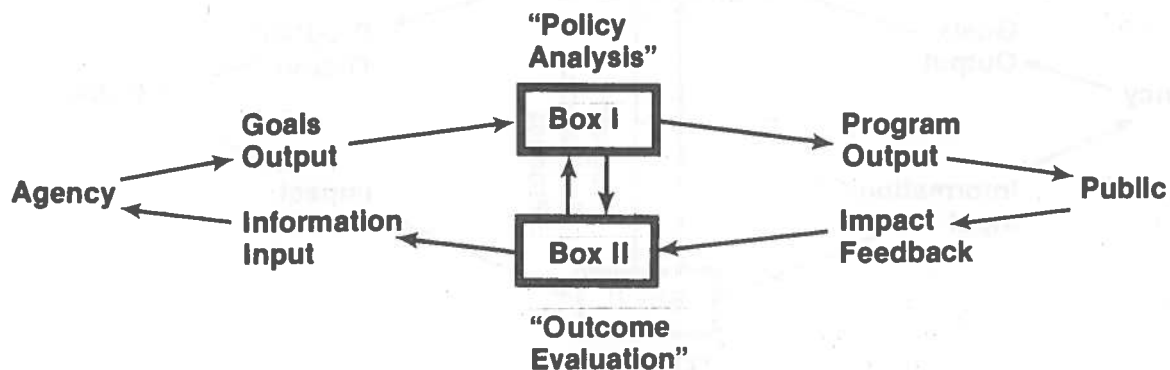
Formative evaluation of the Senior Citizen Watercolor Workshops might include regular, informal conversations with the artist and individual participants. Feedback of ideas generated from theses conversations would be used immediately to increase effectiveness in the implementation and scheduling of workshops. For example, variation in type, size and location of workshops, plus a difference in the combination of these variations, could be systematically planned at the outset of the Senior Citizen Program. This would help to evaluate which combinations seemed most effective and

comfortable for both artist and participant. Formative evaluation is specific and immediate. It is a process of continual, minimal development. It does not and can not evaluate the overall extent to which the workshops achieve the broad, initially stated goal of the program, that being: "to develop in Senior Citizens a better sense of their own worth through the encouragement of creative skills and talents." Since formative research requires continual program modification and interchange between policy analysis and outcome evaluation, both components of the evaluation planning process, it is generally more effective if done in-house as opposed to contracting outside professional help (except perhaps for an initial consultancy.) The key to successful formative research lies in systematic variation. In other words, plan systematically the modifications of the program in order that their specific effects can be determined. Too many variations in the program at one time can cause confusion in determining which modifications are responsible for which outcome.

B. Outcome Evaluation

If the object of evaluation is to assess the broad impact of an entire program's long-term effects rather than "fine tuning" an already successful program in an effort to further develop its effectiveness, a more distinct separation between the policy analysis and outcome evaluation components of the program evaluation process is necessary. The diagram on page 12 exhibits this separation and distinctiveness. Researchers define this as summative evaluation. In contrast to formative, summative is generally conducted most effectively outside the programming agency. This is the case for several reasons. First, summative evaluation is often highly political. For example, if the Senior Citizen's Workshop was discontinued due to in-house evidence

14

that the program "doesn't work," ultimately questions and problems will arise from workshop participants and supporters who want verification of the "doesn't work" statement and discontinuation decision. Second, summative evaluation must be free from bias and the evaluator must be able to openly examine what he/she feels are unproven or unquestioned program assumptions. This can only be accomplished if fear of repercussion or fear of jeopardizing the program or one's own position can be eliminated. If the evaluation is conducted in-house this is usually not possible. Third, the program must be evaluated as it stands, or as it exists at a given point in time, without further modifications. Immediate modifications for improvement are the purpose of formative evaluation, not summative. If an inside programmer conducts the summative evaluation, there is the temptation, as the evidence is accumulated during the evaluation process, to make suggested program modifications. This jeopardizes the purpose and outcome of the summative evaluation. To conclude, the purpose of the evaluation effort must be established and understood prior to deciding and/or choosing and an outside evaluator.

To summarize:

1. Evaluation is only one of many resource tools to use in obtaining information or input necessary for the total program/decision-making process.

2. Evaluation is irrelevant or unnecessary if:

    a. an alternative program is not available
    b. the program can't be changed and remain consistent with its purpose and goals
    c. the future of the program is predetermined politically

3. The purposes of evaluation are:

    a. political
    b. monitoring
    c. assessment

4. Program goals must be defined and established; if the purpose of the program is not clear, evaluation is impossible.

5. If the purpose of the evaluation is for program development or monitoring (formative), evaluation is best conducted in-house.

6. If the purpose of evaluation is impact assessment (summative), the evaluation is best sub-contracted to a professional researcher.

7. Evaluation is a process not a product, a component of the larger decision-making process; be flexible but remember the ultimate purpose of the evaluation.

IV. Policy Analysis

Choosing effectively among alternative courses of action has for many years been the prime concern of management science. It is not the purpose of this manual to detail at length that area of research. However, the techniques used by policy analysts to delineate the goals as program outcomes will be reviewed. As previously stressed, if program goals remain unspecified or undefined, meaningful evaluation is impossible. Furthermore, until program goals are established in measurable terms of concrete outcomes, it is virtually impossible to approach the primary question of impact assessment: What actually happened compared with what was expected to happen?

The best definition for an outcome is any concrete change in the environment resulting from a program, whether that change was intended or unintended. While goals are often phrased in broad, not easily defined, conceptual terms, outcomes or the changes must be stated in concrete terms relating to "real things." For example, if providing dinner for a family is a goal, four full stomachs is an outcome. It is important to note that while the outcome is concrete, four full stomachs, it is not necessarily quantitative: How much is a full stomach? Another example: an outcome of an artist-in-residence program may be a change in attitude in the minds of the

people who come into contact with that artist. While this change in attitude may be only indirectly detectable in a variety of ways, it is, nevertheless, a concrete real change in the way a group of individuals think and behave.

One of the goals of the Senior Citizens Program was to help senior citizens develop a better sense of their own worth through the encouragement of creative skills and talents; the strategy for fulfillment of that goal was a series of watercolor workshops conducted by local professional artists. The evaluation of the success of the workshops depends on whether or not senior citizens develop a better sense of their "own worth" which, though concrete, is not stated in terms that can be measured. How can the workshop participants' "self-worth" be measured? How is it defined in measurable terms? One dimension of "self-worth" might be an individual's feeling of self-confidence which may be expressed and manifested by his/her willingness to ask questions during the workshop sessions, to participate and share work with others and to continue the same behavior in similar workshops in different locations. Another method of measuring the increased sense of "self-worth" might be the participants' use of their prior experiences as themes expressed and developed in their art work. In both cases it also becomes necessary to consider methods of stating and measuring the degree to which these outcomes occur. However, this concern is secondary to the initial step of thinking of goals and relating them to terms that define, in real things, a concrete change of state, relating "self-worth" to questions asked in workshops or relating "self-worth" to paintings of prior experiences.

Once all anticipated outcomes of a program have been listed, it becomes necessary to prioritize, to decide which outcomes are preferred. Tools which can be useful to policy analysts in making these decisions are cost/benefit analysis and various rank order techniques.

A. Cost/Benefit Analysis

Consciously or unconsciously, most decision makers tend to use a cost/benefit approach to the problem of prioritizing anticipated outcomes. The principle of cost/benefit is easy to state, but often quite difficult to put into practice: the "best" outcome is the one which gives the maximum benefit for a given cost, or which provides a given benefit at a minimum cost. If, for example, a community arts council's proposed membership drive is likely to yield $10,000, whereas a corporate giving campaign is likely to yield $15,000 for the same amount of staff time and administrative support, it is clear what the decision should be, all other things being equal. Yet, there may be other reasons why we should consider the membership drive over the corporate giving campaign; a membership drive may provide increased publicity or political support and these intangibles should also be considered a "benefit" on the positive side of the equation. On the negative or "cost" side of the equation, acceptance of corporate funds may lead to some undesirable side effects, such as increased dependence upon a narrow funding base and, therefore, greater difficulty in maintaining decision-making independence. Any small organization that has accepted a CETA employee learns quickly that the benefits of that extra staff employee are often outweighed by the cost of becoming administratively entangled in a mass of government bureaucracies. Clearly, such intangible costs and benefits are very real considerations in the program decision-making process. Though they are not as susceptible to quantitative treatment as monetary costs and benefits, assigning value to intangibles is a primary concern in policy analysis.

B. Rank/Order Techniques

Most cost/benefit analysts attempt to assign monetary value to intangible outcomes--outcomes such as increased publicity and community service. The usual method of doing this is to roughly assign them market values or utilities. Market values are how much you would have to pay someone to supply you with that benefit and utilities are how much you would be willing to pay for that benefit. A useful tool for facilitating group consensus in prioritizing a program's anticipated outcome is to request that each member develop a budget based on an arbitrary amount (usually a thousand dollar budget) and state how much they would allocate for each anticipated program outcome. These individual rankings can then be immediately translated into group priority rankings. In addition to providing a simple first, second, and third ranking of priority outcomes, this method also indicates the amount of difference, in degree, between first and second or second and third ranking priorities. For example, there may be a greater gap between second and third priority rankings than between first and second priority rankings, something that simple rank order would not indicate. Normally, individuals can meaningfully rank order only about seven things, therefore this method of prioritizing outcomes in a group situation offers the advantage of handling and prioritizing a larger number of anticipated outcomes.

Another technique used to prioritize complex or large numbers of anticipated program outcomes is to systematically choose between pairs of anticipated alternative outcomes--alternatives such as: $15,000 and no publicity, or $10,000 and a week of local media exposure. (Note again that it is important to define these choices in concrete measurable terms--terms like, a week of local media exposure as opposed to simply stating, increased

publicity.   Similar to an elimination process in a wrestling tournament in which winners of one round take on winners of the next, this prioritizing technique involves an elimination process whereby the winning priority outcome from one pair of alternative outcomes would be compared and voted against the winning priority outcome from another pair of alternatives.   Another technique, similar to a round robin, is one in which each outcome is voted against every other alternative.   Both of these techniques have flexibility and, as such, tend to be used in conjunction with more conventional group discussions or ranking procedures.   Neither should be undertaken without a good deal of preliminary or subsequent discussion.

Other techniques for prioritizing and ranking anticipated outcomes are more structured.   Nominal Group Technique (NGT) is one formal method for discovering and mapping group priorities.   It is especially useful in discovering outcomes that had not been articulated or established initially, something that the less structured techniques, as mentioned above, are unable to do.   The Delphi Technique is similar and since it is based on a written questionnaire it can also be sent out to obtain input from a broader geographically disbursed decision-making group.   For further information on Delphi and NGT see Group Techniques for Program Planning listed in Appendix A.

There are other techniques just as appropriate that can be utilized to prioritize and to translate goals into concrete measurable terms for program evaluation.   Currently, program planners in arts development often use a combination of such techniques that can then be easily modified to fit the program evaluation process.   If the program to be evaluated is a product or outcome of a well developed planning process, evaluation at BOX I , Policy Analysis, may prove redundant.   Remember, the object of policy analysis in the evaluation process is to answer the question, "what to do?" (p. 11)   If goals

have already been articulated and prioritized in the overall <u>planning process</u>, it may seen unnecessary to do so again in the <u>evaluation process</u>. Yet, in some cases, the initially stated program goals and their associated outcomes may have "drifted" considerably from their original intent, so it is wise to incorporate goal articulation techniques into the first phase of the program evaluation process, that is, policy analysis. Again, <u>policy analysis serves to monitor and check the program</u>, insuring that it is directed toward achieving the originally stated goals.

To summarize:

BOX I : <u>Policy Analysis</u> in the Program Evaluation Process.

1. To insure that a program is designed to facilitate accurate program evaluation, the primary concern is to state goals of the program in terms of concrete outcomes.

2. A goal is a general statement of program intent; an outcome is a change of state in "real things" (individual or organizational) occurring as a result of the program.

3. It is important not to restrict measurement of outcome to tangible benefits alone; <u>measurement of intangible benefits must be considered</u> and therefore the main concern in program evaluation is to translate these intangible benefits into concrete terms.

4. Identifying cost and benefits of anticipated outcomes facilitates decision making between alternative outcomes.

5. Monetary value, <u>market</u> or <u>utilities</u>, can be assigned to intangible benefits to facilitate measuring anticipated outcomes.

6. Formal techniques for goal/outcome ranking and prioritizing facilitate group decision making, but should be use flexibly and in combination with unstructured group discussion.

7. Program planning and program evaluation overlap continually; if the <u>planning process</u> is well organized and designed, the <u>evaluation process</u> will be equally effective.

8. Goals are not absolute; as evaluation questions are developed and as the first indicators or a program's

effects are perceived, goals may have "drifted" and may require redefinition or may need to be restated.

V. Outcome Evaluation

Once a program's goals have been established and prioritized into the most desired outcomes, concerns shift to $\boxed{\text{BOX II}}$ of the evaluation process, _outcome evaluation_. What were the results, the _actual_ outcomes of the program compared with the anticipated or expected outcomes? Did the program actually do what it was designed to do? These seem simple questions, but the answer is not in simply finding out what actually happened.

For example, assume a Summer Theater Program introduces a new method of ticket promotion. As the summer wears on, there is a noticeable increase in attendance. Is the new method of ticket promotion responsible for the increase, or are there other factors, like weather, which are responsible? In order to analyze the influence of these outside factors on the increase of attendance, they must be somehow compared with a _control_. How does this summer's attendance differ from last summer's? How does it differ from that of other theater programs in other places doing the same kind of things? Are these differences a result of outside factors which are unrelated to the ticket promotion scheme? If we analyze all of these outside factors and still find that they are not sufficient reasons responsible for increased attendance, it can be stated, with the greater degree of confidence, that the new ticket promotion method is the factor which is responsible. This process is the essence of _outcome evaluation_.

A. Controlled Experimental Design

As stated, a general model for policy analysis is a cost/benefit approach; for outcome evaluation a similar general model is the _controlled_

experiment. The logic is simple: If two situations which are identical in all respects but one produce outcome that are different, that difference can confidently be attributed to the one distinguishing factor that was different. The difficulty is that outside of a laboratory setting, situations that are identical in all respects but one are rare. (Though experimental psychologists claim otherwise, even in their field, the occurence of such situations are not the norm.) Returning to the example, the only way to insure the credibility in concluding that the new ticket promotion is indeed responsible for increased attendance is to run two parallel seasons, one with the new ticket promotion and one without, and to make sure that all persons attending the one season are not aware of the other. This would be a model experiment but quite difficult to put into practice. Nevertheless, the principles behind such controlled experiments are what are important and they should always be approximated in outcome evaluation. Though most experimental designs are impractical in a field setting, it is necessary to make every effort to replicate these designs as closely as possible. The closer the designs are replicated, the greater confidence in stating evaluation outcomes.

The classic experimental model for program evaluation has three distinct elements: 1) (stimulus) program, 2) (response) outcome, 3) controlled environment.

Experimental Model

```
                              ?
 (Stimulus) Program ─────────────────────────→ (Response) Outcome

 ◄───────────────── Controlled Environment ─────────────────►
```
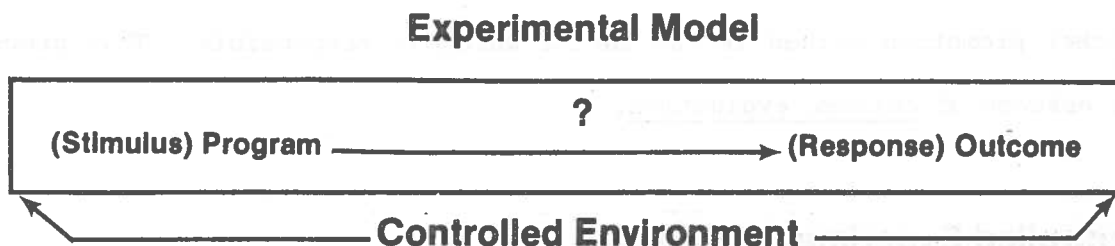
Each of the three elements of a classic experimental model should be examined independently.

1. Program: Most programs can not be defined as single, measurable "stimuli" but are, in fact, a complex combination of elements or factors extremely difficult to separate and distinguish. For example, was the favorable "response" to the local art council's Watercolor Workshops due to the input of the artist/instructor, to the informal interaction among participants themselves, to the facility or physical environment, to the program's psychological effect of providing legitimacy to the participant's personal expression or was it due to various combinations of these factors? Although it may be possible to prove that a program, in its entirety, produced a given response or outcome, it is hard to determine which elements or combination of elements had a greater influence than others in that final outcome. Determining this becomes an easier task if, during the planning process, goals were translated and articulated as concrete outcomes and if they were translated into a range of indicators that can be used to measure those outcomes. Disentangling or separating all elements that make up a total program is not always possible, but it is important that the evaluator is aware of all the elements of the program when trying to assess its impact.

2. Outcome: Outcomes are changes in state of real things as a result of the program. The measurement of the outcome is difficult especially when these are stated in terms of intangibles. In the Summer Theatre example, relating to its

24

new method of ticket promotion, the outcome is straightforward and measurable; increased attendance can be counted. In the Senior Citizen Program, the outcomes were not as straightforward; increased "self-worth" is not as easily measured or counted. This is the challenge of outcome evaluation--translating the intangibles like "self-worth," into a range of observable indicators that can be counted. Table 2, p. 25a, lists a range of observable indicators of "self-worth." The format is a useful device for structuring and facilitating impact assessment in the evaluation process. It is most important to determine whether these indicators validly represent increased self-confidence; do they, in fact, measure "self-worth?" Table 2, Indicator 1a, the "increased willingness of participants to ask questions" may not be a valid indicator of increased self-confidence but could simply indicate greater ease in student/instructor interpersonal communications due to increased familiarity over time. Just as important as determining validity, is determining whether these indicators are reliable measures of self confidence; do they repeatedly and consistently indicate a change in self-confidence? Table 2, Indicator 1e, the fact that "a participant consistently stays and paints after hours on scheduled days of the workshops" may simply be due to the fact that the participant, on those days, waits for a ride home.

Since there may be doubt about the validity or reliability of a single indicator, it is best to provide a broad range or indicators, each measuring the impact from a different

Table #2 • Impact Assessment Outline

Goals	Outcomes	Indicators
(What is to be accomplished?)	(What are the "changes in state" of things as a result of the program?)	(What proof is there of "changes in state?")
Assist senior citizens to develop a greater sense of self-worth through partici-patory arts experience.	1) Class participants increase self-confidence	1a) Increased willingness to ask questions of the instructor
		1b) Increased willingness to share and dis-cuss artwork with each other
		1c) Over time, responses to questionnaires mass mailed to senior citizens indicate fewer respondents answering "too old to be interested," and fewer questionnaires being returned unanswered.
		1d) Participants who are residents in homes for the elderly begin insisting on more creative and recreational activities in their places of residence
		1e) Over time, participants continue to work after hours on their art works on scheduled workshop days
		1f) Participants respond on evaluation ques-tionnaires that the benefits of "height-ened self-confidence" exceed their expectations
	2) Class participants become more self-expressive; they articulate their values based on personal and prior experiences	2a) Artistic themes indicate a move from objective topics or symbols to symbolic representations based on subjective events and experiences
		2b) Self-expression expands to other art forms and disciplines--oral history, music, sculpture or poetry

perspective. Assessment of impact then becomes a matter of "weighing the evidence" from a variety of sources. An indicator can be either obtrusive or unobtrusive. If program participants are aware of the indicator, it is obtrusive; if they are unaware, it is termed unobtrusive. Participants' increased willingness to ask questions, or participants working on their paintings after hours at workshops are both unobtrusive indicators. Table 2, Indicator 1f, "rating of one's own heightened self-confidence" on an evaluation questionnaire, is considered obtrusive. A classic example of combining both obtrusive and unobtrusive indicators is a study initiated by a museum director to determine which exhibits were most popular among six-year old audiences. (See Webb, Campbell, Schwartz and Sechrest, Appendix A) The obtrusive indicators were interviews conducted with every tenth child; to collect unobtrusive indicators, maintenance staff were instructed to observe and record the amount of hand and nose prints on the glass exhibit cases. Taken independently, dirty and smudged exhibit cases have many explanations, but combined with the data from interviews with children and the data from interviews with teachers plus the evidence from a range of indicators cited in the study, the museum director was able to make a fairly valid and reliable judgment as to six-year old audience preferences.

3. Controlled environment: Gaining experimental control insures greater confidence in stating that the "stimulus" or program is

26

actually the cause of the "response" or outcome. In most experimental situations, control means comparison. Program outcomes must be compared with one another; utilizing comparisons relative to a concrete rather than abstract standard assures validity when stating one program's outcome is better than another. Comparison is imperative in proving a point in question, yet this is not recognized by the majority of consumers, creating a situation advertisers use to their advantage. Consider the frequently used slogan: "Brand A antacid consumes thirty-seven times its weight in stomach acid." This may be a true statement; Brand A may, in fact, consume thirty-seven times its weight in stomach acid. It is not known, however, whether something else, roast beef or hot cross buns perhaps, could consume more stomach acid, ounce for ounce, than Brand A. No comparison has been made with some concrete standard and so it can not be assumed that Brand A is, in fact, that great an antacid. Regrettably, similar statements are often used in reports on the impact of arts programming. For example, what information is obtained from the statement, "76% of those surveyed favored an increased public support of the arts?" The statistic is not that impressive unless it is known that respondents were asked to choose among alternatives and indicate which they were most in favor of supporting with an increase in public funding. It is also important to use standards of comparison that are appropriate. Consider another advertising slogan: "75% of the women who use Brand X dishwashing liquid have hands as soft as

women who don't do dishes." The comparison is made between women who use Brand X and women who don't do dishes. Unfortunately, as consumers, the choice is not between purchasing a brand of detergent and/or choosing instead not to do dishes. The choice is, should the consumer purchase Brand X or another brand of dishwashing liquid? The advertising statement has little value; it does not state or make a comparison which proves Brand X to be a better choice over another brand of dishwashing liquid.

In continuing the discussion on "experimental" design, to decribe particular approaches, the following notations will be used to designate research design components:

$$X = \text{(stimulus) program}$$

$$O = \text{(response) observation of outcome}$$

The ideal experimental design, complete with comparison group, would be presented as indicated below:

EXPERIMENTAL CONTROL GROUP   O X O

COMPARISON GROUP   O   O

In the Summer Theatre example, suppose the new ticket discount program needs to be evaluated. Using the above model, X would designate the ticket discount program and O Observations prior to program implementation and after implemention.

In principle, following the controlled experimental design would require measuring the impact of the new ticket discount program by offering it to a given group of people, the experimental control group, and not to a similar group of people, the comparison group. Ticket purchasing patterns of both groups would be observed and recorded prior to introducing the new program to

the experimental control group and again observed and recorded after the introduction of the new discount program. In practice, this may prove difficult. If the comparison group found out about the new discount program, they could object to unfair treatment and demand the discount option as well. This illustrates a major paradox in outcome evaluation of public programming. Public programs evolve out of a need from a specific public and are usually a benefit to that public, yet political pressure demands that the benefit be extended to all publics. If all publics are exposed to the program, a situation arises that makes a true or valid assessment difficult because there is no base or standard for comparison--no comparison group. Evaluators try to solve this problem by taking a political position and arguing for application of the program to one specific group while withholding it from another, but this must be handled with caution. An example; it may be feasible to argue that the ticket discount program should only be extended to senior citizens or to arts council members, but this then jeopardizes the validity of the experiment. It then is not known whether the increase in ticket purchase is due to the new discount program, or due to some other factor associated with being a member of these particular groups. Ideally the most valid method for choosing two groups would be to conduct a lottery termed random sampling in applied research. If implemented sensitively and creatively, it can be politically feasible. Therefore, in practice, it is usually difficult to follow the ideal controlled experimental model described above. It then becomes necessary to replicate the true or controlled experimental model as closely as possible by using "quasi-experimental" designs for evaluating response or outcome.

B. Quasi-Experimental Designs

Quasi experiments are research designs in which the groups involved in the program have not been scientifically selected by random sampling and where there is no comparison group. Instead of comparison groups it is often necessary to make comparisons between programs, either the same program under different conditions, or different programs applied to same subjects.

1. <u>One-Shot Post-Test</u>   X   O

Program evaluation in the arts, and elsewhere, can follow a wide range of these experimental designs. Although weak, one of the frequently used designs is the One-Shot Post-Test noted above, X being the program, O the response or observation of the outcome. This indicates that after the completion of the program, observation of the effects is conducted but there has been no observation made prior to the program, no <u>pre-test</u>, and there is no <u>comparison group</u>. This design is one used frequently by arts evaluators since the decision to evaluate is usually initiated after the program has been implemented. The One-Shot Post-Test must be used due to inadequate program planning. Though unfortunate, this is a typical situation. Lack of a <u>pre-test</u> and a <u>comparison group</u> means that it becomes increasingly difficult to validate or prove that the observed outcome (O) is in fact the result of the program (X). Nevertheless, this design is appropriate and must be used in certain situations: a) If evaluation was not pre-planned as part of the total decision-making process, and b) If a comparison group is not needed or relevant to the situation.

   a.   <u>If evaluation was not pre-planned, this research design may be the only feasible one and it can provide some valuable information.</u> To be effective, the One-Shot Post-Test requires that the evaluator obtain as much detailed information as possible about the program and its outcomes. A "soft" narrative report which encompasses and

reflects all areas of the programming position and situation is more valuable than a "hard" statistical report which may be difficult to obtain under these conditions. A narrative report, while not definitive, can serve as a tool to structure thinking about the program. One key to writing effective narrative reports is to have available all resource material pertaining to the program; expenditure reports, correspondence, posters and press releases can all be used to obtain raw data which can be converted into indicators of the program's effects. Workshop attendance records can yield the obvious data; the number of people who attended and who were served by the workshop program. But attendance records can yield valuable additional information. If data pertaining to the location of workshops, the scheduled day and times, the instructors and other factors are all analyzed in greater detail, inferences about their effects on the program's outcome can be determined. If One-Shot Post-Tests are used frequently, always establish a formal evaluation file for each program where all relative documents and resource material can be kept for "soft" narrative evaluation reporting.

b. <u>A comparison group may not be necessary if determining the effects of the program on a specific population is the purpose of evaluation.</u> Audience surveys or membership surveys normally do not require comparison groups since the purpose is to document characteristics or profile of that specific group. In this case, One-Shot Post-Tests are appropriate. As noted in paragraph a. above, detailed investigation and reporting is vital for effective evaluation. If, when one evaluates the audience, distinct sub-groups of the population can be identified, (the elderly, professional persons, women, minority groups or whatever) these sub-groups within the audience or membership can be used as comparison groups against one another and will aid considerably in effective evaluation. Survey evaluation can be used effectively as a substitute for a "pre-test" by asking program participants to indicate what their expectations of the program had been prior to its implementation as well as asking questions about their assessment of what actually occurred.

2. <u>Time Series</u>:   O X_1 O X_2 O . . .

Another quasi-experiment, the Time Series model is noted above. There is no comparison group and, as stated earlier, the program itself, in quasi-experiments, serves as the method of comparison. In the Time Series, the program is observed over time as different program variations, noted by

the sub-numbers (X_1 and X_2), are systematically introduced and their effects observed. This design demands <u>systematic</u> planning, action and recording. Changes or variables should be stated, should be clear and simple, and observations should be taken and recorded following identical procedures each time.

Program evaluation using the Time Series design frequently makes use of existing statistics (numbers) as a data base (resource of information). For example, if a granting agency recently introduced new program guidelines for grant applicants, it may be interested in what identifiable changes this creates in the character or profile of grant applications. Use can be made of existing records as a data base. Grant proposals on file from the last five or six years can provide a statistical base of information. Examined closely, these data can be coded according to the level of funding requested, the type of organization, the location or destination of funds, creative or innovative characteristics of the programs or other criteria. These existing statistical data can be used to compare the effects of the new program guidelines. If significant changes in the above mentioned data occur after introduction of the new guidelines it can be inferred that the guidelines are responsible.

Since prior records became a <u>base of comparison</u>, any change in administrative record keeping procedures should not be taken lightly. Though changes in administrative procedures are sometimes necessary and beneficial, they do have their effect on this type of evaluation, since the base of comparison for <u>systematic observation</u> is then jeopardized--it is no longer consistent. If change in administrative record keeping is necessary the solution would be to thoroughly restructure the new method in a manner consistent with evaluation concerns, and then maintain that system as long as possible without change. Time Series design is particularly appropriate in

policy analysis or formative evaluation during the program development and planning process.

    3.   <u>Static Group</u>:   Program Group X O

                       Comparison Group   O

Static Group comparison is a design in which a <u>comparison group has been carefully selected</u> for observation (O) to aid in assessing what would otherwise be a One-Shot Post-Test. Similar to One-Shot Post-Test, it is evaluation initiated after the fact, after the program (X) has been implemented. In contrast to Time Series, used more often during "formative" evaluation, this design like One-Shot Post-Test, can be used more effectively for assessing the total impact or final outcome of a program.

An example: an artist-in-residence program has been implemented for one year in seven communities; assessment of its effect or impact on the population's attitudes toward the arts is needed. To observe the impact or outcome (O), a survey questionnaire is developed to determine attitudes towards the arts. (Other tools to measure impact could be used just as well; telephone surveys, personal interviews, or collection of unobtrusive indicators are among a few.) Using the Static Group design, the questionnaire is not only administered to the seven program or "test" communities, but to a number of communities specifically selected to "match" the characteristics of the "test" communities. In other words, comparison groups are selected that have demographic, educational, artistic, behavioral, or as many characteristics that "match," as closely as possible, the test community's profile. After one has analyzed the results obtained from questionnaires administered to both "test" and comparison groups, if there is a significant difference in attitudes toward the arts, there is some basis to infer that the artist-in-residence program had an effect. Like the One-Shot Post-Test, this

design is weak, but must be used under certain circumstances and can yield some valuable information. If circumstances are similar to those stated earlier in the One-Shot Post-Test section (p. 30), the Static Group design is one of the better designs to use for impact assessment in "summative" program evaluation. However, due to the expense and for reasons stated in III. Program Evaluation Process (subsection B., p. 14), outside research designers should generally be consulted to design and implement this type of evaluation.

C. Base Line Survey Research Design

Experimental "Control" Group   (O) X O

Comparison Group   (O)   O

Looking at the model, baseline surveys are developed to function as the observation (O), or pre-test, prior to implementation of the program X. This baseline survey, in the total program planning process, is generally undertaken as a needs assessment; what type program is needed? If the survey is designed with concerns for evaluation, observations O after the program can then be compared against the baseline survey results. If a comparison group had been selected and pre-tested using the baseline survey, even though they are excluded from the program, the additional support data is valuable when evaluating the program's outcome. (An example of a baseline survey is included as Appendix C.)

To illustrate: Assume the artist-in-residence program is the outcome of a needs assessment survey. In other words, a survey questionnaire was designed and administered to a multi-county area in an effort to determine the needs of the area. Based on that survey or pre-test, the local arts council planned and implemented the artist-in-residence program. If the information requested on this baseline survey is designed with the evaluation of the

34

program's effects in mind, it provides a base to compare the actual effects of the program on the selected communities and provides a base of comparison for those excluded as well. This again is a reminder of the importance of pre-planning, of thinking carefully about evaluation and the information it requires during the total planning process, not after.

To Summarize:

BOX II: Outcome Evaluation of the Program Evaluation Process.

1. The essence of outcome evaluation is determining which factors of a program caused which effects by controlling the influencing factors; control requires comparison.

2. A program is a combination of factors that need to be identified and defined.

3. To evaluate or measure program outcomes effectively, identify many different indicators or ways of assessing the results. This gives credibility to evaluation results.

4. The One-Shot Post-Test can be effective in certain situations but requires collecting and reporting as much detailed contextual information as possible.

5. The Time Series Design is most effective for formative evaluation and depends heavily on sound and consistent record keeping and administrative procedures.

6. Static Group Design is best used for impact assessment or summative evaluation and depends on the selection of comparison groups.

7. Evaluation and planning are part of the same process; think about evaluation before implementing the program.

VI. Evaluation and Common Sense

Good management and decision making requires a working knowledge of both formal and informal evaluation principles. The common sense of daily decision making and planning has always required the use of informal evaluation procedures. The purpose of this paper has been to define, in detail, evaluation principles in the decision making process without losing the

practical, common sense of evaluation and decision making. Following a set of guidelines is important to a degree, but this must be balanced by flexibility; understanding this balance is the key to good evaluation. Technically, certain approaches are more effective than others, but, if purposes and goals have been stated, the technical problems can be overcome. Although intensive summative program evaluation often causes tremendous impact and attention, the most effective evaluations are not single events but rather long-term, continuous, low level assessments. Evaluations conducted in this manner are not thought of as single entities or activities, but as an integral part of the total planning and implementation process. Evaluation principles should be integrated into all steps of all program development--during goal identification, needs assessment, target setting, record keeping, program monitoring and continual program refinement. The key point is that it must be systematic. Make sure that in each step of the planning process, evaluation is firmly grounded so that it provides a proper foundation for its application and for its roots in the next step. Each element of evaluation must be appropriate to the purpose of the total program plan. Again, follow guidelines but be flexible and experiment with the different tools to measure and assess impact. (Don't always think of surveys, for example.) Be innovative in constructing comparison groups; keep common sense channels open for effective evaluation methods. In the decision making process, determine what is useful information and what is not; design a valid evaluation report that serves a useful purpose as well as a decorative one. In the final analysis, be a critical thinker!

Appendix A

BIBLIOGRAPHY

The seven books listed below are useful tools in planning and implementing program evaluation research. Though technical in part, they are understandable. None is directed specifically to arts programming, yet each is applicable and provides additional detailed information pertaining to various areas of evaluation research if further data is required or desired. The majority are available in paperback.

1. Donald T. Campbell and Julian Stanley, Experimental and Quasi-Experimental Designs for Research, (Chicago: Rand McNally, 1966). A classic text on experimental design. Examines numerous alternative experimental approaches and discusses threats to their validity.

2. Andre L. Delbecq, Andrew H. Van de Ven, and David H. Gustafson, Group Techniques for Program Planning. (Glenview: Scott, Foresman, 1975). Thorough and understandable manual on techniques for group goal setting. Nominal group and Delphi techniques are detailed and examined; applicable, practical and useful text.

3. Kenneth M. Dolbeare (ed.), Public Policy Evaluation, (Beverly Hills: Sage, 1975). Contains evaluation theory and case studies focusing on large scale federal programs. James Coleman provides a useful overview on applied research principles.

4. Peter H. Rossi and Walter Williams, Evaluating Social Programs, (New York: Seminar Press, 1972). Focuses on large scale federal programs; theoretical articles by Rossi and Glennan are particularly useful.

5. Edith Stokey and Richard Zeckhauser, A Primer for Policy Analysis, (New York: W.W. Norton, 1978). Popular, intense and highly sophisticated text directed toward professionals in research evaluation and policy analysis. Contains excellent, well organized bibliography.

6. E. J. Webb, D. T. Campbell, R. D. Schwartz, and L. B. Sechrest, Unobtrusive Measures - Non-Reactive Research in the Social Sciences, (Chicago: Rand McNally, 1966). Enjoyable, well written and understandable text focusing on unusual observational measurement techniques. Try it on your friends!

7. Carol H. Weiss, Evaluation Research - Methods of Assessing Program Effectiveness, (Englewood Cliffs: Prentice-Hall, 1972). Directed toward graduate students in public policy; a current overview of evaluation methods. Contains good current bibliography on evaluation.

Appendix B

The following procedures have been designed as guidelines for structuring program evaluation. They are presented sequentially but can be altered depending on specific situations.

1. Determine the purpose of evaluation.

    a) Define the audience; define its political context.

    b) Determine if the outcome will be used; can and will the program be changed depending on the evaluation outcome?

    c) Is the purpose formative: to find ways to develop and improve a program? Is the purpose summative: to find out if the program is effective and should be continued?

2. Construct a program monitoring plan.

    a) Articulate objectives in concrete measurable terms: the what, when, how and how often of the program objectives.

    b) Develop a systematic plan for these activities; determine the necessary factors required to meet scheduling of their occurence.

    c) Establish an audit system: quarterly or bi-monthly, the frequency depending on the specific situation. Are audit forms needed? Analyze results: why were scheduled dates missed, prematurely achieved or overdue?

    d) Summarize audit results in a summary report. State reasons for problems in logistics and suggest preventative methods.

3. State expected or anticipated outcomes.

    a) Articulate goals of the program in broad conceptual terms.

    b) Set concrete outcomes for each stated goal. What changes in the state of real things (individual or organizational) are expected to occur as the result of the program?

4. Define indicators for each outcome.

    a) Assess validity of indicators. Has the indicator changed significantly enough as the result of the program to insure that it is in fact a true measure of program outcomes?

    b) Assess reliability of indicators. Do indicators consistently occur as a result of the program or are they a result of something other than the program?

c) Determine a number or quantity of indicators to measure each single outcome. If a single outcome or effect can be measured by a number of different indicators, statements as to program outcomes can be articulated with a greater degree of confidence.

d) List the different components or stimuli of the program. Can the effects of each be distinctively recognized and assessed?

5. Determine existing resources of information (data) and determine what information (or data) is needed.

a) Gather all existing program records: budget, expenditures, reports, attendance records, correspondence, appointment calendar, etc.

b) Determine the indicators that can be recorded or measured unobtrusively and establish record keeping procedures for each.

c) Determine what information must be obtained obtrusively through interviews or survey questionnaires.

6. Design survey or new data gathering techniques

a) If the program has been completed, determine if a comparison group or situation is available. Design the survey or interview to include questions to measure prior expectations of the program participants and comparison groups. (A simulated pre-test)

b) Determine baseline information in order to accurately measure the changes in state from a point or indicator prior to program implementation to some point or indicator at completion of the program.

c) Determine innovative or creative methods which function as credible comparison groups or situations to accurately measure outcome or progress.

7. Establish a data collection process.

a) Keep separate evaluations files for documentation, memos, records, etc.

b) Schedule regular monitoring of programs and updating of evaluation files.

c) Be consistent, systematic and determined.

Appendix C

FOR THE PURPOSES OF THIS STUDY, THE ARTS REFERS TO ANY CREATIVE ACTIVITY WHICH ONE
MIGHT OBSERVE OR PARTICIPATE IN. EXAMPLES INCLUDE MUSIC, DANCE, THEATRE, AND THE
ARTS AND CRAFTS, BUT ARE NOT LIMITED TO THESE.

1. How important are the arts in community life? Very Important:_____ Moderately
    Important:_____ Not Very Important:_____ Completely Unimportant:_____

2. How important do you think the arts should be in community life? Very Important:_____
    Moderately Important:_____ Not Very Important:_____ Completely Unimportant:_____

3. When you were in school, did you participate in dramatic performances, concerts,
    exhibits, or other arts activities? Yes:_____ No:_____
    How frequently did you do so? Regularly:_____ Occasionally:_____ Seldom:_____
                                 Don't Recall:_____

4. When you were in school, did you attend dramatic performances, concerts, or
    exhibits? Yes:_____ No:_____
    How frequently did you do so? Regularly:_____ Occasionally:_____ Seldom:_____
                                 Don't Recall:_____

5. Do you have school-age children living at home? Yes:_____ No:_____

6. How much are the arts stressed in the public schools of your community? A Lot:_____
    A Moderate Amount:_____ Not Very Much:_____ Not At All:_____ Don't Know:_____

7. How much do you think the arts should be stressed in the public schools? A Lot:_____
    A Moderate Amount:_____ Not Very Much:_____ Not At All:_____

8. Are you satisfied with the ways you can spend your leisure time in your community?
    Yes:_____ No:_____
    If no, what kinds of activities are missing?

9. How do you feel about the amount of arts activities (excluding TV) that is available
    to you in your community? Too Much:_____ Just Right:_____ Too Little:_____

10. How do you feel about the amount of arts activities (excluding TV) that is available
    to your children in your community? Too Much:_____ Just Right:_____ Too Little:_____
                                        Does Not Apply:_____

11. What kinds of arts activities have you participated in?

12. What kinds of arts activities seem to be most readily available to you (and your
    children)?

13. From your point of view, what kinds of arts activites seem to be most needed in
    your community?

14. Please rank the following sources of arts activity according to how important you
    think they are in your community. Put a 1 beside the source you think most important,
    a 2 beside the source you think second most important, etc. Leave blank the sources
    that make no contribution.
    _____Schools
    _____Area Colleges
    _____University of Illinois Cooperative Extension Service
    _____Local Clubs and Organizations
    _____Area Churches
    _____Professional Touring Groups
    _____Libraries
    _____Other (Please Specify:_____)

    Now put an "X" beside the source that you think should be most important.

15. Have there been live drama performances in your community in the last year? Yes:_____
                                                                               No:_____
    Have you gone to see any? Yes:_____ No:_____
    If yes, how frequently? About once:_____ About 3 - 4 times:_____
                            About once a month or more:_____
    How do you feel about the amount of live drama available to you? Too Much:_____
    Just Right:_____ Too Little:_____

16. Have there been live concerts in your community in the last year? Yes:_____ No:_____
    Have you gone to see any? Yes:_____ No:_____
    If yes, how frequently? About once:_____ About 3 - 4 times:_____
                            About once a month or more:_____

How do you feel about the amount of live concerts available to you? Too Much:_____ Just Right:_____ Too Little:_____

17. Have there been any arts or arts and crafts exhibits in your community in the last year?
Yes:_____ No:_____
Have you gone to see any? Yes:_____ No:_____
If yes, how frequently? About once:_____ About 3 - 4 times:_____
About once a month or more:_____
How do you feel about the amount of arts or arts and crafts exhibits available to you? Too Much:_____ Just Right:_____ Too Little:_____

18. Have you taken an art or crafts class in the last year? Yes:_____ No:_____

19. Do you think there are sufficient opportunities for arts instruction in your community?
Yes:_____ No:_____
If no, what is missing?

20. Please rank the following factors according to how likely they are to prevent you from attending an arts event. Put a 1 beside the factor you think is most important, a 2 beside the factor you think is second most important, etc. Leave blank those that are not important factors to you.
_____Event too far away
_____Not enough information about the event
_____Not interested in the arts
_____Not enough time
_____Poor quality of the event
_____Not enough interest in the specific art form
_____Too expensive
_____Other (Please Specify:_____)

21. Suppose that a professional theatre company visited your community to perform at reasonable prices:
Would you attend performances? Yes:_____ No:_____ Uncertain:_____
Would you encourage your children to attend? Yes:_____ No:_____ Uncertain:_____
Would you help with publicity? Yes:_____ No:_____ Uncertain:_____
How much would you pay for such a performance?

22. Suppose that, instead of an out-of-town professional company, live performances were offered regularly with people from your community doing all of the acting, directing, technical work, and sponsorship, again at reasonable prices:
Would you attend performances? Yes:_____ No:_____ Uncertain:_____
Would you encourage your children to attend? Yes:_____ No:_____ Uncertain:_____
Would you participate? Yes:_____ No:_____ Uncertain:_____
Would you help with publicity? Yes:_____ No:_____ Uncertain:_____
How much would you pay for such a performance?

23. How important is it to attract tourists to your community? Very Important:_____
Moderately Important:_____ Not Very Important:_____ Completely Unimportant:_____

24. How important is it to attract new business to your community? Very Important:_____
Moderately Important:_____ Not Very Important:_____ Completely Unimportant:_____

25. If you think more tourists or business should be attracted, how effective do you think good local arts programs could be in doing so? Very Effective:_____
Moderately Effective:_____ Not Very Effective:_____ Completely Ineffective:_____

26. What are your favorite hobbies and pastimes?

27. What is the most enjoyable event of any kind that you have attended in your community?

28. What is the most enjoyable arts-related event that you have attended anywhere within the last year?

Zip Code:_____ Male:_____ Female:_____

Marital Status:_____ Occupation:_____

How long have you lived in your community? _____

Age: Under 18:_____ 18-25:_____ 26-35:_____ 36-45:_____ 46-65:_____ Over 65:_____

Please check all of the following schools that you have attended:
Grade School:_____ High School:_____ Vocational School:_____
College:_____ Graduate School:_____ Art School:_____