A FRAMEWORK MODEL OF EVALUATION

Charles W. N. Thompson

This research was supported in part by a sub-grant from T.R.U.S.T., Inc., (then the Council On Population And Environment, Inc.) of Chicago, which was part of a grant from the Law Enforcement Assistance Administration, grant 75DF-99-0040

For presentation at the National Conference on Criminal Justice Evaluation, Washington, D.C., February 1977. (Panel 32 - Evaluation Paradigms: Other Approaches)

Northwestern University

Revised, 2/7/77
(pp. 1, 5, 9, and 10)

Abstract

There are two related purposes to which this paper is directed. First, to provide a framework for identifying and examining the relation between a) evaluation and related programmatic or administrative activities, e.g., planning and budgeting, program management, and implementation; b) evaluation and other forms of inquiry, e.g., research, systems analysis, monitoring, audit, measurement, and control; and c) evaluation and its variations, e.g., descriptive, retrospective, formative, and summative. Second, to provide a framework for identifying and examining a specific proposed or completed evaluation in order to a) understand its characteristics, what it is, what it claims, etc.; b) discover its problems, potential, and limitations; c) evaluate it in comparison to its proposed objectives or purpose; and d) compare it with similar evaluations.

The model uses an information or decision based framework in which all variables of interest are characterized as events in space and time; these events are "selected" for examination through identification of the observer (participant, actor, decision-point) and specification of his relationship to the event(s) in terms of the concepts of confidence and utility. As a primitive, or definitional model, it provides not only stability and a direct one-to-one relationship with real world phenomena but also flexibility in application to a wide variety of current problems in evaluation. The model itself is presented in verbal, mathematical, and graphical forms to facilitate its use.

The framework model is based on prior research and was developed, in part, during the course of carrying out a specific evaluation of an LEAA sponsored project.

1. Introduction

The term "evaluation," even when limited to its application to the "evaluation of social programs," e.g., criminal justice programs, provides an umbrella which covers a wide range of activities for a wide range of purposes [2, pp. 16-21]. In some cases, the activity identified as an "evaluation" is a simple, and relatively straightforward description of what happened with respect to a single intervention and its direct "effect"; at the other extreme, the activity intended includes an incredibly (perhaps, impossibly) complex and comprehensive comparison of "all" of the events occurring in or related to a significant major intervention, against some (nearly) equally complex and profound set of "standards." Similarly, the purpose may be simply to find out whether a particular effect occurred "after" the intervention was introduced; or the purpose may be to assess "all" of the subsequent effects of a major intervention against an extensive set of standards (measures, goals, objectives, criteria, etc.) [10, p. 31].

In those cases where all of the parties concerned (or involved) agree on the activity (and purpose) which is intended (and/or accomplished), it would seem clear that no immediate problem remains. For many activities and purposes there is an abundance of checklists, paradigms, and the like, available, ranging from modest and informal accounts to relatively rigorous and sophisticated formats [1; 5; 6; 7; 8, pp. 5-14]. It is, however, where there is not an "agreement" that these paradigms appear to be limited, and it is proposed that the extent and nature of these areas of disagreement are sufficiently non-trivial to suggest the usefulness of examining potential alternative solutions.

The approach to be presented here is to provide a framework within

which can be described and located any particular evaluation of interest in terms of the activity contemplated and its purpose. The framework provides a stable set of referents and explicit transforms to allow for consideration of related (or alternative) activities and purposes. The presentation which follows is in three parts: the framework, some applications to characteristic or common problems; and some summary comments.

2. Framework or Model

2.1. Defining Evaluation

In a dictionary or common sense, to evaluate means "to set down or evaluate the value of (something)" or "to examine or judge concerning the worth, quality, significance, amount, degree or condition ..." [cf. 4, p. 2]. This would seem to require at least three conditions: first, some "thing" to evaluate (for our purposes, let us call it a "program"); second, some measure or standard of worth, quality, value, etc., against which to compare it; and third, some process (or activity) for carrying out the comparisons (and/or establishing the result of the comparison) [cf. 1, p. 43; 10, pp. 28-29]. There may be some other useful (or even necessary) conditions, and these will be introduced later. Let PR represent the program to be evaluated, e.g., a criminal justice intervention and its intended (or hypothesized or resulting) effect; and let X_{pr} represent the set of one or more characteristics of interest of PR. Let X_s represent the set of one or more desired standards or measures of worth, value, etc., e.g., descriptive characteristics, level of results desired, efficiency, effectiveness, etc. Let X_i——X_j represent the relationship between the ith X and the jth X. We are now ready to define evaluation as follows:

(1) $$\text{Evaluation} = PR + X_s + (X_{pr}——X_s).$$

2.2 The Basic Program

There are many kinds of activities which may be described as programs (or projects) [cf.14, p. 24], but the present interest is in those which involve, at least, some planned or intended intervention in the real world, e.g., a change in patrol procedures, and the resulting (intended, expected, hypothesized) consequence, e.g., an increase in arrests, or reduction in crime, or improved "morale" of the police (or citizens) [e.g., 3]. Let X_{iv} represent the intervention, and let X_{dv} represent the consequence. As a matter of convenience we shall use a separate notation, the colon(:), for the relationship between a specified program (and, later, other specified activities or persons) and some other symbol or expression; the expression (a:b) is read "from a's point of view" or "a's b" or "b with respect to a." We now have, for some program PR, the following:

(2) PR: X_{iv} = independent variable(s), or intervention(s)

(3) PR: X_{dv} = dependent variable(s) or consequence(s) (intended, expected, hypothesized, resulting).

Even at this simplistic level, this two-variable model may not be adequate to describe a program PR.

First, only the most simple program (if that) would be characterized by two variables, and a single level of analysis. While, clearly, other (more detailed) levels of analysis could be accomplished by iterative two-variable steps, it is proposed to identify the set of "other" variables included in, or between, X_{iv} and X_{dv}, with the notation X_{pr}, and where necessary, the notation will specify whether X_{iv} and/or X_{dv} are included within X_{pr}. We now have, in addition, the following:

(4) PR: X_{pr} = variables included within the Basic Program.

Second, there may be other variables (particularly in a real world situation) which are present and interact with X_{iv} and/or X_{dv} in ways which make it difficult or impossible to do much more than to establish the existence (if that) of X_{iv} and X_{dv}. These "unwanted" variables, particularly those which (may) interact with X_{dv}, may make it difficult to establish the relationship between X_{iv} and X_{dv}, i.e., raise plausible rival hypotheses (for which the term "parameter" may be used); so, let X_{par} represent these "unwanted" variables. We now have:

(5) PR: X_{par} = parameter(s), some "unwanted" variables we (may) have to account for.

We now have several "different" possible descriptions of a specific program, with a nominal description, as follows:

(6) PR = $X_{iv} + X_{dv}$, defined in terms of (planned and/or achieved) "existence" of intervention and consequence

(7) PR = $X_{iv} + X_{dv} + X_{pr}$, defined in terms of (planned and/or achieved) "existence" of intervention and consequence, plus some further (more detailed) level of analysis

(8) PR = $X_{iv} + X_{dv} + (X_{iv} — X_{dv}) + X_{par}$, defined in terms of (planned and/or achieved) "existence" of intervention and consequence, plus the relation of the intervention to the consequence and the parameters.

It should be noted that there may be more extended descriptions of programs which require additional considerations, and these will be discussed in 2.5 below.

2.3 Measures of Value

We shall, for our present purposes, distinguish only two kinds of standards or measures of value, as follows:

(9) X_{gs} = a generic standard (or measure), for which there is no apparent or assumed preference, e.g., a foot rule, or weighing scale

(10) X_{ps} = a purpose or value oriented standard (or measure) for which it may appear that there is a preference in direction, etc., e.g., reading speed, number of crimes committed, dollars earned (or spent)

(11) $X_s = X_{gs}$ and/or X_{ps}.

It is not proposed, at this point, to examine either the conditions under which X_s appears, or the several forms X_s may take; further, it should be noted that the distinction above is made for convenience in presentation and is not intended to suggest more.

2.4 The Process

The process of evaluating is proposed to be nothing more or less than comparing one or more characteristics of the program, i.e., X_{pr}, with one or more appropriate standards or measures, i.e., X_s, and identifying the result of that comparison.

2.5 Extended Description of a Program

As was noted in 2.2, it may be useful to consider a more extended description of the concept of a program. To the extent a program is imbedded in the real world, there is, potentially, a nearly unlimited set of possible relationships which might be included with respect to any particular concept of a program. For our purposes, we will limit ourselves to a set of what appear to be common and useful relationships to consider. Using the

notation X_{pr} to include all of the terms in the Basic Program, i.e., X_{iv}, X_{dv}, X_{pr}, X_{par}, we shall now add the following items:

(12) $PR = X_{pr} + X_{ap}$, which adds consideration of the relationship with the set of one or more a priori states

(13) $PR = X_{pr} + X_{rc}$, which adds consideration of the relationship with the set of one or more resources and constraints

(14) $PR = X_{pr} + X_{so}$, which adds consideration of the relationship with the set of one or more spin-offs (or other outcomes)

(15) $PR = X_{pr} + X_{se}$, which adds consideration of the relationship with the set of one or more subsequent effects.

The several relationships are presented in Figure 1. The Basic Program includes those variables (or events) within the first two enclosures, with the minimum set including only X_{iv} and X_{dv} (the innermost enclosure) and the maximum set including, in addition, X_{par} (the second enclosure). The Extended Program includes, in addition, X_{ap}, X_{rc}, X_{so}, and X_{se} (the outermost enclosure).

2.6 A Brief on the Evaluator

In the discussion, to this point, the framework has employed three constructions, as follows:

(16) X an event, or variable, which may be defined as "any conceptually held change in state," or, for convenience, "any conceptually held state" (the distinction being in the level of analysis) [12]

(17) X_i——X_j the relationship between X_i and X_j

(18) PR a collective term for a specific set of X.

No specific accounting has been made of "who" has chosen and/or defined these terms, and, at this point it may be assumed that the writer (or other originator) of the statement(s) is the one whose conceptual holdings are relevant. If the reader (or others) shares (agrees with, understands, matches, etc.) the conceptual holding, there would appear to be no immediate problem. It is, however, the introduction of (possible) differences in the conceptual holdings (including differences held, or internal inconsistencies, by one individual) which presents many, if not all, of the problems with which this framework is intended to deal.

It is when a specific individual, let DP represent him, is concerned with a specific evaluation of a specific program (in the real world) that some further consideration must be given to the several relations which now appear. For our purposes, we shall identify these relationships in terms of the potential problems and the assumed solutions involved in achieving an acceptable and/or useful correspondence between the conceptual holding and its real world counterpart, as follows:

(19) With respect to the program PR, there must be indicators or other evidence of the state of X_{pr} available to DP through observation, interviews, questionnaires, records, or the like

(20) With respect to the standards X_s, there must be available to DP the dimensional state for X_{gs} and, in addition, the preference state for X_{ps}

(21) With respect to the process for comparing X_{pr} with the appropriate X_s, there must be available suitable analytical processes, at least for other than simple

comparisons, and both require, as a minimum, dimensional equivalency between X_{pr} (and X_{ap}, X_{rc}, X_{so}, and X_{se}) and X_s.

Each of the above assumptions presents, in some cases, profound problems in both theory and method, and it is proposed (but not presented here) that this framework can be extended to include these assumptions. For our purposes, we shall note those cases where these assumptions may be specifically critical.

There is one more, and essentially separate, relationship between the DP and the real world, and that is "time." To the extent there is a consensus, it appears that the term "evaluation" is limited to those comparisons which occur normally after the event X_{pr} to be evaluated [6, p. 8], and other terms, such as systems analysis, are used for the evaluation of plans for future events, and the like. This distinction, while perhaps useful, is not required by the proposed framework; the indicators for many events, and particularly X_s, may be indirect, included removed in time, and the test is really in the meeting of the above assumptions (19) and (20).

3. Some Applications

3.1 Defining the term "evaluation" and its variations

Perhaps the simplest or minimum content covered by the term "evaluation," in the context of the evaluation of social programs, is that of identifying, or establishing the existence of, or describing the intervention and the resultant (or expected or hypothesized, or intended) effect of the intervention. This is sometimes called a "descriptive" [but cf.10, pp. 76, 84] evaluation, and, in our terms, consists of measuring (or comparing) the real world event against some descriptive standard or scale. Let EV represent the evaluation, and we now have

$$\text{(22)} \qquad \text{Descriptive EV} = X_{iv} + X_{dv} + X_{gs} + (X_{iv} - X_{gs}) + (X_{dv} - X_{gs}).$$

It should be noted that, at a different level of analysis, a more extended, or otherwise different, set of comparisons can be made, in general between X_{pr} and X_{gs}. For example, "effort" [10, p. 61] may require only X_{iv}—X_{gs}. And, if the results of (22) are then compared with some preference scale, X_{ps}, describing what was expected, e.g., a plan, we have a "process" evaluation [2, pp. 18-19].

A second type of evaluation may occur, particularly in programs of extended duration, prior to the time of X_{dv}, and, where the results of the evaluation are (intended to be) used as inputs to making further changes in PR, the term "formative" [9, pp. 43, 51] evaluation is used. We now have

(23) Formative EV = $X_{pr} + X_{gs} + (X_{pr}—X_{gs})$ at $t < X_{dv}$.

A further term, "summative" [9, p. 43] evaluation, is used to identify one or more of several different types of evaluations, which may include (22), which occur at or after X_{dv}. The first of these is an "evaluation" to determine the relationship between X_{iv} and X_{dv}, i.e., did the intervention "cause" the effect observed? For this purpose it is necessary, to some extent, to account for plausible alternative explanations for the observed effect, i.e., parameters X_{par}. We now have

(24) Summative_1 EV = $X_{iv} + X_{dv} + X_{par} + X_{gs} + (X_{iv}—X_{dv}) + (X_{par}—X_{dv})$ at $t \geq X_{dv}$.

In a strict sense, the two relationships imply a previous comparison of the X_{pr} and X_{par} with the respective X_{gs}.

A second sense of "summative" is an evaluation of the effect or effectiveness of the program (or intervention) and this may mean several things. The simplest case of "effectiveness" [6, pp. 12-13] would be to compare X_{dv} against some preference standard X_{ps}, as follows:

(25) Effectiveness_1 EV = $X_{dv} + X_{gs} + X_{ps} + (X_{dv}—X_{ps})$ at $t \geq X_{dv}$.

A perhaps more common expression would be to combine the requirements of (24) and (25) in an "impact evaluation" [2, pp. 19-20], or a "performance evaluation" [10, pp. 62-63], as follows:

(26) $\left.\begin{matrix}\text{Summative}_2 \\ \text{Effectiveness}_2\end{matrix}\right]$ $EV = X_{iv} + X_{dv} + X_{par} + X_{gs} + X_{ps} + (X_{iv} - X_{dv}) + (X_{par} - X_{dv}) + (X_{dv} - X_{ps})$ at $t \geq X_{dv}$.

A further combination, (22), (24) and (25), has been referred to as a "comprehensive evaluation" [2, pp. 20-21].

The next kind of evaluation is that with respect to "efficiency" which introduces several possibilities. The simplest case in where there is a dimensional similarity between X_{iv} and X_{dv}, e.g., both can be measured in dollars. We then have,

(27) $\text{Efficiency}_1\ EV = X_{iv} + X_{dv} + X_{gs} + \left(\frac{X_{dv}}{X_{iv}}\right)$

Where X_{iv} and X_{dv} are not dimensionally similar, or where efficiency is with respect to some alternative X_{iv} and/or X_{dv} [10, pp. 64-66] or some preferred standard, X_{ps}, a different expression for efficiency is indicated. For convenience we shall assume that X_{ps} can also be used to express the alternative X_{iv} and/or X_{dv}. We now have

(28) $\text{Efficiency}_2\ EV = X_{iv} + X_{dv} + X_{gs} + X_{ps} + [(X_{iv} + X_{dv}) - X_{ps}]$.

All of the above are related to the Basic Program described in 2.2. Consideration of the Extended Description of 2.5 adds some additional possibilities in evaluation.

The concept of resources and constraints introduces consideration of the relationship of X_{rc} to X_{iv}. In simplest of terms a "resource" may be described as possible or acceptable changes in X_{rc} and a "constraint" as a limit on possible or acceptable changes. Similarly, the concept of spin-off introduces consideration of the relationship of X_{so} to X_{iv} and/or X_{dv}, and these relationships may also be both possible and/or acceptable changes

and limits on possible or acceptable changes. Finally, similar extensions are introduced by the concepts of X_{ap} and X_{se}. We shall not here develop the several extensions to the set of evaluations which may occur, but will note that the form will, generally, involve additional X_{gs} and X_{ps} and various combined or aggregated comparisons.

3.2 Evaluating an "evaluation"

There are several different types of "evaluation" of an evaluation, comparable to those described above. Let EV_1 represent the evaluation to be evaluated, and EV_2 represent the evaluation of the evaluation. We can now treat EV_1 as a program, and substitute EV_1 for PR in the several descriptions above. For example, if we wish to "understand" what was done, i.e., describe the EV_1, we can either repeat the process of EV_1, examining the process for internal consistency, or we can carry out an evaluation, treating EV_1 as a PR [1, pp. 101-103].

3.3 Relation to Other Terms

The relationship of evaluation to the program provides a framework for establishing further relationships to activities associated with the program, e.g., planning and budgeting (which is concerned primarily with the time relationships of X_{iv} and X_{dv}, and with X_{rc}), program management (which is concerned primarily with X_{iv}), and implementation (which is concerned primarily with X_{dv}, but sometimes also with X_{iv} and X_{iv}——X_{dv}).

Other terms used to describe inquiries may be related to this framework [11;13]. For example, exploratory research (which is primarily in the form of identifying X_{pr} and comparing it to X_{gs}), hypothesis testing (which is primarily concerned with X_{iv}——X_{dv}), systems analysis (which is, arguably, the same process as evaluation but less constrained in both time and in the set of X_{pr}), monitoring (which would appear to be in the form of X_{pr}——X_{gs} and/or X_{pr}—— X_{ps} on an on-going basis [6, p. 12-14 p. 27]), audit (which would

appear to be in the form of X_{pr}——X_{gs} and X_{pr}——X_{ps} after the fact), measurement (in the form of X_{pr}——X_s), and control (in the form of X_{pr}——X_{ps} at t_1 and the further relationship of X_{pr}——X_{ps} with X_{iv} at t_2).

4. Comments

This is a necessarily brief presentation of an underlying framework for evaluation. It is proposed, but not presented, that the framework can be extended to examine, on a similar basis, the several assumptions, particularly in 2.6, the considerations introduced where several different individuals are involved, and the whole question of establishing appropriate X_{ps}, i.e., the question of value or utility.

It is here offered that the relatively simple set of constructs, essentially a set of variables X and their relationships (——), provides a stable, complete, and traceable framework for examining relatively complex sets of real world events.

REFERENCES

1. Adams, Stuart, Evaluative Research in Corrections: A Practical Guide, National Institute of Law Enforcement and Criminal Justice, Washington, D.C., March 1975.

2. Bernstein, Ilene N., and Freeman, Howard E., Academic and Entrepreneurial Research, Russell Sage Foundation, New York, 1975.

3. Browne, Jane M., "An Evaluative Case Study of the TRUST Project (An Interactive Media Citizens Participation Program)," unpublished Ph.D. dissertation, Northwestern University, Evanston, Illinois, June 1976.

4. Caro, Francis G., Ed., Readings in Evaluation Research, Russell Sage Foundation, New York, 1971.

5. Glaser, Daniel, Routinizing Evaluation: Getting Feedback on Effectiveness of Crime and Delinquency Programs, National Institute of Mental Health, Rockville, Md., 1973.

6. Hatry, Harry P., Winnie, Richard E., and Fisk, Donald M., Practical Program Evaluation for State and Local Government Officials, The Urban Institute, Washington, D.C., 1973.

7. Maltz, Michael D., Evaluation of Crime Control Programs, Law Enforcement Assistance Administration, Washington, D.C., 1972.

8. National Criminal Justice Reference Service, Criminal Justice Evaluation: An Annotated Bibliography, National Institute of Law Enforcement and Criminal Justice, Washington, D.C., June 1975.

9. Scriven, Michael, "The Methodology of Evaluation," in Tyler, Ralph W., Gagne, Robert M., and Scriven, Michael, Perspectives of Curriculum Evaluation, Rand McNally & Company, 1967, Chicago, Illinois, pp. 39-83.

10. Suchman, Edward A., Evaluative Research, Russell Sage Foundation, New York, 1967.

11. Thompson, Charles W. N., "Administrative Experiments: The Experience of Fifty-Eight Engineers and Engineering Managers," IEEE Transactions on Engineering Management, Vol. EM-21, No. 2, May 1974, pp. 42-50.

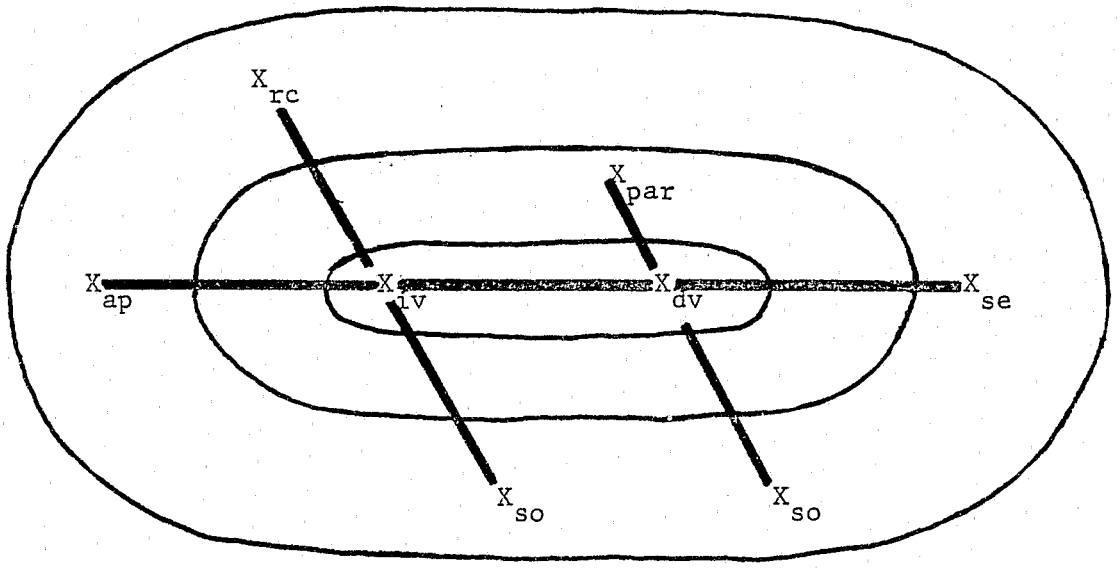
12. Thompson, Charles W. N., "Model of Human Performance Reliability in Health Care Systems," Annals of Assurance Sciences, Vol. 7, No. 2, 1974, pp. 335-339.

13. Thompson, Charles W. N., and Rath, Gustave J., "The Administrative Experiment: A Special Case of Field Testing or Evaluation," Human Factors, Vol. 16, No. 3, June 1974, pp. 238-252.

14. Wholey, Joseph S., Scanlon, John W., Duffy, Hugh G., Fukumoto, James S., and Vogt, Leona M., Federal Evaluation Policy: Analyzing the Effects of Public Programs, The Urban Institute, Washington, D.C., 1971.

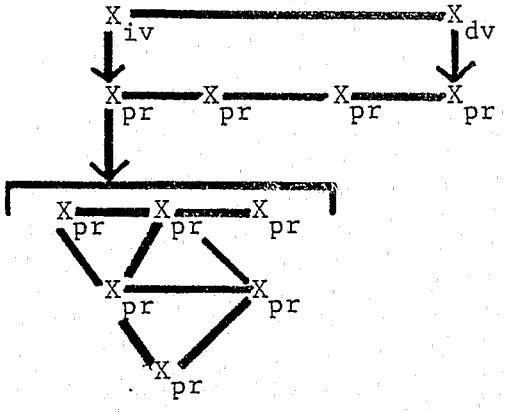
Figure 1

Basic and Extended Program
Variables (or Events)



Note: Basic includes first two enclosures; extended includes all three.

Note: Other levels of analysis include:



END