Principles and Experiences with Evaluation

(Preliminary Summary, May 1979)
CONTENTS

1.1 The history of evaluation within the Federal Administration (unwritten)

1.2 Evaluation in HUD (unwritten)

2. Function and role of evaluation

3. Types of evaluation

3.1 Performance evaluation

3.2 Impact evaluation

3.3 Program evaluation

4. Steps in evaluation

4.1 Identification of the information user

4.2 Focusing the evaluation questions

4.3 Identification of program goals

4.4 Choosing the design and methods and data collection

4.5 Interpretation of the findings and formulation of recommendations

5. Conclusions

Appendix

Framework of the intended work
This paper is a preliminary summary of my work done as a Junior Fellow at the Center for Metropolitan Planning and Research, The Johns Hopkins University. It is based mostly on literature about evaluation of educational and other social programs, and emphasizes the underlying principles based on practical experiences in evaluation over the last few years rather than the theoretical aspects. As there is little literature about evaluation in urban and regional planning, the author interviewed different people within HUD. These findings, and a review of evaluation work being done there, are not yet included in this paper. Together with additional interviews in city planning offices and regional offices of HUD, this work will form part of the final paper to be finished in late 1979.

Finally, I want to express my appreciation to Professor J. Fisher and his staff at the Metro Center for their help within the maze of Planning and the many interesting contacts they provided me.
2. **Function and role of evaluation**

In a general sense, evaluation aims to provide basic information in order to reduce uncertainties during the decision process. This definition of evaluation includes two different understandings of evaluation: the more scholarly conception, which emphasizes general valid results as contributions to basic research; and the direct conception which focuses on utilization, delivering results under very special conditions to improve a program in implementation.

The main functions of evaluation are:

-- to determine to what extent program activities are achieving the desired objectives;

-- to determine the reasons for success or failure of a program;

-- to examine the efficiency and adequacy of a program's activities and organization; and

-- to develop guidelines and set standards of performance.

In a broader sense, evaluation is understood to:

-- provide information on how a program can be transferred to other areas under different conditions, and make suggestions about necessary modifications;

-- enlarge scientific knowledge by testing specific hypotheses under controlled conditions (experimental or quasi-experimental situations);

-- indicate hypotheses to be tested in the future;

-- inform the public about successful programs; and

-- develop a critical attitude among administrators of programs and set up the information flows necessary to better coordinate programs.
3. **Types of evaluation**

The use of 'evaluation' in American literature is rather confusing. Evaluation is used in certain cases for nearly every kind of report about activities from highly subjective judgments, on the one hand, to scientific research on the other hand. At the same time, large areas of policy analysis cannot be separated from evaluation, considering the methods and subjects.

So a lot of the dissents and discussions are based on a lack of distinction between different kinds of evaluation based on different methodical approaches and underlying questions. Assuming that evaluation is basically a comparison between activities (e.g. planned - real) or outcomes (e.g. intended - real), it is therefore useful to distinguish between the following three types of evaluation:

- **Program Evaluation** asks whether a program has met its objectives by comparison of intended with real consequences, and tries therefore to determine the extent to which a given program is achieving its desired results. Program evaluation includes, in certain cases, relative effectiveness evaluation, measuring the comparative effectiveness of alternative strategies. It is mainly output-oriented and thus gives little information about why a certain program has been successful or not. For this information, studies of cause and effect relationships or the process of implementation have to be done; these studies are the basis of the two following kinds of evaluation.
- **Impact evaluation** attempts to define and explain the real consequences of program activities; the latter include also unintended effects, which make it necessary in certain cases to study general conditions (external variables). The question which underlies impact evaluation is very similar to that of scientific research in social and natural sciences, which tries to verify certain hypotheses, i.e. cause-effect relations; in the case of evaluation, the cause is a program activity. There are a lot of assumptions about activity-result relationships made during a planning process, which can be proven to be wrong afterward. Another reason for the failure of a program can be a lack of implementation, which may be evaluated in a third kind of evaluation.

- **Performance evaluation** compares the planned activities with those which really took place. This type of analysis may give useful information about the success or failure of a program.

The three kinds of evaluation are strongly dependent: program evaluation is very comprehensive, giving some kind of balance. The reasons for the success or failure of a program may be inappropriate assumptions about cause-effect relations, which are studied in impact evaluation, or inadequate performance of a planned program, analyzed in performance evaluation.

Following are some more details about the three different kinds of evaluation, which are made from a pragmatic point of view rather than a theoretic one.
3.1 **Performance evaluation**

The questions addressed by performance evaluation are: Has the program been implemented as planned? If not, what are the reasons? Performance evaluation consists therefore in a comparison of planned with real program activities. Successful implementation is characterized by a process of mutual adaptation due to incomplete knowledge about external conditions. Especially within public administration programs, implementation often changes as it drifts down the bureaucratic ladder.

There are two different kinds of performance evaluation: one focuses on the quantity and quality of activities that take place, and refers mainly to staff time, activity and commitment, and material resources. That kind of evaluation, also called monitoring, is relatively simple but nevertheless necessary in cases of lack of communication between the planning or funding agency and the implementing body (e.g. federally funded programs); it is also important where a general program must be adapted to special local conditions. A less output-oriented kind of performance evaluation focuses on the internal dynamics and actual operations of a program, and searches therefore for explanations of the success, failures, and changes in a program. This type of evaluation asks why certain things are happening, what is their context, and how people perceive the program. This information may be very important in interpreting the success or failure of a program. In addition, in order to make utilization-oriented recommendations, it is necessary to consider these aspects in detail.
3.2 Impact evaluation

The main question addressed by impact evaluation is: Does the implemented program lead to the desired outcomes? The major concern is therefore with the relationship between implemented program activities and their effects. These cause-effect relations, together with assumptions about the existence of special conditions, form the basis of any program and its underlying theory of action linking together means and ends. In the case of failure of a program, they have to be tested.

The basic question of impact evaluation is very similar to the question of scientific research. But in comparison to natural science, with its classical situation of experiment, social and spatial planning has to deal with a lot of uncontrollable factors and has to be useful for the decision maker rather than objective and generalizable. Although scientific research always represents a compromise between scientific requirements and administrative needs and resources, aspects like relevance, time, and resources are far more meaningful in planning. Basic research and impact analysis, as far as it is utilization-oriented, exclude each other to a certain degree.

As basic research aims to discover knowledge under the most generalizable conditions, impact evaluation should test the application of knowledge to influence special variables in a highly specific situation. For the latter, the needs of the information user and decision maker are very important. His or her lack of information are far more important than general scholarly interest. The nonuse of a lot of evaluation studies for the immediate decision process proves how far the opinion as to what should be evaluated differ between evaluator and information user. Therefore, cooperation between the two is necessary from the first stage (definition
of the problem). Also, in many cases, intermediate results may invalidate certain questions and lead to a redefinition of the questions evaluated.

In particular, impact evaluation is used to:

-- indicate and measure the outcomes of a program as a whole. For that kind of statement, many assumptions have to be made; usually, it is quite difficult to use the results to improve the program. One reason for this kind of impact evaluation may be the lack of explicit goals to be evaluated.

-- point out possible effects of certain activities. Such an attempt helps to identify and fill gaps in the theory of action underlying the planning process. There are problems with this, because one activity usually has several effects, and one effect may be caused by different activities. In addition, there are in every situation very special conditions ruling, which are in many cases unknown and whose contribution are difficult to determine.

-- make recommendations for the improvement of a program; in most cases, this depends also on the investigations within a performance evaluation.

-- make recommendations for future research or systematic variations in order to provide additional knowledge.

3.3 Program evaluation

The main question in impact evaluation is: To what extent has a program met its underlying objectives? By a comparison of the intended with the real outcomes, it gives a kind of balance, without considering the reasons for success or failure in detail. This general evaluation looks for an overall judgment about the effectiveness of a program and reflects, therefore, the situation of a black box. In many cases, this kind of evaluation is asked in connection with continuing or stopping a
program. By giving some possible outcomes of a program, it may also be the initial point for a cost-benefit analysis.

4. Steps in evaluation

The following notions try to summarize some of the experiences done by evaluators in practice; less emphasis is put on theoretical background, general concepts, and definitions. Evaluation is characterized by a wide variety of research designs, ranging from an intuitive and unsystematic 'Is everyone happy' approach to a complex experimental approach with the use of sophisticated statistical methods. The result is a lack of comparability and cumulativeness of evaluation findings.

Two different ways of proceeding through the whole process of an evaluation are given below. One proposal is from HEW*, the other is from Patton**. They reflect two different understandings of what evaluation should be and under what conditions it should be done. The former is mainly based on the concept of experimental situation and therefore emphasizes standardization of measurement, whereas the latter is based more on subjective judgment and is utilization-oriented, and therefore puts the weight on identification of the information user and the focusing of the evaluation question.


In particular, the two proposals studied recommend the following steps for performing an evaluation:

<table>
<thead>
<tr>
<th>HEW</th>
<th>PATTON</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Identification of the goals to be evaluated.</td>
<td>• Identification and organization of relevant decisionmakers and information users.</td>
</tr>
<tr>
<td>• Analysis of the problems with which the activity must cope.</td>
<td>• Focusing the evaluation question, choosing the type of evaluation.</td>
</tr>
<tr>
<td>• Description and standardization of the activity.</td>
<td>• Identification and clarification of the goals.</td>
</tr>
<tr>
<td>• Measurement of the degree of change that takes place.</td>
<td>• Choosing between alternative paradigms of evaluation measurement and design.</td>
</tr>
<tr>
<td>• Determination of whether the observed change is due to the activity or to some other cause.</td>
<td>• Data collection.</td>
</tr>
<tr>
<td>• Some indications of the durability of the effects.</td>
<td>• Interpretation of the data.</td>
</tr>
</tbody>
</table>

The following section concentrates on certain steps from a rather pragmatic point of view, pointing out some general problems and principles.
4.1 Identification of the information user

Evaluation which is not primarily scholarly in nature, but tries to improve a program directly depends strongly on a precise identification of the person who is in charge of the program. His sometimes very specific information needs have to be taken into consideration; in most cases, a general determination of the possible audience is not enough to ensure the use of the results. The identification of the problem or question of evaluation is very dependent on the specific information needs of the administration. Accordingly, many practical experiences demonstrate that the more precisely the future information user has been determined and the earlier he has been involved in the problem definition and the design of the evaluation, the more results of evaluation studies are considered in the general decision process. All program improvements have to be done by distinct people and are therefore to a large extent dependent on their understanding of the problem. As the results of evaluations and their interpretation require special skills, a whole learning process is supposed. This process has to begin with the definition of the problem and continues through the whole process of evaluation and ends with the interpretation of the data and the formulation of recommendations. One-shot results usually fail to initiate this kind of learning process and therefore lack an adequate influence on the decision process.

A continuous cooperation between evaluator and information user gives some guarantees that the evaluation is still of interest when the results come out. At the same time, it prevents the evaluation from reflecting only the views of the evaluator. Often the problem viewed and the methods are strongly determined by the evaluator; for the same problem,
an economist may propose a cost-benefit-analysis, the sociologist a survey, the anthropologist field methods and the psychologist a test.

4.2 **Focusing the evaluation question**

The question being evaluated has to be chosen in cooperation with the information user and has to be revised in many cases on the basis of intermediate findings. The evaluation question determines the type of evaluation to be done:

- the question about the fulfillment of the main program goals requires a program evaluation;
- the question about the reason for certain outcomes requires an impact evaluation;
- the question about an adequate implementation requires a performance evaluation.

Despite the possibility of changes in the evaluation question during the evaluation process, it is useful during conceptualization to try to interpret possible results. If it is not possible at this stage, it would not be possible in most of the cases when the results are present. Such a procedure, which can take the form of a pilot study, prevents the evaluator from collecting a large amount of irrelevant data which allow only few conclusions and recommendations and usually end up as descriptive studies. Conclusions about the improvement of a program are in such cases impossible, arbitrary, or very general, having only very little in common with the results. In such cases, the lack of utilization orientation may be compensated for by the argument of basic scientific knowledge, which is a strong and undefeatable argument. It leads, in certain areas, to many
evaluation studies with little or no influence on the improvement of existing programs. Therefore, the evaluation question has to be formulated as precisely as possible. Unskilled evaluators and enthusiastic proponents of evaluation, who believe that evaluation is so important that it should be applied to everything, tend to formulate the question too comprehensively. The same may happen if the evaluator and contractor cannot reach a consensus. As a consequence, the results may be very general or some things may end up getting more attention than others in a rather arbitrary way.

4.3 Identification of program goals

The identification of program goals is, in most cases, not as easy as it may seem. In many cases, they need additional interpretation to be used for program evaluation. Because their interpretation is always very subjective, cooperation with the planning agency is necessary. If the goals, or their respective weights, have changed meanwhile, important information concerning the interpretation of goals may be available from the agency, which is responsible for implementation.

The goals clarification by these people has to be done in a very sensible way by the evaluator, in order not to be dominated by a prefixed idea of the latter. Otherwise, the goals clarification becomes rather like a twenty-questions game played at parties: the evaluator collects statements and tells the people how close they have come to the goal he has in mind. That tactic ends mostly in a general feeling of frustration among the participants and usually means the end of any kind of cooperation, depriving the evaluation of its basis. Another possible reaction of
the participants with a similar result consists in making uncertain former statements about valid goals, ending up in an endless goals clarification. As the goals and their weight may change after a certain time or even as a result of the implementation of a program having unforeseen disadvantages, it makes a big difference in results, if the original program goals or the actual goals are evaluated. The latter may be of importance in some cases where implementation has changed drastically.

Whereas goals are crucial for program evaluation, they may be of less importance for impact and performance evaluation, which may be determined to a greater extent by information needs of a decision maker.

4.4 Choosing the design and methods and data collection

There is a strong relation between results, methods, and general design of the evaluation. Design and data collection choices are far from being neutral, objective, or rational. Each design and every method has its own strengths and weaknesses. These have to be considered, combined and evaluated against specific information needs. This is also the reason why a hypothetical attempt at interpretation of possible results should be done at the beginning as part of the problem definition. It may be necessary again before the extended collection of data to assure the greatest possible use out of them. The same is valid in cases where standards have to be set: they should be set before data collection. Discussions about interpretation of data and required standards must be set in an atmosphere that is not charged with defensiveness, rationalization and justification.
The design of an evaluation is most important in the case of an impact evaluation. Here, there are a number of different fundamental possibilities of experimental design, as follows:

--- One-shot case study based on an evaluated observation after the exposure to the program. Although the results are highly subjective, this kind of evaluation is very common.

--- One group, pre-test, post-test design is also very common and is based on a comparison of a state before and after. It may lead to some assumptions about possible changes, although it may be difficult to measure the influence of external factors and special conditions.

--- Static group comparison takes an influenced control group into consideration and assumes the same external conditions for both groups.

--- Pre-test, post-test, control group design is again based on a before-after comparison. The introduction of a control group makes it possible to measure the influence of external factors and special conditions. This is the classic experimental design and in most cases of social and spatial planning it is difficult to obtain. The same applies to an even greater extent to the Solomon Four-group design and the systematic variation of a specific program component.

--- Longitudinal study design evaluates at different points in time, checking the progress toward the objectives. One problem with this design is the carry-over effect from previous tests, which may change the object being evaluated.

4.5 Interpretation of the findings and formulation of recommendations

There are two main opinions as to what the final result of an evaluation should be, again reflecting the more science-oriented or utilization-focused points of view:

-- the goal of evaluation is mainly to deliver data which are the basis for judgments and final conclusions by the decision maker.

-- the evaluator's responsibility is to draw conclusions from data and make judgments about the evaluation results.

A combination of these two views is proposed by PATTON (op. cit. p. 258/9): "The evaluator's job includes making judgments and recommendations, but decisionmakers and information users are first given an opportunity to study the data analysis without the intrusion of the evaluator's interpretations." This means that the data are easy to understand and that methods, especially their strengths and weaknesses, are known to the information user and decision maker.

5. Conclusions

The following principles can be considered as the quintessence of my preliminary review of the evaluation literature:

-- There will always be a conflict between design and goals of basic scientific research, which attempts to obtain the most generalizable information, and utilization-focused evaluation, which tends to influence a program's implementation in a very direct and specific way. There will always be a compromise to combine utility with objectivity.
To ensure commitment and interest of the information user, he and the administrators of the program have to be involved in the evaluation from the beginning and at different stages. This and observing the time limits are the two main factors responsible for an effective evaluation.

The methodological quality tends to be overweighted in its meaning for the practitioner by the evaluator. Especially the utilization of evaluation findings does not depend on broad and scientific studies. Cohen and Weiss even concluded that the use of increasingly rigorous designs and more sophisticated analytical techniques "...led to more studies that disagree, to more qualified conclusions, more arguments and more arcane reports and unintelligible results." Therefore, in most of the cases, simple and specific utilization-oriented evaluations have to be given preference.
Subject: Evaluation of programs with emphasis on urban and regional planning.

Evaluation can refer to different parts of a planning process and its outcome. For the purpose of this research it means:

-- the examination of the extent to which a program and its implementation succeed in reaching their underlying goals.

In this context it does not mean:

-- the (ex ante) evaluation of alternative programs and their assumed outcomes prior to the final selection and implementation.

-- the comparison of a program's outcome with actual goals.

-- the evaluation (ex post/ex ante) of a program's outcome on the basis of general criteria (e.g., impact analysis).

-- the analysis of the change of general indicators as a result of implemented planning and/or general development.

The following questions are considered important:

-- which problems in literature and practice are considered crucial for such an evaluation?

-- which solutions (pragmatic principles, methods, techniques, etc.) are proposed to deal with these problems?

-- how useful are these principles and methods in practice (urban and regional planning)?
I will develop the research in the following steps:

1) Definition of evaluation;
   Evaluation in the process of planning and implementation;
   History of evaluation;
   Types of evaluation;
   Steps in conducting an evaluation.

Sources (in order of their assumed relevance):
--literature
--periodicals

2) Description of the main methodological problems in evaluating a program's outcome with emphasis on urban and regional planning, if possible.

Sources:
--literature
--periodicals
--evaluation reports
--people (evaluators)

3) Outline of approaches, principles, options and methods for step 2.

Sources:
--literature
--periodicals
--people
--evaluation reports