EVALUATION: MANIFESTATIONS OF A NEW FIELD

EUGENIE WALSH FLAHERTY
Philadelphia Health Management Corporation

JONATHAN A. MORELL
Hahnemann Medical College and Hospital

ABSTRACT

An examination of three characteristics of evaluation reveals significant divisions of opinion, suggesting that it is premature to seek a defining conceptual framework for this still evolving field. Therefore an alternative approach toward understanding evaluation is followed: empirical manifestations of the history and current state of evaluation are examined for evidence of growth and integration.

A review of the history of evaluation suggests four causes for its recent growth: new accountability requirements, greater interest among social scientists in social relevance, a scarcity of resources for the traditional social sciences, and an expansion of methods useful for research in applied settings.

Three empirical signs of the new field are described: the concepts and strategies employed in evaluation efforts, the discipline of practicing evaluators, and dissemination of evaluation information. The conclusions suggest that there are both signs of cohesiveness and immaturity in the current state of evaluation.

INTRODUCTION

The term 'evaluation', used to describe a growing body of efforts sharing certain common characteristics, is increasing in use and seems to connote approximately similar meanings to most of its practitioners. However, the consensus breaks down when attempts are made to explicitly define the characteristics of evaluation, leading us to believe that it is premature to seek definitions in this still evolving field (see also Guba 1972). The purpose of this paper is to take another approach: instead of seeking the unique conceptual framework which defines the field of evaluation, we will look at the empirical growth of whatever it is that people call evaluation. This inductive approach can provide us with the pieces which may fit into a conceptual framework in the future. Meanwhile, we can examine the historical and descriptive manifestations for evidence of the growth and integration of evaluation (Roger Brown, 1965, defends a similar approach for the study of social psychology). The companion to this paper, "The Development of Evaluation as a Profession: Current Status and Some Predictions" (Morel1 and Flaherty 1978), uses the manifestations described in this paper to demonstrate that the growth of the field of evaluation fits into sociological models of professional development. These models assist us in understanding the process of development of the field, and provide predictions for evaluation's future.

We first review three basic characteristics of evaluation, and the evidence for the existence of significant schisms concerning each. Following this demonstration of the futility of attempts to define evaluation at this time, we begin the descriptive approach employed in this paper by tracing the historical pressures on the growth and recognition of evaluation as we currently understand it. The causes for the recent expansion of evaluation efforts are hypothesized and examined. The current state of evaluation is described in terms of three empirical manifestations which sociological theories suggest are among the signs of a new profession (Morell and Flaherty 1978).

THREE CHARACTERISTICS OF EVALUATION

Although we seem to share an implicit common frame of reference when we speak of evaluation, this consensus breaks down when attempts are made to explicitly define the characteristics of evaluation. Divisions of opinion

The authors wish to thank Donald Lund, Sherri Sheinfeld, and Jeffrey Travers for their constructive critiques of early versions of the paper, and Marcia Guttenag for her support and encouragement of the ideas developed here. Barbara McKim, Peter Gallagher and Charles Ings were of assistance in collection and preparation of material for the paper.

Request for reprints should be sent to Eugenie Walsh Flaherty, Philadelphia Health Management Corporation, 530 Walnut St., Philadelphia, Pennsylvania 19106.
exist on at least three characteristics: the role of the scientific method, the goal-attainment approach versus the systems model, and the use of evaluation information. Suchman (1967) represents one school of thought concerning the role of the scientific method, considering it an essential characteristic of evaluation. He defines evaluation research as "the utilization of scientific research methods and techniques for the purpose of making an evaluation... Evaluation research is first and foremost, research, and as such must adhere as closely as possible to currently accepted standards of research methodology". He specifically states that he does not view "the field of evaluation as having any methodology different from the scientific method", (p. 12).

Suchman’s high regard for the scientific method in evaluation is not unanimously shared. Weiss (1972a) describes three criticisms of the experimental model by Guba and Stufflebeam:

- It requires holding the program constant rather than facilitating its continual improvement.
- It is useful for making decisions only after a project has run full cycle and not during its planning and implementation.
- It tries to control too many conditions, making the program so aseptic that it is ungeneralizable to the real world.

There seems to be two 'camps' of evaluators concerning this issue, and discussions can become quite vehement about the appropriateness and the utility of the experimental approach.

The second characteristic of evaluation at issue is, in some respects, a consequence of the first. There are two schools of thought on the kind and the timing of information which evaluation should provide: the goal-attainment approach and formative evaluation. The former essentially defines the primary function of evaluation as the measurement of the achievement of a program’s predetermined goals. Schulberg and Baker (1969) call this approach the goal-attainment model, and they lucidly differentiate it from the second approach described below. Deming (1975) calls this the analytical approach, and Seriven (1972) summative evaluation. The goal-attainment approach in operation usually means that information is provided only after program termination, when the attainment of ultimate objectives has been measured. Implicit in this approach is the belief that all program goals can be specified before program implementation.

The second approach has been variously called the systems model (Schulberg and Baker, 1969), enumerative evaluation (Deming) and formative evaluation (Weiss); Rossi and Wright (1977) subdivide it into process evaluation and implementation. Each represents approaches with somewhat different emphases, but all share certain common features. Formative evaluation, as we will call it, is founded on the belief that the users of evaluation information are best served by (1) a comprehensive description of how the program actually functioned and (2) feedback about the attainment of program objectives, about unanticipated results of program efforts, and about the relationship between program functions and the attainment of objectives at continuous points during the program’s course, rather than after program termination. The proponents of this approach share the belief that the goals of a program cannot be fully specified before the program begins (Cronbach 1977). Formative evaluation guides ongoing management intervention and decision-making, internal to the program, rather than summarizing its external impact. Formative evaluation and the goal-attainment approach are by no means mutually exclusive; operating together they would provide the most comprehensive evaluation of all. However, evaluators usually operate under conditions of limited resources and are thus forced into a choice. The choice is not dictated solely by resource limitations or the intended uses of the information, but also reflects the evaluator’s concept of his role and of his relationship with program staff.

The third characteristic of evaluation about which evaluators differ in practice is the use of evaluation information. Explicit recognition of the role that evaluators can play in the utilization of the data produced is a recent phenomenon. Lip service has always been paid to the evaluator’s responsibility to make information useful, but recent recognition of how difficult this process is (Davis and Salasin, 1975, Weiss, 1972b) has forced attention to this aspect of evaluation. Edwards, Guttentag and Snapper (1975) consider information utilization an integral characteristic of evaluation:

"Evaluations ... exist (or perhaps only should exist) to facilitate intelligent decision-making ... If evaluation does not improve the basis for decisions about the program and its competitors, then it loses its distinctive character as evaluation research and becomes simply research." (p. 140).

Few would disagree with this definition in principle; however, disagreement might be more apparent when one examines specific evaluation efforts. How many evaluation efforts actually incorporate methods to increase information utilization into their design from the beginning? On the other hand, how many evaluation efforts exhibit concern with utilization only as the data collection and analysis draw to a close — or perhaps not even then? An empirical examination of evaluation efforts would undoubtedly divide the efforts into two groups in terms of their adherence to information utilization as an essential characteristic.

**HISTORICAL PRESSURES ON THE GROWTH AND RECOGNITION OF EVALUATION**

Evaluation has been around a long time. Early examples frequently mentioned are the studies by Kurt Lewin and his associates on the forces governing attitude and behavior change (1948). Newcomb’s study of the effects of an experimental undergraduate program on students is a classic (1943). Chapin (1947) describes several early evaluations of social welfare programs.

Although program evaluations were certainly con-
ducted earlier, there was a marked growth in both case studies of evaluations and in reviews of evaluations in the mid-1950's. Riecken's memorandum on program evaluation in 1953 suggests that he found evaluation a flourishing field at the time:

"There are a large number and variety of studies being carried on for the purpose of examining and suggesting change in programs of intentional social action, and ... the demand for such studies from social practitioners is increasing." (1972, p. 85).

For an excellent review of evaluation studies in mental health to 1954, see the report by the National Advisory Mental Health Council of the National Institute of Mental Health (1955). Herzog (1959) provides a very thorough exposition on the methodology of evaluation, from which we can still learn a great deal. Getting (1957) reviews the extensive evaluation literature for public health programs. Jahoda and Barnitz (1955), Klineberg (1955) and Paul (1956) each provide good brief descriptions of the nature of evaluation at that time.

This very cursory review of the early evaluation literature suggests that evaluation as a term and as descriptive of studies having certain characteristics in common has been in use at least since the early 1950's, and probably a good deal earlier. Why, then, the sudden recognition? Why have books and journal articles dealing with evaluation proliferated in the 1970's? Why is Suchman's book in 1967 sometimes considered an early classic, when clearly there was much written on evaluation before? Why have several graduate and undergraduate programs in evaluation developed in the last five years, and not before? Why do those who carry out evaluation call themselves evaluators now, when previously they called themselves researchers or social researchers? (Two of the earlier uses of the term evaluator are in Riecken 1972, and in Wright and Hyman's 1964 chapter, The Evaluators.)

We propose four causes of this sudden spurt in the growth of evaluation: greater requirements for accountability in publicly funded programs, increasing interest among social scientists in social relevance, an increasing scarcity of resources for the traditional social sciences, and an expansion in the social science methodologies appropriate for research in applied settings. Obviously none of these causes operated independently, and some probably played a more important role than others; our retrospective look does not permit isolation of the role of each cause.

The increased spending on social service programs in the early 1960's left as its heritage a new concern for accountability in the spending of these funds; this concern remains with us even as the funding for such programs slows down. The legislation for community mental health centers is a good example of this sustained demand for accountability. In the original legislation in 1963, the research and evaluation service of a CMHC was considered one of the service elements necessary for the provision of adequate mental health services, but was not designated an "essential element" (Federal Register, May 6, 1964). No specific portion of the funds were designated for the evaluation service. In 1970 new legislation for community mental health centers required centers to provide two percent of their annual operating expenses for continuing evaluation of the effectiveness of its programs...and...for a review of the quality of the services provided by the center" (Public Law 94-63, Sec. 206). This specific requirement that a constant percentage of center funds be allocated to evaluation suggests that for CMHC centers, and probably for other service programs in the future, accountability will be an integral component of the funding pattern in the future. The increased demand for accountability may put a strain on the existing technology of evaluation, and we can expect to see many new developments and a growing sophistication in these techniques.

The marked increase in social service programs of the 1960's also served to heighten the interest of social scientists in doing research in applied settings. The excitement generated by the concept of the Great Society and by the assumption that with enough money and with the right efforts we could decrease the social ills of this country, served to turn the eyes of the social scientist away from simulated experiments toward these exciting natural experiments. The burgeoning social programs of the Great Society were not solely responsible for the new interest in social research, however: if they were, we would have seen the genesis of evaluation as a field in the days of the New Deal. There have always been social programs to stimulate research interest. Only in the 1960's did we see social scientists as a group turn toward these new programs as exciting areas for research. This supports our thesis that the recent growth of evaluation cannot be explained in terms of any one cause, but only as a result of several causes operating together.

Concurrent with the increased interest of social scientists in applied research, financial backing for these scientists decreased: both academic jobs and the funding for traditional research projects have shown a downward trend in the 1970's. As the growth of college enrollment leveled off and the operating costs of colleges and universities soared, it became increasingly difficult for a recent graduate to obtain a job in academic settings, and even more difficult to achieve tenure and keep that job. The young social scientist who obtained a job discovered when he wrote a grant for research funds that the competition was fiercer than it used to be. Thus, the young and eager social scientist was without the academic job he or she was educated for and even if he had the job, was unable to obtain funding for his traditional research project. On looking around, he or she found that he might be able to do research, albeit of an applied nature, in the social programs which are now required to demonstrate accountability. On the surface the methods required seemed not so different from the experimental methodology in which he was trained. So the social scientist became an evaluator (and learned that in reality the field bears only a remote resemblance to his former discipline!).

The fourth cause proposed for the recent growth in evaluation is the development of new social science methods which can realistically be used in evaluation studies. We left our young social scientist in his or her new role as an evaluator with the dismaying discovery that those elegant research designs he learned in graduate school cannot be implemented in the evaluation of service programs: what does he do? Evaluation has been able to
grow in quantity and in sophistication because evaluators have new tools and new methods with which to work. Perhaps the best known examples are Campbell and Stanley’s quasi-experimental designs (Campbell and Stanley 1966) and interrupted time-series design (Campbell 1975). Isaac and Michael (1974) and Weiss (1972) offer comprehensive descriptions of research designs for evaluation. The increasing availability and flexibility of computers permit management of complex designs and large data bases. There has been a proliferation of new instruments and strategies for evaluation: the National Institute of Mental Health has recently produced an entire handbook on new techniques for the evaluation of community mental health centers and review of it makes one realize how many of those techniques have been developed very recently (Hagedorn et al 1976). However, the methods currently available are not adequate for the job, and we can expect to see further development of the social science methods, as well as the evolution of a body of tools designed specifically for this new field.

This confluence of four causes for the growth of evaluation provides a preliminary explanation for the visibility of evaluation as a field in the 1970’s, and not earlier. These four sources provided the motivation (by means of increased interest in applied research and decreased funds for academic social sciences) and the tools (through the development of social science methods). As we said earlier, these sources are surely not independent; for example, the new interest in social relevance may be a consequence of decreased funds for traditional efforts. We believe, however, that only because all four forces came into play concurrently did evaluation become a field of its own.

**EVALUATION AS A FIELD: EMPIRICAL MANIFESTATIONS**

There are several sociological models of the creation of a new discipline (Morell and Flaherty 1978); although the sociologists differ among themselves, all agree that there are specific ‘signs’ of a mature discipline. We will look at three of these signs with the intent of examining evidence for the growth of evaluation as a distinct discipline or field.

- The development of specialized concepts and strategies for use in evaluation.
- The discipline of practicing evaluators.
- Information dissemination in evaluation.

**The Development of Specialized Concepts and Strategies for Evaluation**

All edited books on evaluation published since 1968 and the NIMH handbook (Hagedorn et al 1976) were reviewed for the concepts and strategies being used in evaluation. The findings suggest that the concepts and strategies are borrowed from six different fields: economics, epidemiology, management science, psychotherapy research, and the social sciences of psychology and sociology.

Evaluation has appropriated from economics methods of assessing the costs and benefits of services provided. Because it is often difficult to put a market value on the outcome of most social service programs, cost-effectiveness analysis rather than cost-benefit has been seen as more appropriate for evaluation. In cost-effectiveness analysis, program outcome is not translated into a monetary value, but is linked directly to the cost of achieving any given level of that outcome. According to Levin (1975), cost-effectiveness analysis was developed by the military for their evaluations of weapons systems. Although the approach itself has actually been applied in relatively few evaluation studies, Carter and Newman (1976) have developed a client-oriented cost-outcome system which has been applied to several mental health programs.

From epidemiology evaluation has borrowed two tools: social area analysis and survey techniques. (Survey techniques also owe aspects of their development to other fields.) Struening (1975) defines social area analysis as: "a set of integrated procedures designed to study characteristics of people who live in defined geographical areas...the interest is in characteristics or attributes which meaningfully differentiate areas or their resident populations, and have important relationships with the health status of residents" (pp. 526-527)

Social area analysis seems most useful as a planning tool, for the assessment of needs and for the selection of location and format of programs to serve those needs. This tool has been used frequently for needs assessment (although in a relatively simplified manner); its popularity is partially due to the fact that it uses existing data and does not require the expensive procedure of collecting original data.

The survey method in epidemiology includes both the field survey and secondary analysis of existing data collected for purposes other than the present study (Susser 1975). Although field surveys have limited utility for evaluation, because of their expense, secondary analysis of existing data has great potential, although it is not without its problems (Webb 1966).

From management science, evaluation has borrowed the management information system and adapted it to its own needs for program monitoring. Although good case studies of management information systems in mental health are available (Taube 1969), for general guidelines on the entire process of designing and implementing a management information system there are relatively few sources. (See Hagedorn et al (1976) for a summary of the current “state of the art.”)

Quality assurance in physical and mental health care bears a strong similarity to the concept of quality control in management science (Way, Lund and Attkisson, in press.) Although the concept of quality control has probably had a minimal influence on quality assurance to date, we can expect its influence to grow as health care deliverers are required to participate in quality assurance procedures. A third contribution of management science may have been the management by objective approach, a possible precursor to the goal attainment approach (Marsh 1978).

Psychotherapy research is an influential precursor of
evaluation, and many psychotherapy studies could serve as the model for a goal-attainment approach to evaluation. The boundary between research and evaluation becomes particularly blurred here, because research in psychotherapy by its very nature takes place in genuine service programs rather than in the laboratory. One would have to look very closely at the objectives of the particular research project to make the distinction: if the objectives are the increase of knowledge about the patient, personality or therapist characteristics, it is a research project. On the other hand, if the objectives are to assess the effectiveness of the particular psychotherapy program, then it is evaluation. Using these criteria, many psychotherapy research projects would probably qualify as both evaluation and research.

Evaluators have learned much about the application of experimental research designs in applied settings from psychotherapy research. Gruenberg (1966) contains several case studies of early evaluations of mental health service effectiveness in which the necessity for deviation from experimental research design is illustrated. Psychotherapy research has also contributed a large array of instruments for the assessment of status and change in the patient population (see Waskow and Parloff 1975 for useful critique of many of these instruments).

Evaluation has appropriated heavily from psychology and sociology. Perhaps the most important contributions are the experimental and quasi-experimental research designs (Campbell and Stanley 1966; Cook and Campbell, 1976). From social psychology and sociology come interviewing techniques and empirical evidence on the sources of bias in interviewing. Both psychology and sociology contribute personality and social attitude tests; the latter have been especially useful for evaluation of training programs.

Evaluators are heavy employers of the psychometric methods of psychology and sociology, especially for test construction and scaling approaches. Psychometric theory has also contributed methods of assessing reliability and validity of measurement instruments. Thorndike and Hagen's *Measurement and Evaluation in Psychology and Education* (1961) is an invaluable source for these areas; Nunnally and Wilson (1975) and Nunnally and Durham (1975) consider the problems of measurement, test construction and scaling specifically in the context of evaluation.

The systems model described earlier is also borrowed from sociology. This model is especially appealing to those concerned about the lack of utilization of evaluation findings. *Carried out thoroughly, however, the systems model is much more demanding for the evaluator in terms of time and knowledge and much more expensive in terms of manpower* (Schulberg and Baker 1969). We will probably see more application of this approach in the future, although the specific steps involved in its implementation are still unclear (Rossi and Wright 1977).

The evidence to this point supports Struening's statement that:

"... the field of evaluation has not produced a well-developed applied science complete with a comprehensive set of constructs and their reliable and valid measures. ... Nor has it developed a systemati-

ically acquired body of knowledge acceptable to the scientific community and sufficiently general to apply in a variety of settings without question ..." (1975, pp. 519-520).

Is evaluation at this point no more than an amalgam of concepts and techniques from different fields? The answer at this point is perhaps a qualified "yes", qualified because there are a few concepts and techniques which seem to have been created specifically for the needs of evaluation.

The first is goal-attainment scaling (GAS). Kiresuk and Sherman's 1968 article is by now a classic, and goal-attainment scaling has been extended from its original application, the measurement of progress of individual patients in therapy, to comparison of program effectiveness and the achievement of program and administrative goals. From its initial use in mental health settings, GAS has been used in drug abuse programs, rehabilitation services, and programs for the retarded.

A second new concept has developed from the increasing concern with the utilization of evaluation findings: the A VICTORY technique. A VICTORY is an acronym for the seven factors which one should consider in planning for the utilization of evaluation results: Ability, Values, Idea, Circumstances, Timing, Obligation, Resistances and Yield. This technique has been applied in numerous evaluation projects and is reported to be successful in increasing utilization (Davis and Salasin 1975).

A third concept was developed to facilitate the role of evaluation in decision-making: the decision-theoretic approach by Edwards, Guttentag and Snapper (1976). (See also Guttentag 1976.) This approach provides a framework for the total process of decision-making; Edwards et al discuss the value of the framework for the dissemination and utilization of evaluation results.

A fourth concept that has become uniquely evaluation's is the distinction between process and outcome. Donabedian (1969) describes the relationship between process and outcome as an "unbroken chain of antecedent means followed by intermediate ends which are themselves the means to still further ends", (p. 189). Process variables are generally assumed to be those variables which make up the delivery of program service as it is actually carried out, not as intended to be carried out. Weiss (1972b) suggests that the process approach "directs attention to the essentials ... allows a test of the theoretical linkages and enables the evaluator to say useful things about the stage where things go awry and adjustment is needed", (p. 324). The distinction between process and outcome measurement has even been applied in goal-attainment scaling (Kiresuk and Lund 1975). This concept, although perhaps not new in evaluation (it is difficult to trace its origins) has become essential in the framework of an evaluation project.

Lastly, evaluation has taken as its own the concept of needs assessments as a category of techniques aimed at a particular problem. The assessment of type and extent of needs is an important aspect of program evaluation, and several new techniques have been developed for it. The key informant approach and the community forum are two new and practical impressionistic approaches for assessing a geographical area's needs (Hargreaves et al
1975. See also Warheit, Bell and Schwab 1974). Several of
the needs assessment techniques were borrowed from
other fields; social area analysis and surveys, for example,
come from epidemiology. The contribution of evaluation
is the conceptualization of these techniques as members of
a single category with the objective of assessing need.

The Discipline of Practicing Evaluators

Until recently, there were no undergraduate, graduate
or post-graduate training programs in evaluation.
Hagedorn et al. (1976) reports that as of 1976 there were
one undergraduate program, three masters programs, fif-
teen doctoral programs, and eight post-doctoral or ad-
vanced doctoral programs in evaluation or a closely re-
lated field.

Because these training programs in evaluation are rel-
atively new, it is safe to assume that most working
evaluators were trained in some field other than evalua-
tion. We have already seen that many of the concepts and
techniques used in evaluation came from the fields of
economics, epidemiology, management science,
psychotherapy research, education, psychology and
sociology. Is there a parallel variety in the backgrounds of
those who practice in this new field?

Three disparate sources were available to the authors
on the disciplines of working evaluators:

1) The academic disciplines of the directors of 236
evaluation research projects funded in the fiscal
year of 1970 by the federal Departments of
Health, Education and Welfare; Labor; Justice;
Housing and Urban Development; and Agriculture
as well as the Office of Economic Opportunity
and the National Science Foundation, (Bernstein
and Freeman 1975).

2) The academic disciplines of 54 staff members in
the state mental health agencies of 14 southern
states in early 1974, (Heighton and McPheeters
1976).

3) The academic disciplines of all authors in five
well-known edited books in evaluation; Caro
1971, Russi and Williams 1972, Schulberg, Sheld-
don and Baker 1969, Struening and Guttentag
1975, and Zusman and Wurster 1975.

Each of these sources has an inherent bias, and use of any
single one would certainly misrepresent the total picture.
If anything, these sources together provide an overrep-
resentation of individuals who are well-known and who
have received evaluation research funds; the working
evaluator who has not received federal funds for a specific
evaluation project in his name or who has not published
will be underrepresented. On the other hand, there is a
compelling argument in favor of obtaining just such a
biased view. The evaluators in our sample are likely to be
those with the greatest impact on the development of
evaluation as a field. We have already seen that there
have been a multitude of influences on the concepts and
techniques in the field; it is not unreasonable to think that
influential evaluators will introduce into evaluation the
techniques they have been trained in. Therefore, the find-
ings of this somewhat cursory survey may very well have
implications for the future course of evaluation.

Psychologists play a dominant role in each of the three
groups; (Table 1) sociologists play a lesser but still impor-
tant role. Psychiatrists are important in federally funded
projects and publications, but are completely missing
from the state mental health agencies. On the other hand,
statisticians make up seventeen percent of the state staff,
and are minimally if at all in the other two groups. Does
this apparent dichotomy between the state mental health
staff and the evaluators directing evaluation projects and
writing book chapters represent the beginnings of two
specializations in this new field? The small numbers in

<table>
<thead>
<tr>
<th>DISCIPLINES OF THREE GROUPS OF EVALUATORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directors of federally funded projects</td>
</tr>
<tr>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Psychologists</td>
</tr>
<tr>
<td>Sociologists</td>
</tr>
<tr>
<td>Medicine</td>
</tr>
<tr>
<td>Psychiatric Social Work</td>
</tr>
<tr>
<td>Public Health</td>
</tr>
<tr>
<td>Economics</td>
</tr>
<tr>
<td>Statisticians</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

† Not represented in categories
2 For state mental health evaluation staff, the economics category is made up of accountants.
this comparison prohibit much speculation, but the tasks which the two groups perform are certainly different and require different skills. It may be reasonable to predict the growth of two bodies of knowledge and two sets of techniques in evaluation, one for the staff of government agencies and a second for the directors of large scale evaluation projects (who, we suspect, are also the authors of the book chapters to some extent). Will one of these bodies of knowledge and techniques be useful to the evaluator who is not at all represented in our sample: the evaluator working in the evaluation unit of a community mental health center, drug or alcohol abuse program, rehabilitation program or any other service delivery program?

Is there a relationship between the concepts in evaluation borrowed from other fields and the academic disciplines of these evaluators? Psychology, sociology, psychotherapy research, epidemiology (as public health), and economics, five major contributing fields, are represented in the backgrounds of the evaluators who wrote book chapters, which would be a major source of influence on the concepts and techniques. Management science, however, is not. Management science, whose main contribution to evaluation has been the concept of 'quality control', and the "Management by objective" approach system, may be represented by the economists, who made up six percent of all authors.

There is a strong relationship between the academic backgrounds of prominent evaluators and the fields from which evaluation has so richly borrowed. What does this mean for the future? Can we predict that as more and more of the working evaluators are professionals with academic training in evaluation, there will be less borrowing and more creativity in the development of new techniques for evaluation?

Although the foregoing analysis makes it clear that today's evaluators were trained in a variety of established disciplines, the twenty-seven new evaluation training programs suggest that this trend will diminish greatly in the future. Some of these programs are contained within departments of other disciplines and some exist as unique programs in their own right. Although course content varies greatly, the programs share the objective of providing specialized training to people who will be experts in evaluation. It is a reasonable assumption that as these graduates begin to compete with their more traditionally trained colleagues for jobs in evaluation, the importance of specialized training will assert itself. Thus, we can project a shift in the make-up of evaluators from those who started their careers in traditional disciplines to those who began with a specific interest in evaluation, with career goals in the field, and with professional training specifically focused on evaluation issues. Unfortunately, we may in the process lose the unique multidisciplinary aspect of evaluation.

| TABLE 2 |

JOURNAL ARTICLES AND BOOKS CATEGORIZED AS EVALUATION  
1960 - 1975

<table>
<thead>
<tr>
<th>Year</th>
<th>Psychological Abstracts</th>
<th>Sociological Abstracts</th>
<th>Books in Print</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>4</td>
<td>/</td>
<td>2</td>
</tr>
<tr>
<td>1961</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1962</td>
<td>0</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>1963</td>
<td>4</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>1964</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>1965</td>
<td>26</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>1966</td>
<td>34</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>1967</td>
<td>109</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>1968</td>
<td>156</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>1969</td>
<td>142</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td>1970</td>
<td>199</td>
<td>55</td>
<td>7</td>
</tr>
<tr>
<td>1971</td>
<td>112</td>
<td>44</td>
<td>9</td>
</tr>
<tr>
<td>1972</td>
<td>104</td>
<td>37</td>
<td>10</td>
</tr>
<tr>
<td>1973</td>
<td>251</td>
<td>Cease</td>
<td>11</td>
</tr>
<tr>
<td>1974</td>
<td>407</td>
<td>separate</td>
<td>20</td>
</tr>
<tr>
<td>1975</td>
<td>520</td>
<td>listing in 1973</td>
<td>11</td>
</tr>
</tbody>
</table>
A sign of the increasing numbers of those who call themselves evaluators is the recent proliferation of professional associations specifically formed for evaluators. A recent survey discovered a number of local evaluation organizations around the country (Evaluation 1975). Two national organizations have been formed: The Evaluation Research Society and the Evaluation Network. Evaluation associations are a sign that these professionals, from different backgrounds, are engaged in work of a nature and with problems so similar that they feel they would profit from communication.

Information Dissemination in Evaluation

In order to assess the extent of dissemination of information on evaluation, three publication sources were examined from 1960 to 1975: Psychological Abstracts, Sociological Abstracts, and Books in Print. All articles or books under the heading of 'evaluation' or, in the case of books with 'evaluation' in the title, were included. The results show a marked increase in articles cited in Psychological Abstracts, a moderate increase in Sociological Abstracts, and a slight increase in books on evaluation (Table 2). The preparation time and the publication lag for books are so much greater than they are for journal articles that the difference in rate of increase is not surprising.

When did evaluation publications show their greatest growth? Articles in Psychological Abstracts seemed to show the sharpest increases in 1967, in 1973, and again in 1974, with another jump in 1975. Sociological Abstracts showed the greatest increase in 1970, and books in 1974. No pattern seems to exist, except for a steady increase in the 1960's and a sharp increase in the 1970's.

Another indication of the growth of dissemination of evaluation information is the appearance of journals and newsletters specifically dealing with evaluation. There are now three journals in evaluation: Evaluation (Minneapolis Medical Research Foundation), Evaluation Quarterly (Sage Publications), and Evaluation and Program Planning (Pergamon Press); the first began publication in 1972 and the last two in 1977. There are numerous newsletters concerned with evaluation: see the first two issues of the newsletter Evaluation News (Evaluation Network) for a comprehensive list of the other newsletters.

CONCLUSIONS

Having reviewed a varied assortment of aspects of evaluation, can we draw any conclusions about the present state of the art of evaluation? Certainly only tentative ones. First, there are several signs of cohesiveness: the development of a few concepts specifically for evaluation, the recent establishment of some training programs in evaluation, the increase in dissemination of information about evaluation, and the very recent proliferation of associations of evaluators. Second, there are some genuine signs of immaturity: the lack of agreement on basic characteristics of evaluation concepts and methodologies, the fact that most techniques used in evaluation are borrowed from other fields, and the diverse

academic backgrounds of most practicing evaluators.

The close examination suggests that evaluation is a heterogeneous mix of borrowed techniques, of a very diverse set of practitioners, and of real disagreement about what evaluation is. Whatever the reality, the external appearance of evaluation makes it look more and more like a unified field: why else evaluation associations, evaluation training programs, and evaluation books and articles? Does this external appearance of unity herald the genuine development of a field? Or will evaluation always be a heterogeneous mix, with evaluators bound together only by the applied nature of their research efforts?

REFERENCES


EDWARDS, W., GUTTENTAG, M., and SNAPPER, K. A decision-theoretic approach to evaluation research. In Handbook of


