

THE EVALUATION OF R&D PROGRAMS AND PERSONNEL

Review of the Literature

Thomas E. Clarke, M.B.A., M.Sc.

Stargate Consultants Limited

Nanaimo, B.C.

December, 1986; Revised 2000

PREFACE

Although this report contains some minor modifications since it was originally written in 1986, it reflects our state of knowledge at that time. In the author's opinion, little has changed in our understanding of the difficulties in effectively evaluating the impact of R&D work, or in judging the quality of work produced by an individual contributor. No new "magic bullets" have been developed since 1986 to improve our ability in this evaluation activity.

INTRODUCTION

In the past, the evaluation R&D programs received relatively little attention by authors or management researchers interested in "Research-on-Research". As a result few articles or reports dealt specifically with the topic. This was in sharp contrast to the R&D project evaluation and R&D personnel evaluation on which articles, books, etc. abound (Clarke, 1974; Clarke and Reavley, 1993, 2000). With few exceptions, the literature that did exist on evaluating R&D programs dealt with the evaluation of industrial R&D programs and usually in the context of measuring R&D productivity.

The lack of literature also reflected the relatively low level of interest governments' had in evaluating their R&D programs and expenditures to determine both their effectiveness and their contribution to national innovation goals and priorities. This has now changed in that companies and governments are seeking to justify the considerable investment they make in R&D. As a result many reports and articles on the topic have been published since 1986 (including some hilarious Dilbert cartoons), (e.g., Apt and Watkins, 1989; ARA Consulting, 1993, Freedman, 1995, Phelan, 2000; Sen and Shailendra, 1992 and Wolff, 1989). A major contribution to the literature is the recent book by Eliezer Geisler entitled, "The Metrics of Science and Technology" (Geisler, 2000).

Because there is considerable overlap between personnel and program evaluation, this literature review will deal with both areas. R&D program evaluation concerns will be examined from the point of view of government and private sector R&D programs.

In the R&D management literature many authors state that we do not have precise measures to determine the innovativeness or productivity of an R&D program or an individual scientific contributor. This should not lead a research manager to conclude that such evaluations should not take place for, as Dr. A. H. Rubenstein notes in his review of technology management, "technical managers must do a better job of evaluating their own activities, or this will be done for or to them" (Rubenstein, 1985). Concern over who will develop the R&D performance measures is echoed by Thomas Phelan (Phelan, 2000), "If we fail to develop acceptable performance measures ourselves, we should hardly be surprised if others develop them for us".

R&D PROGRAM EVALUATION METHODS

Areas of Evaluation

R&D PROGRAM = R&D MANAGEMENT + SCIENTIFIC/TECHNICAL ELEMENTS

An R&D program consists of two distinct elements: a managerial element which consists of technical managers, organizational structure, and organizational environment; and the technical element made up of scientists and engineers, technical equipment and facilities, and technical projects/problems to be solved. The quality of performance of an R&D program is, therefore, a function of the quality of the management and the scientific/technical resources available in an organization.

It should be possible to evaluate both the managerial and technical activities related to the immediate outputs and longer term impacts of the R&D program. Specific criteria can be developed to assess both the process and output/impact of the managerial and technical stages. Thus, we can evaluate the management process by determining:

- whether management procedures are appropriately specified;
- whether these procedures inhibit or support an environment which encourages creativity/productivity;
- whether managerial styles are appropriate to the R&D situation; and
- whether there are clear and effective communications channels in place to provide the information necessary for timely decision-making.

The results of the management process, as distinct from the outcome of the technical or research process, are also subject to evaluation. The processes of identifying objectives, planning, and project selection will result in the program having a portfolio of research projects that it is conducting either in-house or through contract research to outside organizations. It should be possible to evaluate the composition of this portfolio and/or to evaluate the quantity and quality of proposals being submitted to the program. It should also be possible to evaluate the overall working environment to determine whether it is supportive of high quality research and development, combats technological obsolescence, and provides suitable rewards and recognition to motivate creative/productive performance among the program personnel.

Similarly, it should be possible to evaluate whether the monitoring of projects and evaluation of project reports is appropriate, or whether it is detecting significant departures from project plans in a timely manner to avoid cost or time over-runs or to signal the need for a change in project and perhaps even program objectives

The success of management actions taken to facilitate the diffusion, transfer and use of results might be evaluated by determining what percentage of project results are reaching various target audiences, the number of requests being made to the program for information, the existence or otherwise of a formal technology transfer procedure, or the extent to which cross-fertilization is occurring among projects.

An organization's corporate objectives strongly influence the direction of its R&D program and its sub-objectives. There will not usually be a single R&D impact/objective but usually several being addressed at the same time. If nothing else, an organization concerned about its future capability, will have what can be called maintenance impacts/objectives that are designed to maintain and develop the organization's technical capability. While they may be considered part of the managerial process, their direct impact is on R&D capability. It is this capability which enables the R&D organization to take on new problems and solve them in an effective and timely manner. These maintenance objectives usually include:

- prevention of technological obsolescence;
- prevention of burnout in professional staff;
- development of a network of colleagues both within and outside the organization to facilitate the acquisition of technical information; and
- maintenance and development of a highly motivated professional staff.

The research/technical element of the R&D program can be also be evaluated in terms of process, output and impact. Research/technical process evaluations might consider the suitability of the methodology employed to specific research questions (i.e., use of the latest techniques or outdated procedures), the quantity and quality of scientific/technical personnel, the quantity and quality of technical facilities, the extent to which the project adheres to schedules, or the degree to which the project researchers involve potential users of the research early in the life of the project. A factor known to increase the probability for successful technology transfer.

Process evaluation is especially important for longer-term R&D projects for which outputs are not available during the project. It may also be useful for very complex, or for less structured projects to conduct mid-course evaluative reviews that may aid in restatement of objectives, reorganization, changes in project design or other adjustments.

Evaluation of the research outputs from the program might consider the quantity and quality of the findings. The criteria used in the evaluation of research results must take into account the mission of the R&D program. Basic research programs might be evaluated on their contribution of scientific knowledge while more mission-oriented research and development programs might be evaluated with respect to the relevance of the research findings to improving or enabling an organization to fulfill its objectives.

R&D output generally consists of new information or knowledge in the form of papers, reports, prototypes, or patents. These outputs lend themselves to evaluation by a wide range of reviewers such as fellow scientists/engineers, by "customers" and by employers. Methods and techniques for evaluating R&D outputs, which will be outlined in the next section, are more readily available than for process or impact evaluation. Output evaluation is expected by researchers and, within certain limitations, is probably more readily acceptable.

From the point of view of policy makers, politicians, planners, the scientific/technical community and the general public, it is the impact of the R&D program which is the most important matter (Salasin, Hattery, and Ramsay, 1980).

Nielson and Brazzel (1980), in a discussion of the evaluation of agricultural research, define impact evaluation as addressing the following questions:

1. Were the objectives of the research program accomplished?
2. What were the scientific/technological, economic, social and environmental impacts of the program?
3. Who benefits from the program, how and by how much?
4. Who is worse off because of the program, how and by how much?

5. What were the costs of the program in terms of investments and other opportunities foregone?

6. Who bears the cost?

7. How do benefits compare with the costs?

Thus research impact can be measured in terms of:

Utilization as determined by, for example, surveys of potential user groups, actual user groups or incorporation of findings in educational materials.

Effects as determined through, for example, retrospective studies such as TRACES or HINDSIGHT; direct measures of economic benefits, measures of influence on further research work (citation index measurements), or instrumental and conceptual uses of program findings in policy or program initiatives.

Influence of Program Characteristics

The general approach and specific indicators to be employed in evaluating a given program, and the interpretation of evaluation results, will depend on several characteristics of the program. These include the expected time frame for production and use of results, the maturity of the field, the types of products produced, potential economic/legal/political impacts of the program's work, and the types and loci of research performers.

The time frame in which it is reasonable to expect the program to produce results, and those results to be used, is an important characteristic to consider in determining how the program should be evaluated. Almost all projects or programs produce some results such as technical reports in the short-term. If, however, the mission of the program involves a long-term effort that includes multiple projects, the short-term output may not provide suitable evidence for comprehensive program evaluation. It may be necessary to wait years before determining whether a given line of inquiry pays off, either in terms of increasing knowledge or producing usable products/processes.

Similarly, the time frame for use of the program's results should be taken into account. Findings from basic research programs may not reach practical application within five to ten years after the research is completed. Research results designed for intergovernmental use may require that the program not only produce results, but that it also disseminate the results to potential users, train these users, assist in building the capacity of municipal or provincial governments to use the results, and wait for the use to be institutionalized at a local level before deciding on whether the research was successful. Different criteria and measures for evaluation will be appropriate for the short and the long term period.

Various areas of research differ in their maturity, and in the ways in which research findings are disseminated and applied. Research in a mature disciplinary area can usually be evaluated based on generally acceptable standards for work in that area. In new interdisciplinary areas, there will most likely be no acceptable standards. Mature fields have established publication patterns, making it more reasonable to compare programs in these fields based on publication practices (e.g. number of publications, citation patterns, prestige of journals in which publications appear) than in fields where publication patterns are less-defined, or where results are disseminated more through informal than formal communication channels. In rapidly changing technologies such as microelectronics, informal channels of communication are more likely to be used such as personal exchanges of information between members of a communications network since by the time a publication is in print the results are out-of-date.

Both the nature of the products produced by the research (outputs such as publications, technical reports, patents, hardware), and the suitability of these outputs for immediate use or commercial development, must be considered in defining an evaluation strategy. Programs whose projects result in prototype hardware devices or demonstration/information service programs may be evaluated with respect to whether the technology works or the information solved problems. Is it an improvement over existing methods? Programs whose projects are suitable for commercial development might be evaluated, in part, on whether the technology was successfully transferred, and commercialization occurred. Do project reports disseminated by the program contain the information needed by potential users. Appropriate criteria may range from the operational test and evaluation of hardware/software under field or actual operations (providing cost/effectiveness measures for comparison with existing hardware/software) to peer assessment of scientific output or user assessment in the case of problem solving information.

The structure of the program's user community is important to understanding the results of any evaluation based on criteria related to utilization. Thus, for example, the existence of an infrastructure to facilitate utilization (e.g., consortium of private companies, provincial government departments that provide the same service) is an important aspect to consider when evaluating a program.

The evaluation of a R&D program should also consider the extent to which activities of the laboratory are intended to develop and maintain a pool of researchers in a given area. Departmental responsibility for the health of science in a given area will vary.

To summarize, the evaluation of R&D programs can be based on the degree to which the R&D outputs have had an impact on the environment external to the laboratory. Research outputs can be measured in terms of the number of:

- papers published in refereed journals;
- patents filed/awarded;
- internal reports prepared;
- presentations made at national/specialized conferences or symposia;
- new techniques/processes developed;
- operational equipment developed;
- operational problems solved;
- new professionals trained; and
- papers published in trade journals or popular literature.

The impacts of an R&D program can be evaluated in terms of the degree to which they:

Basic Research

- advance knowledge in a specific field of science;
- raise important theoretical issues;
- resolve a recognized scientific controversy;
- advance knowledge in research techniques;
- result in a pool of highly qualified researchers; or
- develop an R&D capability in a university, government or other organization.

Mission-Oriented Research and Development

- advance knowledge in a specific field of application;
- enhance the utility/usefulness of products, software or hardware developed;
- enhance the utility/usefulness of processes developed;
- result in technical/scientific information that can be applied to solve an operational problem in a cost effective manner;
- advance knowledge of research techniques;
- result in technical information that assists in decision-making, or new policy or regulatory formulation;
- produce information that contributes to the education of the general public to scientific issues;

- result in a pool of highly qualified, applied researchers;
- improve an organization's technological innovation capability; or
- improve the organization's ability to successfully transfer technology to an adopting organization.

EVALUATION OF THE R&D MANAGEMENT PROCESS

The attitudes or opinions of scientists and engineers about their employing organization will affect their level of creativity and productivity. A positive attitude towards the organization would signify good organizational health or a good working environment which has been shown by numerous studies to be fundamental to encouraging creativity and productivity in the R&D setting. Walton (1961) suggest that one technique for gauging organizational health in a government laboratory is to administer a simple questionnaire consisting of incomplete sentences, cartoon incomplete sentences, and semantic scales. The R&D staff are asked to complete the sentences or rate an item in the organizational climate such as quality of management. The rating is to include not only "the ways things really are", but also "the way things ought to be". The difference between these two ratings is a measure of dissatisfaction. Examples of the incomplete sentences include:

Promotion in this organization is ...
 My manager is
 Morale in this laboratory is ...
 What I like least/best about working here is ...
 Recognition of scientific accomplishment by senior management is ...
 This laboratory's reputation is

Walton considers that the use of this Organizational Self-Inventory provides an accurate picture of the working environment as perceived by the scientists and engineers.

Other methods to determine the state of the work environment include more formal questionnaires, workshops to discuss environmental factors in the work place and personal interviews conducted by outside consultants. Discussion of the situations portrayed in Dilbert cartoons would also be a method of determining the attitudes of bench level technical personnel to their managers.

Frank Andrews (1979), in a UNESCO study of scientific productivity, proposes examination of "organizational, managerial, and psycho-sociological conditions for successful R&D". He believes this approach to be "more rewarding and closer to reality" than cost benefit analysis or other input-output techniques. He states "this approach gives policymakers and research managers a whole set of indications concerning how to act on important variables that relate to R&D performance".

The MITRE Corporation report suggests that a checklist prepared as a result of a Hughes Aircraft Company study of R&D productivity may prove useful in assessing individual and organizational effectiveness (Ranftl, 1978).

Numerous studies of R&D management show that the quality of R&D management is a major factor in determining whether the investment in R&D pays off or not (Pelz and Andrews, 1976; Badaway, 1982; Isenson, 1965). Thus if the results of an evaluation show that objectives have not been met, the quality of management would be a prime candidate for explaining the shortfall in performance.

TECHNIQUES TO EVALUATE R&D OUTPUT/IMPACT

R&D program evaluation techniques can be categorized into quantitative and qualitative methods as follows (Burgess, 1966; Roman, 1980; Lynn, 1978; Martin and Irvine, 1983; Takei, 1981):

- peer evaluation of the contribution of the R&D program to scientific or technical knowledge;
- counting the number of publications in refereed journals or conference proceedings, in a particular time period;
- counting the number of citations a scientific or technical publication or report receives;
- counting the number of awards or honours the laboratory or professional staff receive in a given time period;
- counting the number of technical reports produced in a given time period;
- counting the number of potential patents disclosed or patent applications filed in a given time period;
- determining the degree of commercial success that the R&D output (i.e., products, processes, or services) has had, usually in terms of dollars (if, applicable);
- cost benefit analysis;
- client/customer evaluation of the usefulness, timeliness or effectiveness to them of the R&D output (products, processes or information); or
- comparison of the quantity or quality of R&D output with other similar R&D programs.

Quantitative methods are neither superior nor inferior to qualitative methods for evaluating R&D programs. Both methods have their place and should be used together to properly and effectively evaluate an R&D program. For example, cost benefit analysis could form part of client evaluation or laboratory comparison.

In a review of procedures to measure R&D productivity, Pappas and Remer (1985) suggest that the more quantitative methods are more appropriate for the development/product improvement end of the R&D spectrum while more qualitative (e.g., peer review) techniques be used to evaluate basic or applied research activities.

In their study of evaluation of government R&D programs, Salasin et al (1980) group five evaluation techniques into three categories. First are those techniques that are intended primarily for evaluating the outputs or impact of a research program. They are retrospective studies of impact, cost/benefit analysis, publication/patent counts, and citation analysis. Second are approaches to the evaluation of program management, which are directed at evaluating process and intermediate management output rather than R&D output or impact. Approaches to management evaluation include checklists, questionnaires and structured interviews. Third, peer evaluation can be applied to assessing process, output and/or impact of a research program.

Retrospective studies such as HINDSIGHT (Sherwin and Isenson, 1966) and TRACES (N.S.F., 1969) have been conducted to examine the contribution of research findings to the development of useful technologies. However, the long time lag between the production of research and its utilization (5 - 30 years), and the questions about the representativeness of the technologies that are selected, makes it difficult to use such retrospective studies for ongoing program management.

The following techniques are more amendable to the evaluation of R&D programs in a reasonable time frame.

Peer Evaluation

Although this technique is not generally considered to be reliable in the minds of many program evaluators who deal solely with non-R&D type programs, it is considered to be a valuable method in the area of R&D program evaluation. A report of the U.S. Committee on Federal Laboratories Task Force on Performance Measures for Research and Development (National Bureau of Standards) underscores the importance of peer evaluation when it states, "The generally recognized best procedure for evaluating research and development is one in which peer and other technical experts, including management, jointly judge the progress towards goals of ever increasing definition and mutual acceptability". Another study conducted at the U.S. Air Force R&D Laboratories showed that professional colleagues are suited to evaluate the innovativeness and productivity of researchers' output (Stahl and Steger, 1977).

The key problem in the use of peer review is that "the effectiveness of peer review depends critically upon the procedure for choosing peers" (Kochen, 1978).

Some of the more obvious problems are:

- identifying knowledgeable peers that are willing to act as reviewers;
- making sure that the selected peers have no conflict of interest; (e.g., receive grants from the program, or co- author papers with R&D program professionals);
- presenting information to the peer reviewers in such a way that interrater agreement is high; or differences can be readily understood; and
- developing a suitable and readily understandable rating instrument.

These difficulties will be compounded if the number of scientists or engineers working in the program area in the country is small. Reviewers from other countries or complementary disciplines may have to be contacted.

A multi-criteria approach to peer evaluation appears to be the route taken by several studies. Some of the factors which could be used by a peer to rate published articles or reports are:

- overall research quality;
- raising of important theoretical issues;
- contribution of an important method;
- resolving a recognized controversy;
- contribution of important data;
- leading the field in a false direction; and
- provision of new ideas for other investigators.

A MITRE Corporation report (Salasin et al, 1980) recommends that if peer or expert assessments are to be used as part of an evaluation, that guidelines should be prepared to describe:

- the viewpoints or characteristics of those stakeholders to be involved in the evaluation process so as to insure adequate representation of the various viewpoints;
- the types of material to be presented to experts for their judgements (e.g., program summaries prepared by the program manager, summaries of individual project findings, collections of project reports, etc.);
- the types of judgements that different groups of stakeholders might be asked to provide (e.g., industry users might be asked whether technical assistance received solved their problems, senior departmental management whether information resulted in the department meeting its overall organizational objectives, etc.);
- the preferred approaches to conducting or implementing evaluation activities (e.g., the types of personnel that should be involved in different aspects of evaluation activities);
- the policies regarding confidentiality and access to the distribution and review of evaluation reports;
- techniques that are suitable for the analysis and summarization of the evaluation data (e.g., ways of handling dissenting opinion); and

- a standardized format for program evaluation reports.

Number of Publications

One of the most common techniques for judging the output of an R&D program is to simply count the number of papers produced during the life of the program or during some given period. Although a relatively easy procedure, it has some important drawbacks which reduces its effectiveness as a valid technique, if used by itself.

Among the most important drawbacks are:

- the "counts" reflect quantity more than quality;
- even the "counts" may be misleading due to multiple articles in different journals or conference proceedings about the same work;
- in some disciplines, especially new ones, the communication of scientific/technical information may not be done via journal articles, etc.; and
- some employers actively discourage open literature publication for reasons of either commercial confidentiality or military secrecy.

Another danger of using publication counts in isolation is that it will encourage professional staff to produce a lot of "hack" articles to keep their numbers up.

A question that must be addressed is that of comparing one publication with another in terms of length. For example, how would one book compare with one journal article (Wilson, 1964; Lightfield, 1971). How would a review article rate compared with an article based on original work?

In another examination of the use of publications as an measurement criteria in evaluating basic research programs, Frame (1983) considers that before an evaluator can begin to collect and analyze data, he or she must answer a number of fundamental questions:

1. What publications are to be included in the evaluation? Only journal articles? Monographs? Research Reports? Conference Proceedings? Some combination of these?
2. What data sources should be used? Self-reporting from scientists/ engineers? Abstracts and /or citation indexes? Journals?
3. What time period should be examined? What time lags should be used to take into account that time elapses between doing the research work and getting it published.
4. Should publications that are only directly associated with the objectives or field of study of the program be included?

5. Have we controlled for scientific discipline to take into account the wide variations in publication rates from discipline to discipline?

6. What control group is most suitable to use for baseline comparisons? Is the data on the control group readily available or does it have to be developed?

Frame suggests the use of "Field Norm Tables" which gives the publication rate for several scientific/engineering disciplines. In most cases these tables would have to be developed by the evaluator to fit the disciplines he or she was interested in.

Frame acknowledges that using counts of scientific papers for program evaluation is a method more applicable to evaluating basic research and is most viable when applied to groups rather than to individuals.

He further suggests that counts of scientific papers can be used as a cost benefit measure. Specifically, how many research papers are generated per research dollar? This could be compared to other comparable research programs if the pertinent data were available. Martin and Irvine (1983) also employed this measure using operating costs of the research centre as the dollar input rather than total cost of the research so as to avoid the cost of equipment in a particular year biasing the result.

Vollmer (1967) warns, that publications as an indicator of performance would only be valid in laboratories where one of the objectives was to contribute to scientific knowledge. In an applied R&D program, evaluation along these factors would likely indicate that the laboratory was performing below par whereas criteria which measure the applications of science to new products or processes would be more important and more in line with corporate objectives.

Citation Analysis

A modification of publication counts is the use of citation analysis techniques. This technique involves simply counting the number of times a particular paper or report has been referred to or cited by other researchers in their publications. There are several drawbacks to using citations as a measure of R&D impact or quality. Among the distorting influences affecting the use of citation measures are:

- inability to cite the output of secret military and commercially proprietary industrial research;
- reliance upon personal communication, rather than on formal communications channels (e.g., reports or articles) for the exchange and transfer of technical information;
- repeated articles by the same author(s) to present the same findings;
- excessive citations of associate investigators; the "I'll cite you and you cite me" routine;
- differences between fields in their citation behavior and publishing practices (as noted by Frame, 1983);
- self-citation;

- only the first author of a multi-author paper is used in the citation system;
- variations in how much emphasis is devoted to disseminating the research; and
- variation in how extensively the article submitted to a journal is refereed.

Martin and Irvine (1983) note the following as additional problems in using citation analysis:

- authors with identical names;
- clerical errors in preparing the citation index;
- incomplete coverage of journals;
- critical citations;
- "Halo effect" citations (citing a well known author); and
- variation of citation rate of a paper due to either initially unrecognized advances, or integration of the major advance into the pool of common knowledge resulting in a lack of formal citation of the original source (e.g., $E = MC^2$).

Citation measures cannot be employed until several years after the publication of research findings because of the substantial time lags involved in publication. This delay was one factor which has led to the growth of technical reports. Because of the time delay, it can only be applied to longer term R&D programs.

Several studies have shown that peer evaluations of the publications quality have correlated positively with the results of citation analysis.

Eugene Garfield (1979), a principal contributor to the development of a citation analysis database warns its users that "citation data is subtle stuff. Those using it to evaluate research performance at any level must understand both its subtleties and its limitations. However, none of the grounds for criticism are insurmountable obstacles in the way of using citation data to develop fair, objective, and useful measures of individual or group performance."

Number of Awards or Honours

Another indicator of R&D program performance is the prestige of the R&D program as seen by outside organizations (Roman, 1980). A measure of this prestige is the awards which the professional staff receive from their scientific/professional societies (Schainblatt, 1982). Some organizations use such factors as staff

elected to national academies and staff invited to sit on government committees as indicators of R&D program quality. This evaluation method is basically a form of peer assessment with the evaluators being drawn from beyond the immediate boundaries of the scientific/technical discipline covered by the R&D program.

Since these indicators are more closely tied to the evaluation of individual scientists or engineers, they can suffer from the problems which have been indicated earlier such as the "Halo Effect" or the "Matthew Effect" (Merton, 1968). The Matthew Effect is said to occur when a researcher's work is not given the recognition it deserves because the researcher is relatively unknown, or a researcher's work is given more acclaim than it deserves because of the prior high reputation of the researcher. For example, a group considering a candidate for an award, may consider the person's or the R&D program's past performance and not necessarily present performance.

These indicators should definitely be used in conjunction with other more direct indicators such as peer assessment, program comparison or citation analysis, and not used alone.

Number of Technical Reports Produced

In organizations where restrictions on publication in the open literature exist, it is felt that the counting of internal technical reports is a suitable substitute. Again, as in the case of publications in the open literature, the number of technical reports is more a measure of quantity rather than quality.

Emphasis on only the number of reports in a given time period could encourage the professional to produce several "hack" pieces rather than one solid, high quality report. If a journal publication or conference paper is written based on a technical report, is the resulting paper given credit also as a publication?

This methodology should be used in conjunction with client/customer evaluation to get a measure of quality.

Number of Patents

Patent counts and/or patent statistics have also been used as an indicator of research output. However, this measure has many limitations for the measurement of the output of R&D activities.

Inventions or other outputs of R&D may not be reflected in a patent statistic because:

- the inventors perceive the inventions lack commercial application and/or are used in-house only and are thus not patented;
- the propensity to patent varies with the ease with which the technical advance can be copied;
- the inventors believe that there is greater secrecy for their intellectual property if a patent is not filed;
- the rate of change of technology may be so fast that a patent would not provide any economic advantage;
- some inventions are not patentable;

- the inventors are more concerned with scientific and technical leadership than patent protection;
- the existence of legal factors is seen as blocking the patent application;
- the cost of patent filing and maintenance fees may reduce the rate of patent filing; and
- patents do not generally reflect the level applied research and inventive activity of federal government laboratories where the propensity to patent is much lower.

Other problems with patent counts as a measure of R&D output are: quality variations in patents that are not reflected in patent statistics, patents pertain primarily to developmental efforts and the differences in patenting behaviour between industries and technical disciplines.

Patent Citations

As in the case of publications citing earlier work, patent applications often cite previous patents. Some authors consider that counting patent citations is a measure of the quality of scientific output and impact.

Narin, Carpenter and Woolf (1984) believe that it is possible to assess technological performance through the examination of patent citation statistics. The more times a patent is cited in subsequent patents, the more technologically important the patent. Their underlying assumption is that, "patents and patent citation analyses are a valid reflection of technical productivity and communication". Patent citation data analysis must take into account patenting frequency norms for the technology class.

Narin, et al, consider that an important advantage of patent citation analysis is that it is unobtrusive. It does not require the active involvement of any of the patent originators. Because patent applications eventually become public documents, it is possible to compare the R&D performance of individuals, companies, government or academic laboratories and countries in a given technical field.

Degree of Commercial Success

While a methodology more appropriate to industrial R&D programs, many government R&D programs in, for example, the National Research Council or the Canada Centre for Remote Sensing, are concerned with the successful transfer of technology to the private sector for commercial exploitation.

Basically this technique involves determining the economic returns from the successful introduction of a new technology in the market place. Patterson (1983), for example, describes a system adopted by Alco Laboratories for evaluating the economic contribution their R&D was having on the corporation.

One of the most common techniques is to measure the fraction of company current sales from products developed in the last five years. A variation on this technique is to determine the level of sales that has directly resulted from a specific new product or process. In the case of government-developed technology, the flow of revenues from licenses could be used as a measure of commercial success.

Commercial success can also be measured in terms of the number of new products or services introduced into the market place, new markets penetrated, technical problems solved, and new jobs/wealth created.

It should be noted however, that what is being measured is not only the impact of the R&D output, but also the quality and efficiency of the commercialization side of the innovation process such as production, marketing and planning in the private firm (Collier, 1977). Thus an R&D program could be extremely effective in developing a particular commercializable output, but the commercialization could fail due to poor technology transfer, bad marketing, actions by a competitor, or an untimely market entry. The evaluation of the R&D output in commercial terms must focus on the R&D element of the innovation process in order to get an accurate appraisal.

Cost Benefit Analysis

Cost benefit analysis (CBA) has been developed as a systematic approach to evaluating alternative projects and can be applied to evaluation of project and program performance.

Cost benefit analysis is most easily applied in systems for which the costs, benefits, and relationships among research, technology transfer or commercialization, and productivity are reasonably well documented.

A participant at the MITRE workshop on evaluation of R&D summarized the difficulties in applying cost benefit analysis as the following: " Cost benefit analysis in science presents you with two problems;- it is terribly difficult, almost impossible, to determine what the costs are, and the benefits are also extremely difficult to assess".

The effective use of cost benefit analysis is more likely to occur when this technique is used in conjunction with patent/publications counts, citation analysis and program comparisons.

Customer/Client Evaluation

Essentially a variation on peer assessment, this methodology would involve interviewing the recipient of the R&D output, to determine whether the technology or information was helpful in terms of, for example, solving or avoiding a technical problem, making a decision, developing a science-based policy or regulation, or preparing for some required action.

This approach is a measure of the degree of satisfaction of a client with the R&D output provided by the R&D program. Each client will have their own "measures of satisfaction" depending on what they want from the R&D program. Some government departments conduct client satisfaction surveys to collect this type of information (Clarke, 1997) In the case of the client being a private company, the R&D output might be evaluated in terms of cost savings or increased market share or profits.

Intended recipients of government R&D output would include senior management in a department, other branches of the government or external private or public organizations or individuals.

R&D Program Comparison

This methodology involves comparing the R&D outputs from the R&D program being evaluated to a similar R&D program(s) elsewhere. If effect, it results in a multiple evaluation of several R&D programs in order to rank the quality of the particular R&D program in question.

In a major study to determine whether this methodology is effective in assessing basic research programs, Martin and Irvine (1983) pose the following questions:

- to what extent can the outputs of research centres or groups working in the same field be compared; and
- is it possible to assess whether one research group has contributed more to scientific knowledge than another, and if so, how?

They suggest that a multicriteria approach can be a valid method of evaluating basic science research programs. The criteria which they call "partial indicators" of scientific performance by a research centre are:

- number of publications in refereed journals or conference proceedings;
- number of citations; and
- peer evaluation of the contribution of the R&D centres to scientific knowledge compared to other similar centres.

While acknowledging the shortcomings of each of these criteria if used on their own, they consider that when used together many of the problems are reduced.

The values obtained from the partial indicators of a particular R&D program or centre under evaluation is then compared with values obtained by measuring the scientific output from other similar R&D programs in the same scientific area in other locations. These other R&D programs are, in effect, control groups, or reference points. Factors such as size of R&D program in terms of total funding and total professional personnel are normalized to remove the "level of effort" influence on the partial indicator results.

Martin and Irvine have applied this technique on an assessment of radio astronomy facilities in England and have found that all three partial indicators converged to the same finding in terms of the contribution of these R&D centres to the advancement of scientific knowledge in the area of radio astronomy. It should be noted that the prime objective of these research centres was to contribute to scientific knowledge and so Martin and Irvine were, in effect, evaluating the degree to which the centres had met their objectives.

The evaluation process recommended by Martin and Irvine is very similar to that proposed by Takei (1981) in evaluating industrial R&D programs.

A common theme which runs through both the industrial and government R&D program evaluation processes is that multiple criteria are used. Vollmer (1967) considers that any evaluation of research effectiveness must use multiple criteria and thus multiple approaches to obtaining data.

EVALUATION OF GOVERNMENT R&D PROGRAMS

As noted in the introduction, only a few studies were found which dealt with the problems of evaluating government R&D programs. One such study is the report of the White House Science Council's Federal Laboratory Review Panel which found several serious deficiencies which resulted in Federal Labs not meeting quality and productivity standards (OSTP, 1983).

One of the major difficulties in evaluating government R&D programs is the problem of identifying the government department or agency's long-term strategic R&D objectives or goals. In many cases these

long-term objectives do not exist, as they change as the government's shorter term priorities change. This makes R&D planning very difficult and in turn makes R&D project selection in support of R&D program objectives chaotic. Without alignment of R&D program objectives in support of the overall objectives of the parent department or agency, R&D output will not be evaluated very favourably. Thus the R&D management process which determines whether this alignment takes place must also be subject to evaluation as part of the overall R&D program evaluation.

In the early 1980's, the Task Force on Science Policy of the House Committee on Science and Technology in the United States asked the Office of Technology Assessment to study, "the models and other analytical tools developed by economists to judge capital investments, and the applicability and use of these models and tools to government funding of scientific research". In 1986, the O.T.A. reported that, "using economic returns to measure the value of specific or general federal research expenditures is an inherently flawed approach". The only exceptions to this rule are certain Federal R&D programs whose specific goals are to improve the productivity of particular industries or industrial processes (O.T.A., 1986, p. 26). The O.T.A. report notes that, "the fundamental stumbling block to placing an economic value on Federal R&D is that improving productivity or producing economic return is not the primary justification for most Federal R&D programs".

The O.T.A. report concludes that, "bibliometric and other science indicators (e.g., statistics on scientific and engineering personnel) can be of some assistance, especially in research program evaluation, and should be used more. However, they are extremely limited in their applicability to interfield comparisons and future planning" (O.T.A., 1986, p. 9).

Peer review, in one form or another appears to be a popular method of evaluating government R&D programs. Fasella (1984) in his review of the evaluation of the European Community's R&D programs notes that peer review, and interviews with R&D program personnel and potential users of the R&D output, conducted by a panel of external independent experts, are the two main R&D program evaluation methods used. An extensive review of R&D program evaluation methods and experience in the European Community and in the U.S. is provided in the proceedings of an evaluation of research and development seminar held in Brussels, Belgium in 1983 (Boggio and Spachis-Papazois, 1984).

Fundingsland (1984) also reports on the extensive use of peer review in the U.S., either in the form of individual experts or panels, to evaluate research from the proposal stage to the output stage in organizations such as the National Science Foundation and the National Institutes of Health.

One major, comprehensive review of the program evaluation process as it applies to government R&D programs was found. Entitled "The Evaluation of Federal Research Programs", it was intended to explore and assess the state of the art of evaluating government R&D programs (Salasin, Hattery and Ramsay, 1980). They consider the following to be the major difficulties in the evaluation of government sponsored R&D programs:

1. Defining Success

There is no straightforward definition of what constitutes successful research for a given program. Research "success" can take different forms. For instance, contributions can be made in the acquisition of new knowledge and/or by making achievements that contribute directly to an agency's mission. Even research that is not immediately productive may present opportunities for later developmental efforts. There may be problems in identifying scientific advances. The "technological" outcome of research (e.g., specific devices or processes) may be more easily identified than the "scientific" outcome (e.g., the research's contribution to knowledge). The "value" of a technological outcome will almost always be easier to determine than the value of a scientific advance. Evaluative methods should distinguish between these two outcomes and be adapted to each.

The "significance" of research is difficult to determine due to the time lag between research output and the impact of the output. (The difficulty of assessing the significance of research work is illustrated by the 32 years it took the scientific community to assess the contribution that Dr. Barbara McClintock made in the field of medicine. She conducted her research in 1951 and received the Nobel Prize for Medicine in 1983.) Research has its major significance, not in the specific findings of a given program or project but in the implications that these findings may have for future work. An effective evaluation

approach should, therefore, include projections about future developments based on current research activities. "Forecasting" should be an integral part of the evaluation.

The identification of output may be a complex and subtle problem, even when a clearly defined and measurable goal has been set. This difficulty occurs because the output from an R&D program is usually multifaceted. For example, along with meeting a specifically defined applied mission objective, the program may also produce major advances in science or engineering. How should these outputs or impacts be treated?

If an evaluation is to ascertain the impact of R&D findings, it must include qualitative as well as quantitative information. When quantification is used in conjunction with qualitative analysis, it provides important support for conclusions and recommendations. But the numbers have no magic of their own and their "objectivity" is an illusion. They are the product of people who subjectively collected them, sorted them, interpreted them and used them to make a point (Siedman, 1977).

It seems unlikely that the aggregation of such qualitative information can be conducted in a routine or mechanical manner. An in-depth understanding of a program's substantive area is needed to assess the meaning and interrelationships of findings from individual projects. [This would reinforce the belief that the evaluator or a member of the evaluation team must have a scientific background].

2. Multiplicity of Objectives or Impacts

Research problems may involve basic sciences, technology, or institutional and personal relationships. The complexity of these problems requires that a program be evaluated with respect to multiple objectives or impacts. For example, an R&D program could have impacts on: policy, program delivery, contribution to knowledge (theoretical or applied), technology transfer, the development of new products or processes, contribution to methodological techniques and thus future research, and/or the training of future scientists.

An unexpected objective or impact may be realized which would have to be taken into account in assessing the "Impacts and Effects" of the R&D program.

3. Aggregating Project Evaluations

Many R&D programs are made up of a collection of individual R&D projects rather than a group of activities aimed at a program objective. The problems encountered in evaluating project based programs may include:

- the contribution of very small projects to program objectives may be difficult to assess;
- the number of projects may be so large that evaluation may be very time consuming and costly;
- projects may be carried out by third parties making the collection of evaluation data more difficult and costly; and
- there is a temptation to simply add together the outputs of the individually identified project.

Evaluation of a program may not be achieved simply by aggregating evaluations of constituent projects. A program evaluation should be related to the program's central theme or objective. The interrelationships and balance of project outcomes as related to the central program mission should be included in a program evaluation. Further, some program activities may not be attributed to any individual projects. Activities that are undertaken to build networks of researchers (for the exchange of scientific/technical information), to provide technical support or training to users, or to coordinate efforts with other research programs may be undocumented efforts of program staff rather than project related activities.

4. Reconciling Political and Scientific Viewpoints

Throughout the workshop conducted as part of the MITRE study, there was a recurrent theme of accountability and political review. At the political level, the oversight or evaluation is generally in the context of national goals, whereas, at the program level the context may be the contribution to much more restricted goals, or to the discipline.

Several rather strong admonitions were presented in the workshop to the effect that the political responsibility and accountability of all federal programs to central policy agencies must be attended to. Researchers may resist accounting for their activities and the quality of their performance to members of the government, however, as users of public funds, this is an unavoidable responsibility. It seems apparent that much needs to be done to reconcile assumptions, criteria, and evaluation contexts at governmental, departmental, agency and program levels.

While difficult, this reconciliation must be made to ensure continued support by the elected officials for the R&D programs.

5. Resistance by Scientists to Evaluation

The barrier to evaluation set up by the scientists and engineers in R&D can be considerable. It has been noted by management researchers that research scientists, in particular, tend to resist the attempts of non-scientists - managers, clients or the general public - to evaluate scientific work. Scientists are generally highly professionalized people. Like physicians, lawyers, and other professionals, they maintain that only a member of their profession is able to evaluate their work. They are also trained like other professionals to keep their judgements about professional skills of colleagues within their own circles (Vollmer, 1967).

The uproar in England over the public evaluation of the research programs of several radio astronomy research programs by management researchers at the Science Policy Research Unit of the University of Sussex is a clear example of what happens when scientist's shortcomings are aired in public (Dickson, 1983).

Thus a government R&D program evaluator should be prepared to face considerable hostility, especially from older scientists, and especially if he or she does not have a science degree.

Several studies have noted the need for the R&D program evaluator to have a technical background, or have someone on the evaluation team with a technical background in the scientific discipline being evaluated (Bennett and Jaswal, 1982).

Fundingsland (1984, p. 111) considers that the most common weaknesses in the evaluation of mission-oriented, government R&D are:

- inadequate definition of goals and expectations;
- failure to anticipate outcomes, including probable users and potential utilization, side effects and risk assessment, and technology transfer and application problems;
- too little involvement of potential users in initial planning (i.e., especially important for those government laboratories whose mission is to support industrial clients);
- failure to delineate criteria for success and establish an independent evaluation team to monitor progress from the beginning;
- failure to establish baseline data and a system for developing essential information and data from the beginning; and
- inconsistent oversight and feedback.

Callaham (1985) in his review of the evaluation of forestry research programs considers that innovation, modifications to innovations, and scientific findings should be used as measures of accomplishment in research and development programs. In an experiment to use such measures, scientific staff at a forestry laboratory were asked to list their achievements over a five year period under three categories:

Invention and Innovation

Fully developed new or useful product, process or technique accepted and currently in use. An innovation generally evolves from a "set" of scientific discoveries, inventions, and real-world problems or opportunities.

Modification of Innovation

Extends the use of an innovation to new geographic areas, species or problems. Also includes refining or adapting existing innovations or making them more cost effective.

Scientific Finding

Contributes to the broad base of knowledge or methodology. Differing from innovations, they are not directly marketable or useful outside of science.

They then rated each achievement against the following sixteen categories of social benefits developed specifically for the forestry sector, to obtain an overall score:

- generated income/employment in the forest industry or regional economy;
- increased utilization of natural resources;
- improved quality of physical/biological environment;
- lowered prices or costs to consumers;
- reduced costs for managing resources;
- improved methods for planning and evaluating alternative investments;
- resulted in new or improved products;
- improved visual environment and related amenities;
- increased resource productivity;
- reduced costs through improved products;
- improved scientific methods and theory;
- improved social environment;
- raised quality, lowered cost of housing;
- enhanced health and safety;
- improved cultural/historical/geological environment; and

- enhanced public involvement in decision-making.

An important aspect of this evaluative approach was the involvement of users or customers of their R&D output as part of the review and evaluation process.

Callaham concluded that, "evaluating the social benefits from past, present and future research is technically difficult, but stimulating and rewarding work".

EVALUATION OF INDUSTRIAL R&D PROGRAMS

Fumio Takei (1981) describes a method by which a company can evaluate its engineering program. It basically consists of comparing the output of the engineering program with that of the engineering program of rival companies in the same product line in terms of:

- the timely introduction of new or improved products;
- product performance;
- product reliability;
- number of patents per engineer;
- sales per engineer;
- number of publications and submissions of technical papers per number of engineers; and
- quality of patents

In addition, in the case of scientific or technical papers, R&D personnel in corporate headquarters compare the quality of papers, etc. produced by divisional R&D programs with those published in the same field by academics and personnel in research societies.

The major difficulty in applying this program comparison methodology is to locate a similar R&D program that provides valid comparisons. For example, an R&D program which emphasized technical reports and hardware development would be a poor comparison for another R&D program which had as its major objective to contribute to scientific knowledge through open literature publications.

As noted in the Martin and Irvine paper, the level of financial support that a program receives must also be taken into account.

Where possible, this methodology would provide valuable information to senior management regarding the overall effectiveness of their R&D program activity.

It has been suggested that the "inventivity" of a firm can be determined by calculating the ratio of number of patents to R&D expenditures. Using this approach, Gilman and Siczek (1985) found that large firms were not as inventive as smaller firms. One drawback of this method is that not all R&D activities are destined to result in patents.

EVALUATION OF R&D PERSONNEL

Because of the considerable degree of overlap between R&D program evaluation and R&D personnel appraisal methodologies, the results of studies dealing with personnel evaluation provide additional insight into the problems faced by program evaluators.

The difficulties faced by evaluators in determining whether a research program has met its program objectives are similar to those faced by R&D managers when evaluating their scientific or engineering staff.

The performance of scientists and engineers is assessed or appraised for the following reasons (Grove, 1985):

- improve performance;
- motivate personnel;
- assess the subordinate's work;
- provide important feedback to a subordinate;
- reward and/or recognize exceptional performance;
- provide guidance on work direction;
- provide discipline;
- justify a raise or promotion; and
- reinforce the organization's culture and environment.

To this list could be added one more reason why many R&D managers evaluate their subordinates: the organization insists on a yearly performance appraisal. If this is seen as the sole reason for the evaluation, the potential benefits that can result from an effective evaluation exercise will be lost and the result will be a poor performance appraisal.

The characteristics of a poor performance appraisal are:

- review comments are too general or superficial;
- appraisal results in little or no action and becomes a meaningless exercise;
- no advice or guidance is provided of how to improve performance;
- supervisor avoids bringing up any negatives to avoid confrontation (self-censorship of criticism);
- no real encouragement by superior to improve;
- supervisor not familiar with work and must rely on employee to tell him/her what he/she has done during the year;
- the appraisal is limited to past activities and does not include future opportunities;
- subordinate gets a nasty surprise; and
- supervisor shows no real interest in the evaluation activity and the career of the subordinate.

It is important to assess not only the results that an employee has achieved but also the procedures used to achieve those results. High performance can be achieved at an unacceptable cost in terms of both financial and human resources. The subordinates efficient use of human and financial resources should be an important element in the performance appraisal.

There may also be a time lag between the process or procedures used to achieve objectives (e.g., managerial style) and the identifiable output such as problems solved or new knowledge created. High quality output today might be the result of past good management, while tomorrow's research output could be threatened by present ineffective management practices. New R&D managers should be assessed on the contribution they make to the R&D unit's performance, not on the inertia built up by the previous manager.

Five basic categories of obstacles have been identified to evaluating or measuring the productivity of scientists and engineers in R&D (Ruch, 1980). These are:

1. The difficulty of defining the output or contribution that is made by a knowledge worker.
2. Overcoming the tendency to measure activities (e.g., number of papers or reports produced) rather

than results or impact of the scientific/technical output.

3. The matching of inputs to outputs or impact within a reasonable time frame. Output or impact may occur years later.

4. Including a quality dimension in the measure.

5. Including the concept of effectiveness as well as efficiency in the productivity measure.

Although these difficulties were formulated in the context of evaluating the performance or productivity of individuals, they also constitute problems in the evaluation of R&D programs.

As Keller and Holland (1982) point out, in the search for valid measures of R&D professional performance, the pendulum often swings between a preference for objective indicators such as number of publications or patents, and subjective performance ratings by peers, superiors or by the individual in question. They conclude, however, that neither objective nor subjective performance measures have an inherent superiority. Each has its problems. The number of patents can be misleading because patents vary considerably in their commercial success, and all publications do not have equal value. Among the problems encountered in subjective ratings by peers or superiors is the "Halo Effect". The Halo Effect occurs when the rater does not differentiate among the several performance dimensions to be rated, but instead evaluates the ratee on a global or overall basis.

One problem that affects both objective and subjective measures of R&D performance is Merton's concept of the "Matthew Effect". Merton (1968) observes that those who have achieved considerable recognition are given even greater recognition in the future, while those who have not made a name for themselves find that recognition tends to be withheld. In effect, a misallocation of credit for accomplishment occurs whereby earlier accomplishments are underrated while later ones are overrated. Thus well-known professionals find publication easier than do unknowns, and awards are given to those who already have awards. In the case of subjective ratings, the prior reputation of the ratee may overwhelm other considerations in the evaluation process or may intimidate the rater.

Many of these problems would have their counterparts in the evaluation of a laboratory program. For example, a rater may be intimidated by the generally high reputation of a laboratory and fail to assign a low rating to a particular R&D program even though it is warranted. Organizational reputations, good or bad, linger long after the cause of that reputation has ceased to exist.

The relationship between input and output in R&D has a variable time lag. It is even hard to know when there has been an output or to what input it can be attributed. In some instances, the results may be almost immediate and their parentage readily discernible. In other situations, return may be obscure, may never come, or may be so far in the future as to defy accurate correlation with the amount of input (Roman, 1980).

R&D Employee Evaluation Methods

In a review of papers concerned with measuring the performance of researchers, Edwards and McCarrey (1973) state that the basic drive for accurate evaluation methods stems from the need to distinguish between an above average scientist and an average one. They pose the questions should a scientist be judged on the basis of his/her contribution to science, to their employing organization, or to both; is it possible to evaluate a scientist completely on objective factors or must one consider subjective evaluations as well?

They conclude that scientific output is multi-dimensional and cannot be satisfactorily measured by any one criterion alone.

Among the factors used by organizations to evaluate the performance of scientific or technical staff are (Taylor et al, 1961; Grasberg, 1959; Cole and Cole, 1967; Edwards and McCarrey, 1973, Phelan, 2000):

- number of publications;
- quality of publications;
- citations by other authors;
- number and quality of technical reports;
- number of papers presented at conferences;
- patents disclosed or granted;
- quality of work as viewed by supervisor;
- peer ratings (both inside and outside the lab);
- reputation in the scientific/technical community;
- overall contribution to organization's goals;
- demonstrated creativity or original approach; and
- awards or honours received from scientific societies.

Many of these factors are also used to evaluate R&D programs, but as noted earlier, there are several validity and reliability problems associated with these measures such as the unequal quality of publications and the Halo Effect in subjective ratings which have to be overcome or neutralized to ensure a valid evaluation.

A study funded by the National Science Foundation focussed upon assessing the quality of manuscripts. Reviewers were asked to prepare a narrative report and complete a rating form. For the narrative report, reviewers were asked to state their understanding of the purpose of the study and to organize the narrative report according to the following criteria:

1. Scientific Merit

Discuss the quality of research performance including:

- (a) analysis and support for conclusions;
- (b) design and methodology;
- (c) coverage of related research;
- (d) quality of application of mathematical principles;
- (e) quality of data, given the use to which they were put; and
- (f) citation of references.

The reviewers were also to compare the quality of the manuscript with journal articles, dissertations, and other related work published during a designated time period.

2. Accomplishments

Describe accomplishments of the work including:

- (a) basic scientific advances such as developments in theory, new information, new technology, systems or procedures;
- (b) advances in the specific field of application; and
- (c) achievement of the research objectives as you understand them, and whether or not a significant advance in the field has been made.

3. Utility

Discuss the utility of the study including:

- (a) usefulness for further research;
- (b) usefulness for policy-making or organizational decisions; and
- (c) feasibility of application of the results, including applications outside the direct field of study.

4. Report Format and Readability

Assess the report's format and readability in terms of clarity of writing and appropriate use of tables, charts, symbols and notations.

5. Dissemination Recommendations

Identify specific individuals and groups to whom the research output should be directed. Recommend the extent of dissemination and specific media which should be used (e.g., journals, conferences, etc.).

The NSF research reviewers were instructed to rate the research reports on the above criteria and to make an overall quality rating. The evaluators observed that differences in the background of research reviewers (e.g., area of training) substantially influenced their ratings. Despite this, it was noted that ratings of scientific merit, accomplishment and utility were usually closely interrelated.

Evaluation Interview

An effective performance appraisal interview relies on the supervisor having good interpersonal communications skills. One aspect of good communications skills is knowing how to deliver the results of an appraisal to an employee. In any situation where either complex, or emotionally charged information is to be communicated, it is best to provide the information in two ways: orally and in writing.

An employee should be provided with a copy of the performance appraisal in advance of the appraisal interview. This allows the subordinate to read, and re-read the appraisal in order to increase his/her understanding. The employee also has a chance to organize their thoughts and questions, and if necessary, to have any emotional outbursts in the privacy of their own homes. The interview should take place within 24 hours of providing the subordinate with the written appraisal.

There are numerous articles available which describe various appraisal form formats. The heart of any form should consist of just two questions:

- What was the employee asked to do, or what objectives had the employee agreed to attain?
- How well did the employee do in these activities and what factors contributed to this performance level?

If the supervisor cannot answer question one, he or she has no business answering question two.

In order to avoid psychological barriers being erected in the first paragraph, appraisals should begin with a description of what the employee was expected to accomplish during the reporting period. The order of these activities or objectives should be such that when answering the question regarding performance, the supervisor can start with some positive supportive statements. To start the evaluation section with criticism will only colour the subordinates perception of the remaining comments, even positive ones.

The subordinate's reaction to an appraisal can run the gamut from enthusiastic approval to outright rejection. In the case of a negative appraisal, the supervisor should not be surprised if subordinate denies the existence of the performance problem, or blames others. The supervisor must be prepared to point out, with clear examples, the negative impact of the subordinates poor performance not only on themselves, but on others in the organization. Since the most important objective of performance appraisal is to improve performance, an effective R&D manager also provides guidance to the subordinate on how the negative performance can be improved.

Until the subordinate agrees that the performance problem exists, and that it is his/her responsibility to correct it, corrective measures will not be taken. For example, a researcher who denies having a writing problem will not take the necessary measures, such as attending a report writing course, to overcome the skill deficiency.

Even when providing a very positive assessment to a high performer, there may be areas where the person can improve, or new skills that the person can acquire in order to advance their professional career.

The major beneficiary of the performance appraisal should be the employee, not the employer.

SUMMARY

Each of the impact factors leads to a series of questions which can be best answered by one or more of the particular data gathering methods. Some of the factors, however, result in a question which is simply a restatement of the factor in the form of a question.

Basic Research

Advances knowledge in a specific field of science

Has this research added to our knowledge of a scientific area in an incremental or "breakthrough" manner?

Raises important theoretical issues

Has the output raised important theoretical issues?

Resolves a recognized controversy

Has the output resolved a scientific controversy?

Advances knowledge in research techniques

Has the research process advanced research techniques?

Will the new techniques enable new lines of research to be explored?

Results in a pool of highly qualified researchers

Has the fact of working on projects within the R&D program resulted in developing a new generation of highly qualified professionals?

Has the program contributed to the maintenance of the scientific expertise of the older staff?

Develops An R&D Capability in a University or Other Organization

Through contracting out research or through staff/student interchanges, has an R&D capability been developed in an external organization?

Evaluation Methods

Among the most appropriate methods for determining the quality of the impact of basic, undirected research are:

- peer review;
- number of publications in prestigious, refereed journals or in the proceedings of prestigious conferences/symposia;
- citation analysis;
- comparison with similar research programs; and
- user/client surveys (if, applicable).

Mission-Oriented Research and Development

Advances in knowledge in the specific field of application

Has this research added to our knowledge of a scientific area in an incremental or "breakthrough" manner?

Has this research added to our knowledge of how to apply the scientific information to a technological problem?

The degree of utility/usefulness of product/hardware developed

Has the R&D resulted in a product or piece of hardware that proves the feasibility of a scientific application?

Has the R&D resulted in a product or piece of hardware that can be used by a client in a cost effective manner?

Has the R&D resulted in a product or piece of hardware that is still needed? (i.e., fits the markets' "window of opportunity")

The degree of utility/usefulness of a process developed

Has the R&D shown that a process can be used at laboratory scale?

Has the R&D resulted in a process that can be successfully, operated at plant capacity, and be cost effective?

Has the R&D resulted in a timely development of the process?

Has this new process resulted in cost savings for the client?

The degree to which information could be applied to *solve an operational problem in a cost effective manner*

Has the information provided from the R&D activity solved a client's problem?

Has the client accepted the information and applied it?

Are the number of requests for R&D advice increasing or declining?

Advances in knowledge of research techniques

Has the research and development process advanced research techniques?

Will the new techniques enable new lines of research and development to be explored, or new applications possible?

The degree to which the information has assisted in decision making, or policy or regulation formulation

Has the information played a major/minor role in organizational decision making?

Has this information contributed to the formulation, design, and/or conduct of other research projects/programs?

Has the information resulting from the R&D been incorporated into government policy?

Has the information resulted in more effective health or safety regulations?

Are the findings of this program potentially relevant to future policy debates or international agreements?

The degree to which the information has contributed *to the education of the general public to scientific issues*

Are the findings of the R&D being communicated in popular literature?

Are the findings being discussed in newspapers or on TV?

Are the R&D results being incorporated in text books?

Results in a pool of highly qualified, applied researchers

Has the program resulted in the development of a new generation of applied researchers?

Are these researchers in demand by other R&D programs?

Has the program contributed to the maintenance of the scientific/technical expertise of the older staff?

Develops a Technological Innovation Capability

Has the program resulted in the development of a new technological capability, either because of contracting out or through staff interchanges?

Successful Transfer of Technology

Has the program successfully transferred technology to the private sector?

Has the transfer resulted in licence revenue for the laboratory?

Has the transfer resulted in a new company, new product, process or service, being marketed nationally or internationally (i.e., wealth or job creation)?

Has the technology transfer had any negative effects on other companies?

Are companies actively monitoring the program's work to seek out commercial opportunities?

Evaluation Methods

Among the most appropriate methods for determining the quality of the impact of mission-oriented research and development are:

- peer review;
- customer/client surveys;
- cost benefit analysis of products/processes developed;
- number of products/processes developed;
- comparison of outputs with similar mission-oriented R&D programs
- number of patents issued;
- number of patents successfully developed into commercializable products;
- number of internal reports;
- number of problem solutions accepted and adopted;
- number of refereed publications (if applicable); and
- importance/magnitude of policies/programs/regulations affected by R&D.

Because mission-oriented research can involve some basic research, most of the methods of evaluating mission-oriented research must include the methodologies for evaluating basic research.

CONCLUSION

After reviewing the literature on the evaluation of R&D programs the following conclusions can be drawn:

- It is possible to evaluate R&D programs.
- The evaluation of R&D programs cannot be a textbook or mechanical process of "turning the crank". The evaluator's judgement is a necessary component and should not be discarded simply because it is subjective. Thus the quality of the evaluators personal skills, experience, and knowledge are of paramount importance.
- Because research and development produces multiple outputs, multiple criteria must be used, and the multiple criteria selected must be chosen so as to reflect the primary and secondary objectives of the R&D program.
- The criteria or techniques employed in evaluating R&D programs must include both "quantitative" and "qualitative" factors. Thus both publication type criteria and peer review are necessary to obtain a valid and reliable evaluation.
- To be a complete and accurate evaluation, the management and research process, the R&D outputs, and the impact of the R&D outputs must be evaluated otherwise the evaluator will be unable to determine the contribution that management or science/technology plays in achieving high or low performance.
- In order to obtain the needed information for evaluation it will be necessary to involve the program staff in the evaluation process to a greater degree than would be found in the evaluation of non-R&D programs.
- Due to the nature of R&D program/personnel evaluation, it is imperative that a person familiar with the management of science and technology be a member of the evaluation team.

In general, a systematic approach to evaluating R&D programs should (Salasin, Hattery and Ramsay, 1980):

- allow assessments of the "significance" of research in terms of its potential future impact; and
- actively involve all groups of "stakeholders" in the design and conduct of the evaluation to help ensure that the evaluation findings are useful to and believed by each group.

It is clear, however, that the quality of the R&D output and the resulting impact will depend critically on whether the operational and strategic objectives of the R&D program have been clearly and unambiguously developed, and communicated to the R&D management and researchers in the program.

To be useful a good R&D program evaluation must be understandable and be acceptable by those people who must make decisions about the program. The end result should be a program evaluation which enhances the ability of the R&D managers and their scientific staff to meet their objectives and goals in the most resource effective manner.

REFERENCES

Andrews, Frank, "Scientific Productivity: The Effectiveness of Research Groups in Six Countries", Cambridge, MA: Cambridge University Press, 1979

Apt, Kenneth E. and Watkins, David W., "What One Laboratory Has Learned About Performance Appraisal", *Research-Technology Management*, Vol. 32, No. 4, July-August, 1989, pp. 22-28

ARA, "Methods for Assessing the Socioeconomic Impacts of Government S&T", Vancouver, B.C.: The ARA Consulting Group, Inc., 1993

Badawy, M.K., "Developing Managerial Skills in Engineers and Scientists", New York: Van Nostrand Reinhold Co. Inc., 1982

Boggio, G. and Spachis-Papzois, E. eds., "Evaluation of Research and Development", Hingham, MA: Kluwer Academic Press, 1984

Bennett, W.D. and Jaswal, I.S., "The Evaluation of Government Research and Development Programs", Report No. PE 17/1980, Dept. of Energy, Mines and Resources, Ottawa, Ontario, February, 1980

Burgess, J.S., "The Evaluation of A Government-Sponsored Research and Development Program", *IEEE Transactions on Engineering Management*, Vol. EM-13, No. 2, June, 1966, pp. 84 - 90

Callahan, R.Z., "Evaluating Social Benefits of Forestry Research Programs", *IEEE Transactions on Engineering Management*, Vol. EM-32, No. 2, May, 1985, pp. 47 - 54

Chelimsky, Eleanor, "Proceedings of a Symposium on the Use of Evaluation by Federal Agencies", Symposium Report: Vol. 1, McLean, VA: Metrek Division, MITRE Corporation, March, 1977, 200 pages.

Chelimsky, Eleanor, "An Analysis of the Proceedings of a Symposium on the Use of Evaluation by Federal Agencies", Symposium Report: Vol. II, McLean, VA: Metrek Division, MITRE Corporation, July, 1977, 51 pages.

Clarke, Thomas E., "Review of Business Development Activities in Government and Private Sector Research Institutes in the UK and Holland, Ottawa, ON: Stargate Consultants Limited, 1997

Clarke, T.E., "Decision Making in Technologically Based Organizations: A Literature Survey of Present Practice", *IEEE Transactions on Engineering Management*, Vol. EM-21, No. 1, February, 1974, pp. 9 - 23

Clarke, T.E. and Reavley, "S&T Management Bibliography - 1993, 2010", Nanaimo, B.C.: Stargate Consultants Limited, 2010 [<http://www.tomeclarke.ca>]

Cole, S. and Cole, J.R., "Scientific Output and Recognition: A Study in the Operation of the Reward System in Science", *American Sociological Review*, Vol. 32, No. 3, 1967, pp. 377- 399

Collier, D.W., "Measuring the Performance of R&D Departments", *Research Management*, Vol. 20, No. 2, March, 1977, pp. 30 - 34

Dickson, David, "Study of Big Science Groups Hits Raw Nerve", *Science*, Vol. 220, April 29, 1983, pp. 482 - 483

- Edwards, S.A. and McCarrey, M.W., "Measuring the Performance of Researchers", *Research Management*, Vol. 16, No. 1, January, 1973, pp. 34 - 41
- Fasella, P., "The Evaluation of the European Community's Research and Development Programmes", in *Evaluation of Research and Development*, G. Boggio and E. Spachis-Papazois eds., Hingham, MA: Kluwer Academic Press, 1984, pp. 3 - 13
- Frame, J.D., "Quantitative Indicators for Evaluation of Basic Research Programs/Projects", *IEEE Transactions on Engineering Management*, Vol. EM-30, No. 3, August, 1983, pp. 106 - 112
- Freedman, Ron, "Evaluating the Impact of Publicly Funded R&D", *Proceedings of an External Review Workshop*, Ottawa, December 9, 1994. The Impact Group, 1995
- Fundingsland, O.T., "Perspectives on Evaluating Federally Sponsored Research and Development in the United States", in *Evaluation of Research and Development*, G. Boggio and E. Spachis-Papazois eds., Hingham, MA: Kluwer Academic Press, 1984, pp. 105 - 114
- Fusfeld, H.I. and Langlois, R.N., "Understanding R&D Productivity", New York: Pergamon Press, 1982
- Garfield, E., "Citation Analysis as a Tool in Journal Evaluation", *Science*, Vol. 178, 1972, pp. 471 - 479
- Geisler, Eliezer, "The Metrics of Science and Technology", Westport, CT: Quorum Books, 2000
- Gilman, J.J. and Siczek, A.A., "Optimization of Inventivity", *Research Management*, Vol. 28, No. 4, July-August, 1985, pp. 29 - 31
- Glass, E.M., "Methods of Evaluating R&D Organizations", *IEEE Transactions on Engineering Management*, Vol. EM-19, No. 1, February, 1972, pp. 2 - 12
- Grasberg, A.G., "Merit Rating and Productivity in an industrial research laboratory: A Case Study", *IRE Transactions on Engineering Management*, Vol. EM-6, No. 1, March, 1959, pp. 31 - 37
- Grove, A.S., "Performance Appraisal: Manager As Judge and Jury", *Research Management*, Vol. 26, No. 6, November - December, 1983, pp. 32 - 38
- Isenson, R.S., "Allowed Degrees and Type of Intellectual Freedom in Research and Development", *IEEE Transactions on Engineering Management*, Vol. EM-12, No. 3, September, 1965, pp. 113 - 115
- Keller, R.T. and Holland, W.E., "The Measurement of Performance Among Research and Development Professional Employees: A Longitudinal Analysis", *IEEE Transactions on Engineering Management*, Vol. EM-29, No. 2, May, 1982, pp. 54 - 58
- Kocaoglu, D.F., "A Participative Approach to Program Evaluation", *IEEE Transactions on Engineering Management*, Vol. EM-30, No. 3, August, 1983, pp. 112 - 118
- Kochen, M., "Models of Scientific Output", in *The Advent of Science Indicators*, Y. Elkana et al, eds., New York: John Wiley and Sons, 1978
- Lynn, L.E., ed., "Knowledge and Policy: The Uncertain Connection", National Academy of Sciences, Washington, D.C., 1978, pp. 18 - 19
- Martin, B.R. and Irvine, John, "Assessing Basic Research", *Research Policy*, Vol. 12, 1983, pp. 61 - 69
- Merton, R.K., "The Matthew Effect in Science", *Science*, January, 5, 1968, pp. 56 - 63
- Moser, M.R., "Measuring Performance in R&D Settings", *Research Management*, Vol. 28, No. 5, September - October, 1985, pp. 31 - 33

Murphy, S.R., "Five Ways to Improve R&D Efficiency", *Research Management*, Vol. 24, No. 1, January, 1981, pp. 8 - 9

Narin, F., "Evaluative Bibliometrics: The Use of Publications and Citation Analysis in the Evaluation of Scientific Activity", Cherry Hill, NJ: Computer Horizons, Inc., 1976

Narin, Francis, Carpenter, M.P. and Woolf, Patricia, "Technological Performance Assessments Based On Patents and Patent Citations", *IEEE Transactions on Engineering Management*, Vol. EM-31, No. 4, November, 1984, pp. 172 - 183

National Science Foundation, "Assessment of the Scientific Quality and Utility of Reports Produced by the International Institute for Applied Systems Analysis", Final Report, August, 1978

National Science Foundation, "TRACES (Technology in Retrospect and Critical Events in Science)", Prepared for the NSF by the Illinois Technology Research Institute, Volume 1 - December 15, 1968 and Vol. 2 - January 30, 1969

Neilson, J. and Brazzel, J., "Evaluation as an Aid to Decision-Making in the Food and Agricultural Sciences", Joint Planning and Evaluation Staff Paper, No. 80-DD-02, Science and Education Administration, U.S. Dept. of Agriculture, Washington, D.C., March, 1980

O.C.G., "Evaluation of Research and Development Programs", Ottawa, Ontario: Program Evaluation Branch, Office of the Comptroller General, March, 1986

O.S.T.P., "Report of the White House Science Council Federal Laboratory Review Panel", Office of Science and Technology Policy, Executive Office of the President, Washington, D.C., May, 1983

O.T.A., "Research Funding As An Investment: Can We Measure the Returns?", A Technical Memorandum, Washington, D.C.: U.S. Congress, Office of Technology Assessment, Report #OTA-TM-SET-36, April, 1984

Packer, M.B., "Analyzing Productivity in R&D Organizations", *Research Management*, Vol. 26, No. 1, January/February, 1983, pp. 13 - 20

Pappas, R.A. and Remer, D.S., "Measuring R&D Productivity", *Research Management*, Vol. 28, No. 3, May - June, 1985, pp. 15 - 22

Patterson, W.C. "Evaluating R&D Performance At Alcoa Laboratories", *Research Management*, Vol. 26, No. 2, March/April, 1983, pp. 23 - 27

Pelz, D.C. and Andrews, F.M., "Scientists in Organizations: Productive Climates for Research and Development", New York: John Wiley and Sons, 1976

Phelan, Thomas J., "Evaluation of Scientific Productivity", *The Scientist*, Vol. 14, No. 19, October 2, 2000

Ranftl, R.M., "R&D Productivity", Carver City, CA: Hughes Aircraft Co., Second Edition, 1978

Ruch, W.A., "Measuring Knowledge Worker Productivity", in *Dimensions of Productivity Research*, Proceedings of the Conference on Productivity Research, Houston, TX, American Productivity Center, April 20 - 24, 1980

Rubenstein, A.H., "Trends in Technology Management", *IEEE Transactions on Engineering Management*, Vol. EM-32, No. 4, November, 1985, pp. 141 - 143

Salasin, John, Hattery, Lowell and Ramsay, Thomas, "The Evaluation of Federal Research Programs", McLean, VA: The MITRE Corporation, NTIS Order Number PB 81234106, \$11.50, June, 1980

Schainblatt, A.H., "How Companies Measure the Productivity of Engineers and Scientists", *Research Management*, Vol. 25, No. 3, May, 1982, pp. 10 - 18

Seidman, E., "Why Not Qualitative Analysis?", Processed. Ms. Siedman is Director, Program Evaluation Unit, U.S. Commission on Civil Rights, May, 1977

Sen, B.K. and Shailendra, K., "Evaluation of Recent Scientific Research Output by a Bibliometric Method", Scientometrics, Vol. 23, No. 1, January, 1992, pp. 31-46

Sherwin, C.W. and Isenson, R.S., "First Interim Report on Project HINDSIGHT", Office of the Director of Defense Research and Engineering, Washington, D.C., October 31, 1966

Stahl, M.J. and Steger, J.A., "Measuring Innovation and Productivity - A Peer Rating Approach", Research Management, Vol. 20, No. 1, January, 1977, pp. 35 - 38

Takei, Fumio, "Evaluation Method for Engineering Activity - One Example in Japan", IEEE Transactions on Engineering Management, Vol. EM-28, No. 1, February, 1981, pp. 13 - 16

Takei, Fumio, "Evaluation Method for Engineering Activity Through Comparison With Competition - Four Years Experience", IEEE Transactions on Engineering Management, Vol. EM-32, No. 2, May, 1985, pp. 63 - 70

Vollmer, H.M., "Evaluating Two Aspects of Quality in Research Program Effectiveness", in Research Program Effectiveness, M.C. Yovits et al, eds., New York: Gordon and Breach Science Publishers, Inc., 1967

Wallmark, J.T. and Sedig, K.G., "Quality of Research Measured By Citation Method and By Peer Review - A Comparison", IEEE Transactions on Engineering Management, Vol. EM-33, No. 4, November, 1986, pp. 218 - 222

Walton, Eugene, "Gauging Organizational Health - A Questionnaire Study in a Government Laboratory", IRE Transactions on Engineering Management, Vol. EM-8, No. 4, December, 1961, pp. 201 - 205

Wolff, Michael F., "How Am I Doing?", Research-Technology Management, Vol. 33, No. 3, May-June, 1990, pp. 9-10

Stargate Consultants Limited (now Clarke-Reavley Consultants)

1687 Centenary Drive

Nanaimo, B.C. Canada, V9X 1A3

Tel: (250) 755-3066

teclarke1@shaw.ca