Evaluating Program and Project Performance – Some Lessons from the Canadian R&D Sector[1]

Chris Jones CA MBA CMC[2]

Abstract

Program evaluation and performance measurement involve the rigorous and independent assessment of the performance and effectiveness of discrete government programs, projects and activities. Every evaluator relies to a greater or lesser degree on the careful assessment of tangible[3] and intangible[4] data to arrive at conclusions about the effectiveness of a program and the nature of its impacts. Automated assessment approaches are rarely a feature of arriving at conclusions.

When available, tangible data is critical for evaluating programs, but one of the inherent problems in program evaluation, particularly in the assessment of outcomes, impacts and effects is that there are some aspects of performance where the numbers cannot give you the full story. They can only be measured by reference to the *intangibles*. The evaluator's credibility often relies on such subjective judgements, which may or may not be the "right" ones, and they may or may not be very accurate. To counter this, evaluators frequently use the opinions of subject matter experts.

Harnessing the views of experts in a rigorous, explicit and objective way is never easy. It was recognised that Canada's scientific research community had a similar problem in evaluating the likely effectiveness and success of research and development projects and programs. Here, virtually all the available performance criteria are “intangible” and experts (usually working in panels) make the key decisions. ProGrid® harvests the expert opinion in a disciplined, rigorous and scientific way. While its methodology is generic, the performance criteria are very specific to the user’s circumstances and it is completely transparent to clients - and to stakeholders. The Government of Canada, three Canadian provincial governments and several non-profits now use ProGrid® for project evaluation - leading to grants and investments of some $2 billion in expenditures every year.

Could evaluators use the methodology for the general evaluation of intangibles where expert opinions are solicited? The answer is “yes” and some ProGrid® users are beginning to use it for post-project evaluation and broader program evaluation applications. Its use in program evaluation is in its infancy, but it is an excellent example of technology transfer from one discipline to another - and a significant tool in the practitioner's 21st century armoury.

---

[1] This paper was presented in abbreviated form to the Annual Conference of the Canadian Evaluation Society, Vancouver BC, June 2003.

[2] Chris Jones is President of FORUM Consulting Group Ltd., a Victoria-based firm of management consultants that specialises in the evaluation and assessment of government programs, and Partner in FORUM Decision Systems. He may be reached at forum@acncanada.net.

[3] Tangible - "of definite shape, corporeal, material".

[4] Intangible - "incapable of being perceived by touch; incorporeal, such as knowledge, skills, goodwill".

FORUM Consulting Group Ltd. and FORUM Decision Systems

FORUM Consulting Group is a management consulting firm based Victoria, BC, Canada, that has specialised in program evaluation, policy analysis and performance assessment for federal and provincial government clients since 1987. FORUM has conducted some 250 professional engagements for governments, most of which were evaluative by nature, and often in highly technical fields. The firm specialises in “first time” evaluations of new programs.

In 2002, FORUM joined with Victoria-based natural resource science/R&D consultants, R Keith Jones and Associates, to form FORUM Decision Systems. FORUM Decision Systems is an affiliate of ProGrid Ventures Inc., owners and developers of the ProGrid® methodology.

Background

Governments use discrete initiatives to implement their policies through *programs* and *projects*, using internal or external resources. A program may be "deconstructed" into its separate components, each of which has its own set of goals and objectives. Together the components combine to achieve the program or project objectives. Good program design demonstrates consistent and realistic linkages between an *agency's corporate and policy objectives* and the *objectives of the individual components* through which these program objectives are achieved. The linkages may be depicted in a "logic model" which is often the starting point for the evaluation of a program or an assessment of program performance.

Measurable program impacts take time to develop. Therefore, the impacts and effects of the overall program itself may themselves be largely intangible and may not be easily discernible at the time a program is evaluated. The evaluator relies on the careful examination of the program logic, followed by the measurement of selected components or projects undertaken or funded by the program. Evaluation findings are then used to provide an indication of the program's overall impacts and effects, and the effectiveness of its direction, compared to the original requirements, goals and objectives.

However well program components appear to result in the achieving their *individual* objectives, there is sometimes difficulty in assessing the extent to which, in combination, they worked together to achieve the *program objectives*. Stakeholder interviews, expert panels, focus groups, case studies and surveys may be used to address the achievement of objectives, with varying degrees of success and rigour. Many of the evaluations we have seen, and indeed some of those we have ourselves performed in the past, defy academic rigour, and so must leave some questions as to their credibility.

The Traditional Approach

Traditional approaches to program evaluation include document reviews, surveys, interviews, case studies, focus groups, site visits and the like. Given an unlimited budget, these approaches taken in combination and applied with proper diligence can provide excellent results. However, our clients never have unlimited budgets and we invariably find that, given budget restrictions, certain parts of the evaluation present more difficulties than do others. This may not be a major problem as long as the client is fully aware of any areas of weakness and is able to make subsequent programming decisions accordingly. However, it is not the most satisfactory solution for either the client or the evaluator.

Some Difficulties with the Traditional Approach

There are several difficulties with the traditional approach to performance measurement and evaluation. Some of them are as follows:

Difficulties in Measuring Intangible Benefits

Government program performance must be judged by reference to the intangibles, because that is how program investment decisions are often made[5]. For example, if a project reports some positive impacts, were they better or worse than could be reasonably be expected? Would they have occurred anyway? Because the benefits (which nevertheless may be very real) are so often intangible, it is difficult to measure them. "The accountants can't count them" and rarely is there any attempt to design a framework of evaluation criteria that will assess the intangibles in a program or project in a disciplined, rigorous, objective and efficient way.

Reliance on Subjective Opinions

The reported degree of success or failure may depend on the subjective opinion of a variety of supposedly objective expert contributors, but none of them are without their own biases. Perhaps more important, the evaluator may also develop a biased view with regard to intangible issues, even after gathering such evidence as may be available from the data collection process. Evaluators are much better at reporting "what happened" or "what happened as a result" than they are at evaluating whether or not this reflects good, bad or indifferent performance. The natural question is "compared with what?" Usually there are few standard comparisons to be made.

Missing the Most Worthwhile Benefits

Some government programs (for example R&D programs, government support for new businesses or industries and financial support for experimental projects) are inherently "high-risk". Projects within them are not homogenous and the success of a small number of high-impact "winners" may be offset by the failure of a larger number of "losers". This is also true of some social programs - but is not in itself a reason for "bad marks".

The selection of discrete projects or components for examination and measurement has the potential to distort the results of the evaluation if not done rigorously and correctly. However, there may be no tangible evidence to ensure that the selection of projects is done correctly. For example, unless the projects are generally homogenous, the selection of projects by random means can miss those very projects with the most positive impacts that make the whole program worthwhile. Some significant and worthwhile benefits may never be disclosed by the evaluation - and neither will the fact that they are not disclosed.

Selection of "Representative" Projects for Case Studies and Detailed Review

The "case study" approach to evaluation and performance measurement typically relies on the selection of a representative number of the "right" projects to evaluate. Once examined, those selected may turn out not to be typical of the whole and so the results obtained from them may or may not be representative of the program. If they are not typical of the program, the wrong conclusions may be drawn and again this fact may not be properly disclosed in the evaluation.

A Common Need

What is there in common about these shortcomings with the traditional evaluation approaches? *It is the absence of a rigorous means of taking the intangibles into account in the assessment of program components*.

[5] It has been said that 90% of government program investment decisions rely mainly on the assessment of intangible factors.

What is lacking is a rigorous, efficient, transparent and *explicit* means of taking the intangibles into account in evaluating program outcomes, in choosing approaches, and in selecting and evaluating projects and initiatives within programs. Although absolute objectivity may still be elusive, the process must be capable of incorporating a significantly greater degree of objectivity and efficiency than is often the case today. Above all, the process must be explicit and the results completely transparent.

The Assessment Tool

Introduction

ProGrid® was originally developed to assess investment decisions in Canada's technology-intensive industries. A range of ProGrid-based evaluation and impact assessment tools has since been developed, often by clients themselves. Because it addresses the intangibles effectively, it represents a major step forward in assessment methodology for government service delivery, performance measurement and evaluation.

Intangibles include systems, processes, technology, leadership, relevance, relationships, values and special knowledge. By definition, they are difficult to measure. If managers had all the information they needed at their fingertips, there would be no such things as "intangibles". Nor would there be any need to measure them in a special way. Decisions could be made more readily and the risks of error sharply reduced. However, the real world is not like this, especially in government. Decisions often have to be made with an incomplete knowledge of the intangibles involved. Moreover, they have to be properly accounted for.

Governments operate through *programs*[6]. What tangible assets does government have as a basis for its ability to satisfy its customers? Not many. Governments do not use factories and rarely produce goods. The tangible physical assets used in the delivery of program services, such as land, buildings and vehicles, are often quite incidental to its main purpose. A government program itself is essentially a collection of intangibles brought together to meet an identified need. In fact, governments and agencies deliver services to citizens through a process that relies almost entirely on effective and efficient choices involving intangible assets.

Systems have been used for deciding between alternatives based on intangible factors, but they tend to be complex and unwieldy or simple and ineffective. ProGrid is a disciplined, rigorous and objective approach to measuring those intangible values which "the accountants can't count". It relies on a unique methodology and software tools applied under the direction of the evaluator and customised to first define and then measure the intangibles inherent in the evaluation process. The criteria it uses are completely explicit.

ProGrid does not make decisions for you. However, it *does* identify which of the various options best meet the criteria that have been pre-determined by your evaluators, in conjunction with the management team.

History

ProGrid was designed and developed by professional research and development portfolio managers and scientists in the Canadian aerospace industry. The developers had many years' experience in managing multi-million dollar R&D budgets and in choosing among competing projects and research strategies. They recognised that most of their decisions had to be made

[6] A set of grouped activities, normally using dedicated resources, designed to achieve defined, measurable objectives consistent with government's purpose.

largely based on intangible values such as technology, relevance, future commercial potential and management skills. However, they were concerned by the inability of the available decision support methodologies to evaluate them. There were rarely any physical, tangible assets available to be measured.

The first version of ProGrid was developed and commercially launched in the early 1990's

Users

Intangible criteria are more common in government than they are in the private sector, so this is where ProGrid has been used the most to date. ProGrid's list of users includes BC's Forest Innovation Investment, the Canadian Foundation for Innovation; the Alberta Heritage Foundation for Medical Research; the Ontario Ministry of Natural Resources; Natural Resources Canada; Defence Research and Development Canada; the Ontario Forest Research Institute; Centre for Research in Earth and Space Technology; Ontario Research & Development Challenge Fund, Alberta Department of Innovation and Science, Alberta Agriculture Research Institute, Alberta Energy Research Institute, Alberta Science Research Investment Program; Alberta Agriculture Value Added Corporation; Alberta Ingenuity Fund, Alberta Agriculture and Rural Development; Climate Change Central and the Alberta Research Council. The Science Council of BC became BC's first ProGrid client in 2000 when it used ProGrid to help Forest Renewal BC to select $14 million of forestry research projects from an initial list of over 250 projects totalling over $100 million.

ProGrid was the single commercially available technology selected for presentation at the Conference Board's 2001 "Innovations Canada" conference.

Applications in Evaluation and Performance Measurement

ProGrid may be used in a variety of applications in program evaluation and performance measurement, including:

Evaluation Design Phase

- Identification, definition and prioritisation of the key performance issues to be addressed;
- Identification and characterisation of expected project impacts and the performance measures related to them;

Project Evaluation Phase

- Selection of projects to be reviewed or evaluated;
- Setting consistent reporting objectives on how individual projects met their expectations;
- Soliciting and evaluating opinions from program users, stakeholders, beneficiaries or experts;
- Selection of case study projects and measurement against rigorous and consistent evaluation criteria.

Program Impact Assessment Phase

- Definition of the criteria to be used to assess impacts;
- Collection and analysis of expert opinions on program or project impacts;
- Reporting the results in graphical form.

Features

Evaluation and Audit

One of ProGrid's most important features is its ability to document and demonstrate in a graphical form the ability of a program or project to meet the performance expectations specified by the agency that funded it. In addition to its decision support features, ProGrid provides an enhanced medium of accountability, complete transparency, a permanent audit trail for future reference, clear documentation for the selection decisions and a medium to communicate the results. Once an agency uses ProGrid to support its assessments, the evaluation process loses much of its complexity, opacity and "mystique" - and often some of its cost.

Data Base Capability

ProGrid has an optional data base capability that holds data permanently on a year by year basis. This enables comparisons to be made of the performance and impacts of program components on an historical basis, for example to monitor improvements made as a result of previous evaluations, to monitor improved results in long term projects, or to compare the expected performance of new projects with the actual performance of older projects. This provides solid evidence for cancelling old projects if the scarce resources are better spent on new ones.

Some Other Applications in Government

As well as performance assessment and evaluation, ProGrid has other actual and potential applications in government.

Applications for funding to governments and their agencies

This is the most common use of ProGrid to date. The values, priorities and expectations of the funding program are established and signed off by the senior managers responsible for the program from their interpretation of the program guidelines approved by the legislature or by the funding authority. The criteria are expressed as a ProGrid Performance Criteria matrix.

ProGrid forces the articulation and publication of the decision criteria. Proponents are asked to rank their applications according to the criteria, and reviewers then validate the assessment using standardised ProGrid assessment forms. The data is digitised and displayed using ProGrid's charts, graphs and assessment tools. Management teams then use the ProGrid reports for choosing the best of the competing proposals. One client was able to choose a few "winners" out of over 1000 legitimate proposals in a few days.

Because each proponent receives a comprehensive report on his/her proposal evaluation, complaints and appeals are virtually eliminated.

Government procurement - outsourcing and professional services contracts

Unlike commodity procurements, outsourcing and professional service contracts can rarely be awarded based on cost alone. The proposals are essentially a set of qualitative statements which, taken together, attempt to meet the agency's contracting or outsourcing objectives. Except for the cost, these statements are usually *intangible*, and there is no objective way of measuring and comparing them, certainly not by using of the simple spreadsheet approach commonly used in government agencies.

The "right" choice between the available proposal submissions is critical because the successful bidder is actually implementing government policy initiatives by different means. The correct

choice depends on the extent to which they meet the requirements and expectations of the agency.

ProGrid may be used to select the winners of contracts to implement government policy initiatives through outsourcing and professional services proposal calls through government RFP solicitation procedures. ProGrid selection methodology offers clients a new and better way to solicit higher quality proposals, select the winners, reduce the inherent risks, avoid controversy and strengthen partnership and accountability relationships with suppliers. The standardisation of the proposal (RFP) process and the transparency of the decision making process also make life much easier for the proponents.

Appointing and promoting staff on a merit basis

One of ProGrid's government clients has used it for several years to make R&D project investment decisions and to evaluate their results. It occurred to the client that, since the performance criteria for their research staff were similar, a ProGrid-based approach could be used for simple and effective staff performance appraisals, career development planning and identifying expertise gaps. The provincial government HR authorities recently approved the methodology for staff performance appraisal.

Using consistent criteria for decentralised decision making

Government agencies sometimes need to use consistent criteria all across their jurisdictions for a variety of decision in different locations. This is difficult for head office to control and co-ordinate. ProGrid will help to design the criteria in the first place, and then help to ensure they are consistently applied and the results consistently reported for oversight purposes. ProGrid will ensure that a decision choice made in one region would be the same in any other region, but regional and local criteria can easily be built into the ProGrid application.

Steps in the Assessment Process

Overview

The principles of accounting have remained unchanged since they were first established five hundred years ago, despite the huge changes in the methodology by which they are applied. Likewise, the principles of the ProGrid decision support process are exactly the same as in any other decision making process. However, as with accounting, the principles are applied with more rigour and inclusiveness and they result in better, quicker and more actionable information.

The typical steps in a ProGrid implementation are as follows:

1. Determine the key performance criteria (usually from three to twelve) – and have them approved by the agency responsible
2. Set out the criteria out in a Performance Matrix format - and have them approved by the agency responsible
3. Design the Language Ladders™ - and have them approved by the agency responsible
4. Collect the data (may be reports, case write-ups, proposals etc)
5. The program/project managers evaluate their performance on standardised assessment forms using Language Ladders™
6. The experts evaluate the data independently on standardised assessment forms using Language Ladders™.

7. ProGrid® software assesses the performance based on the experts' assessments and ranks the projects/initiatives in order of their ability to meet the performance criteria
8. The experts meet to review their individual and collective assessments. The initial assessments are flashed up on the screen. They do a final performance evaluation strictly according to the criteria
9. Standardised format, limited text reports, graphs and charts are prepared for managers, stakeholders and anyone else to whom the program is accountable.

As with accounting, this process can be undertaken manually or on a spreadsheet, and in most organisations, it usually is. However, volumes, complexity, urgency, sensitivity, inclusiveness and the increased need for accountability make this more difficult every day. The ProGrid software facilitates the process and simplifies the assessment procedure enormously. Most importantly, ProGrid can produce performance measurement metrics for later use as needed.

Develop the Performance Criteria

Each customised ProGrid application has a distinct set of evaluation criteria that incorporates the values, priorities and expectations of the organisation that has the responsibility for funding, investing or managing the program. Designing these criteria may be a complex task but it is crucial. The matrix below used in a recent client application but it has been generalised. It focuses on the three "overarching" criteria shown in bold.

Performance Criteria Matrix Example		
How relevant was this project to our corporate objectives?	**How effective was the management and implementation?**	**What was the impact on our client population and stakeholders?**
Linkages with our organisation's strategic objectives	Quality of project strategy and design objectives	Delivery of committed services to clients (outputs)
Fit with recognised knowledge or service gaps	Method of implementation and service delivery	Appropriate and positive outcomes
Advance on current situation (incrementality)	Effective and efficient use of resources to be contributed by this agency	Absence of unintended negative effects
Development of new delivery partnerships and collaborations	Cost benefit relationships	Support to policy makers and service practitioners

Design the Language Ladder Statements

Once the overall criteria have been determined, a series of "Language Ladder Statements" is designed for each cell in the matrix – twelve in the above example. The following example shows the Language Ladder statements for the above "*Quality of project strategy and design objectives*" cell. The statements are mutually exclusive and uniformly separated. Because they are discrete, they are easily answered by the expert panel of assessors.

Language Ladder Example (Select only *one* of the four statements)	
Matrix cell: "*Quality of project strategy and design objectives*"	
This concept had never been tried before this application	A
The key elements of the concept had been confirmed and validated, but an integrated approach including all of them had not been developed prior to this implementation	B
An integrated approach had been developed and tested in the field before implementation	C
An integrated approach had been developed, tested and applied over the full range of relevant applications	D

The assessor selects *one* of the statements on the Language Ladder, and then repeats the process for each of the other cells. As long as the criteria defined in the cells and Language Ladder statements are well designed, the reviewers need not worry about any conditional "noise factors" which may be inherent in the decision.

ProGrid now has a wide selection of Performance Criteria and Language Ladder sets which have been used successfully in client applications. They can be made available to new clients as a basis for developing their own.

Rank the Alternatives - the Assessor Meeting

Most clients have their expert reviewers review the results (or the project proposals) independently and then meet afterwards to compare their findings. ProGrid uses the Language Ladder assessments made by the expert panel to provide an overall assessment. The assessments are digitised and the software uses special routines to assess the alternatives according to the pre-determined criteria. Reports, which may be in the form of a ranking list, are generated for the assessor meeting.

When they come together to discuss their assessments, the key features and the reviewers' assessments are flashed up on the screen. One of the screens indicates the degree of consensus between reviewers for each individual criterion. Where a high degree of consensus already exists, the panel may choose to move on. Discussion is focussed on criteria or projects where little or no consensus exists and changes to the initial assessments may be made as a result. Either way, since all the key information is at hand and the process is shown on the screen consensus is achieved much more quickly than with other methodologies.

Reports – Charts and Graphics[7]

ProGrid's outputs are in charts, graphics and tables. The approach produces a range of client-selected charts and graphs. For example:

- Rankings against the client's established performance criteria;
- Ranking of options in order of their ability to meet the criteria;

[7] Some of ProGrid's standard reports appear in the accompanying PowerPoint presentation.

- For project investment selection applications, options which fall above the "funding line" and options which don't - and specifically why;
- Profile showing how the options fared against each performance criterion;
- Historical view - comparison of criteria rankings of current applications with those of previous applications;
- Degree of consensus among the “expert panel” decision makers.

The chart showing the degree of consensus between the members of a team is particularly useful tool. One agency uses this information to compare and assess the performance of the reviewers over time.

** ** **

Conclusion

The ProGrid tool represents a new, rigorous and disciplined approach to the problem of evaluating the intangibles inherent in the assessment of programs and projects to meet government and agency evaluation and performance measurement objectives. It has application in the design, selection, evaluation and performance measurement phases of a program.

Could evaluators use the methodology for the general evaluation of intangibles where expert opinions are solicited? The answer is “yes” and some ProGrid® users are beginning to use it for post-project evaluation and broader program evaluation applications. Its use in program evaluation is an excellent example of technology transfer from one discipline to another - and a significant tool in the practitioner's 21st century armoury.