
1. Introduction to the handbook

Albert N. Link and Nicholas S. Vonortas

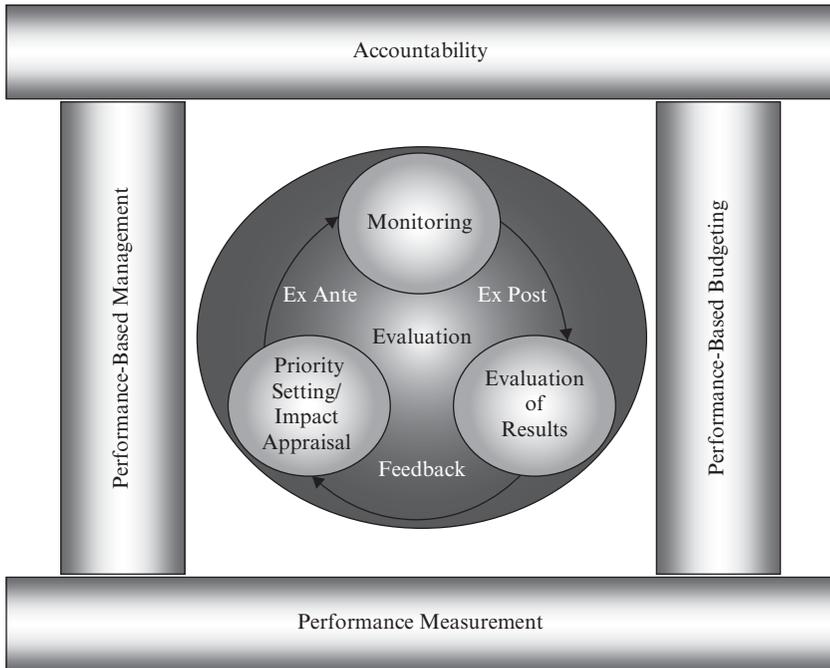
It hardly seems necessary these days to point out the importance of technical advance. We look to it to rescue us from the consequences of exhausting essential natural resources; abate inflation through productivity increases; improve our balance of payments deficit; eliminate famine; and cure cancer, heart disease, and a variety of other ailments. Our faith in technical advance is bolstered by achievements such as the atomic bomb, electronic computers, the landing of a man on the moon, heart transplants, and test-tube babies. We no longer ask if something is possible, but how soon it can be done and at what price. (Kamien and Schwartz, 1982, p. 1)

Thus two important mainstream economists opened their landmark book thirty years ago. The same year several other landmark books were published on the characteristics, organization, and economic importance of technological advancement – such as Caves (1982), Freeman (1982), Nelson and Winter (1982), Rosenberg (1982), Stoneman (1983), to mention a few. They were not even aware of major advances about to penetrate and revolutionize industry, such as genome decoding, mobile telephony and the internet. Little could they also guess about the massive changes in the global political economy that have extensively redistributed innovative activity far beyond the hotbeds of the twentieth century (North America and Western Europe) across Asia and Latin America and countries like Japan, the Republic of Korea, Taiwan, Singapore, China, India, Israel, Brazil, Mexico, Russia, and South Africa, to name the obvious examples.

In these thirty years (1982–2012), investments in public goods, research and development (R&D) investments in particular, have risen around the world so much that governments and the general public in developed and fast-developing countries are very conscious of the opportunities for growth.¹ Both governments and the general public are asking important questions of efficiency, effectiveness and accountability of significant societal investments in science, technology and innovation. In combination with increasing demands for transparency, accountability and performance, this has created a climate ripe for the evaluation of public policies and the assessment of public R&D funds.

Figure 1.1 summarizes the interconnected facets of evaluation and the use of the results in policy decision-making. The facets of the evaluation proper include:

2 The theory and practice of program evaluation



Source: Polt and Vonortas (2006).

Figure 1.1 The facets and uses of evaluation

- priority setting and *ex ante* impact appraisal;
- monitoring of progress – interim evaluation;
- evaluation of results and impacts (*ex post*).

Cumulatively, they aim at:

- measuring performance;
- supporting performance-based management and performance-based budgeting;
- enhancing accountability and transparency;
- improving the communication of program activities and outcomes to policy decision-makers and sponsors.

A wide variety of methodologies exist to evaluate the objectives, pathways, outputs and outcomes. Mainstream evaluation techniques include surveys, statistical/econometric estimation, patent analysis,

bibliometrics, scientometrics, network analysis, case studies, historical tracings, and expert judgment.² The utility of each technique, however, will vary according to the range, specificity and relative importance of the objectives set forth for the public program in question. The applicability of individual techniques will also vary with the R&D program's chronological evolution, as the anticipated technological, economic and societal impacts anticipated from the program entail different gestation periods. Thus, retrospective analyses of previous R&D programs, interviews with prospective participants and end users, and preliminary surveys can be expected to precede techniques useful for garnering data on intermediate phases and indicators of progress, such as patents and innovations, which, in turn, will precede more extensive and diverse forms of data collection and causal analysis of modes of inter-sectoral collaboration and international economic competitiveness. Table 1.1 presents an overview of these methods and a brief description of their contents.

These techniques may be used singly or in combination (mixed methods/triangulation); they may entail collection of primary data or use of secondary data; they may be based on simple linear or complex, non-linear relationships; and they may be directed at one or more of the outputs, outcomes, and impacts associated with a program's objectives. Selection of the most appropriate methodologies, singly or in combination, will depend on the specific objectives and uses for which the evaluation is designed and the audiences for whom the evaluation is intended (Polt and Vonortas, 2006).

Evaluations of large-scale, complex programs frequently employ mixed method designs. The choice of the mix of instruments, the sequence in which tools are deployed, and the ways in which elements derived from the different tools employed are merged in the overall analysis are central to the success of an evaluation (Georghiou and Laredo, 2006). Multiple techniques are combined to offset the limitations of specific approaches, to complement one another such that the findings generated by one approach (for example, surveys) are used as datasets in constructing variables used in other approaches (for example, econometric estimation), and to triangulate findings as a means of increasing the confidence that program impacts have been correctly identified and located (Polt and Vonortas, 2006).

Four inherent basic problems of program evaluation must be kept in mind (Georghiou and Laredo, 2006):

- *Timing* – the effects of research are often manifested long after the research has been completed and the connections obscured.

4 *The theory and practice of program evaluation*

Table 1.1 Overview of evaluation methods

Methods	Brief description	Example of use
Analytical/ conceptual modeling of underlying theory	Investigating underlying concepts and developing models to advance understanding of some aspect of a program, project, or phenomenon.	To describe conceptually the paths through which spillover effects may occur.
Survey	Asking multiple parties a uniform set of questions about activities, plans, relationships, accomplishments, value, or other topics, which can be statistically analyzed.	To find out how many companies have licensed their newly developed technology to others.
Case study – descriptive	Investigating in-depth a program or project, a technology, or a facility, describing and explaining how and why developments of interest have occurred.	To recount how a particular joint venture was formed, how its participants shared research tasks, and why the collaboration was successful or unsuccessful.
Case study – economic estimation	Adding to a descriptive case study quantification of economic effects, such as through benefit–cost analysis.	To estimate whether, and by how much, benefits of a project exceed its costs.
Econometric and statistical analysis	Using tools of statistics, mathematical economics, and econometrics to analyze functional relationships between economic and social phenomena and to forecast economic effects.	To determine how public funding affects private funding of research.
Sociometric and social network analysis	Identifying and studying the structure of relationships by direct observation, survey, and statistical analysis of secondary databases to increase understanding of social organizational behavior and related economic outcomes.	To learn how projects can be structured to increase the diffusion of resulting knowledge.
Bibliometrics – Counts	Tracking the quantity of research outputs.	To find how many publications per research dollar a program generated.
Bibliometrics – Citations	Assessing the frequency with which others cite publications or patents and noting who is doing the citing.	To learn the extent and pattern of dissemination of a project’s publications and patents.

Table 1.1 (continued)

Methods	Brief description	Example of use
Bibliometrics – content analysis	Extracting content information from text using techniques such as co-word analysis, database tomography, and textual data mining, supplemented by visualization techniques.	To identify a project's contribution, and the timing of that contribution, to the evolution of a technology.
Historical Tracing	Tracing forward from research to a future outcome or backward from an outcome to precursor contributing developments.	To identify apparent linkages between a public research project and something of significance that happens later.
Expert judgment	Using informed judgments to make assessments.	To hypothesize the most likely first use of a new technology.

Source: Ruegg and Feller (2003, pp. 30–31).

- *Attribution* – a given innovation may draw upon multiple research projects and a given research project may impact upon multiple innovations. In drawing pathways between them it is also the case that an innovation depends upon many inputs other than research before market or social effects are realized.
- *Appropriability* – the beneficiaries of research may not be the same people or organizations who performed it; it may not be obvious where to look for effects.
- *Skewness of results* – the distribution of impacts in a project portfolio is typically highly skewed. A small number of projects may account for the majority of effects, while a good number often just advance knowledge in a general way. This has implications for sampling strategies.

This handbook presents a state-of-the-art collection in the theory and practice of program evaluation. The following twelve chapters were written by some of the most well-known experts in socioeconomic evaluation surveying various areas of this interdisciplinary field. The methods and applications presented herein are divided into four parts: economic, non-economic, hybrid, and data-driven.

The part on economic methods and applications includes three chapters. The first of these (Chapter 2) is by Albert Link and John Scott; it summarizes the theory and practice of public-sector R&D economic analysis with

6 *The theory and practice of program evaluation*

specific reference to the National Institute of Standards and Technology's (NIST) efforts to document the impact that their in-house R&D has had on society. The authors focus on the analysis of the economic impact of R&D programs. In particular, they are concerned with the comparison of the core question of the social benefits associated with a publicly supported R&D program compared to society's costs to undertake the program. They draw on impact analyses of NIST activities, retrospective analyses in particular. Although some of the NIST impact analyses reviewed in the chapter are partially prospective in nature, the analyses are still retrospective in the sense that they examine NIST programs from an historical investment perspective.

Chapter 3 is by Spyros Arvanitis. It reviews micro-econometric approaches to evaluating public programs that promote technological advancement. Micro level means that the unit of analysis is the firm as compared to the meso level (sector, industry) or to the macro level (economy wide). The chapter focuses on micro-econometric evaluation studies of policy measures aiming at enhancing the R&D capacity of firms in general or of particular groups of firms characterized by certain attributes such as firm size, industry, type of technology, and so on. Policy support mostly takes the form of subsidies to or fiscal incentives for commercial R&D. Other policy tools include government and university research supplying the business sector with substantial R&D inputs. In addition, a second smaller group of micro-econometric evaluation studies deals with government programs supporting the diffusion of new technologies (for example advanced manufacturing technologies) in the business sector mostly for some particular firm categories (small enterprises, firms belonging to high-tech industries, and so on). The micro-econometric evaluation approaches surveyed herein are relevant to all policy interventions that target firms, for example programs that aim to foster R&D activities (tax policy, subsidies, and so on), R&D collaborations among firms, the adoption of new technologies, start-up assistance for new high-tech-firms, and so forth.

Chapter 4 is by Sébastien Gasault, Aard Groen and Jonathan Linton. It deals with the methodologies of R&D project portfolio selection. This chapter considers the current state of knowledge in project and portfolio selection and identifies the connection to the highly developed field of financial portfolio management. Current practice of R&D portfolio management typically involves attempts to group similar projects together in order to optimize resource allocation. Relationships between projects are rarely considered formally, which may lead to balancing problems in the overall portfolio that offers sub-optimal solutions to the organization's strategic goals. Decision makers require better tools to maximize program efficiency. This can be achieved by evaluating projects simultaneously

based on a combination of individual characteristics and the interaction between projects. The chapter, therefore, explores the differences between project and portfolio evaluation techniques. Existing gaps in the literature in terms of portfolio selection are identified with a brief presentation of the current state of the research and possible areas of study in the future.

The section on non-economic methods and applications includes three chapters. Chapter 5 is by Irwin Feller. It looks at what is probably the most widespread evaluation tool of all times: peer review. The terms “peer review” and “expert panels” are used interchangeably to describe procedures used to select competing proposals for funding; accept or reject manuscripts submitted for publication in (refereed) journals; judge whether an individual faculty member’s performance warrants tenure and/or promotion to a higher academic rank; assess whether an agency’s policies or programs have achieved their intended objectives; and related situations. Peer review can be used both in *ex ante* decision-making contexts – providing guidance about the likely best course of action given objectives, constraints, uncertainties and options – and *ex post* contexts – providing formative or summative evaluations of past events designed to assist decision makers to make better decisions in the future. The chapter specifically focuses on peer review procedures to review project and center proposals submitted to government agencies to support academic scientific and technological research. Emphasis is placed on the extent to which the design and implementation of peer review mechanisms simultaneously satisfy two essential societal expectations: allocative efficiency – directing public funds to their most productive use – and procedural equity – assuring that decisions are made in ways that accord with requirements or norms for political accountability and fairness.

Chapter 6, by Gretchen B. Jordan, addresses logic modeling. Logic modeling is a management and evaluation tool to develop a succinct picture of a program’s goals and the strategies for achieving these within a broader context. A logic model is a plausible and sensible model of how the program will work under certain environmental conditions to solve identified problems. The elements of the logic model are resources, activities, outputs, outcomes, and the external contextual influences on the program. The process of developing the model builds a shared understanding of the program and performance expectations as well as a short, clear performance story for those less familiar with the program such as senior managers and Congress. The primary use of logic modeling is to design program evaluations and performance measurement systems.

The third chapter in this section, Chapter 7, is by Barry Bozeman and Gordon Kingsley. They argue that despite significant progress in the ability to conceptualize, measure and evaluate research impacts, a gap

remains in research evaluation methods and technique in terms of the appraisal of the sociotechnical impacts of research. The chapter describes a case study-based approach to sociotechnical evaluation of research – Research Value Mapping (RVM) – and illustrates with applications. RVM combines qualitative approaches (the case studies) with quantitative approaches (data developed from the case studies) in an attempt to develop causal explanations of the social distance traveled by the knowledge products of research from outputs, through outcomes and impacts.

The section on hybrid methods and applications includes three chapters. Chapter 8 is by Nicholas Vonortas; it introduces the social network methodology as a tool for evaluating important aspects of R&D programs. The chapter accomplishes this by expanding on three recent examples of actual application of the methodology for evaluative purposes and, through them, observing some of the standard (and non-standard) graph theory concepts at work. The first case deals with inter-sectoral differences in innovation-related networking activity. Five industrial sectors are analyzed: pharmaceuticals, plastics, computers, electronics and instruments. Both virtual knowledge networks based on patents and patent citations and actual partnership networks based on research alliances in each sector are utilized. The second case of network methodology application presents an in-depth empirical analysis of the emergent networks created and maintained by sustained public funding in information and communication technologies (ICTs) through the European Research Framework Program (FP). The third case also deals with networks built around ICTs through European FP funding. It investigates whether the projects selected by the first two calls for proposals of FP6 (2002–04) in three ICT thematic areas have been effective in supporting network hubs that nurture global knowledge leadership and European cohesion. Three types of data are utilized: European FP IST alliance data, global alliance data in ICT and patent citation data of identified participants in these alliances. The chapter argues that social network methodology is an important tool in the analytical toolbox for evaluating the performance of publicly funded research. Its main potential lies in improving our understanding of the structural and capacity-building impacts of R&D investment.

Chapter 9, by Alan C. O'Connor, Michael Gallaher, Ross Loomis and Sara Casey, reviews categories of environmental benefits and practical strategies for estimating and presenting them. Downstream costs of fossil fuel combustion are not fully captured in market prices, providing a rationale for the public sector to mitigate negative externalities via policy and regulation, promotional campaigns and innovation programs. Such initiatives often create opportunities for the private sector by creating a more favorable risk/reward profile, accelerating new market development

and building on emerging consumer preferences. The ability to measure and monetize impacts on environmental emissions and related health benefits is of increasing interest to program evaluations. These benefits are conceptually important and in practice not insignificant in monetized value. The chapter begins with an introduction to common environmental impact metrics and then reviews how to quantify benefits from avoided energy consumption and energy security, avoided emissions, and environmental health benefits. The chapter reviews an easy-to-use tool for estimating environmental health benefits called the Co-Benefits Risk Assessment (COBRA) model. COBRA provides estimates of health effect impacts and the economic value of these impacts resulting from changes in the physical units of emitted pollutants. It concludes with an in-depth analysis of the real-world example of the environmental benefits of solar photovoltaic (PV) energy systems.

Chapter 10, by Drew Rivers and Denis Gray, describes a methodology for capturing economic impact data from private sector firms involved in cooperative research with universities. The authors focus on Cooperative Research Centers and first appraise the extant Industry/University Cooperative Research Centers' (IUCRC) evaluation strategy in addressing the program's explicit partnership and capacity-building objectives and in measuring outputs and proximal outcomes like publications and patent applications based on center research. They find the extant strategy, however, insufficient to consistently capture more distal (and likely more significant) economic outcomes derived from process, product and service innovations. As a consequence of these findings the authors examined the value of a complementary assessment strategy: confidential interviews with key informants inside firms that had been nominated as potential "high impact beneficiaries" of ideas and technologies generated by established IUCRCs. This assessment strategy proved to be successful in gathering rich information on how these centers contribute to innovation and in obtaining credible, quantitative estimates of the economic impacts of IUCRCs on program stakeholders.

Finally, the section on data-driven applications includes three chapters. The first of these, Chapter 11, is by Diana Hicks and Julia Melkers. It provides a background and surveys bibliometric approaches to the analysis of R&D. Bibliometrics – or the study of publication-based output – is a quantitative method widely applied in evaluation. The most common and accepted use is for the analysis of the output of basic research. However, bibliometrics is also useful as a partial indicator of overall R&D output, and of the productivity and impact of funded research teams and centers. Fundamental issues arise in the use of bibliometric measurements of R&D activities and are discussed herein. For example, a potentially long

time frame is needed to judge the results of research programs fairly, but funding bodies require evidence of success after perhaps two or three years. In addition, R&D is a cumulative, networked process: results most often build upon other results produced by other groups and programs. Success is often difficult to credit in its entirety to a single source, a problem if agencies need to claim ownership.

The chapter by Rosalie Ruegg and Patrick Thomas, Chapter 12, summarizes the use of patent analysis in program evaluation, indicates the research questions it can help to answer, illustrates how patent analysis is performed, and shows its limitations. Patents can be used to indicate that knowledge output has occurred, and patent citation analysis can be used to evaluate the extent to which a given research program has influenced downstream innovation and commercial development. It can also be used to find connections between research programs and other organizations that are building on their research. In addition, it can be used to identify patents that are highly cited by many later patents, signaling that they are likely to contain technological information of particular importance to many new innovations. An application of patent analysis in R&D for advanced combustion in vehicle engines by the Energy Efficiency and Renewable Energy Office (EERE) of the US Department of Energy is used as illustration.

The final chapter in the handbook, Chapter 13, is by John Jankowski. It addresses the topic of innovation surveys. The focus is on the redesigned Survey of Industrial Research and Development of the US National Science Foundation into the new Business Research & Development and Innovation Survey (BRDIS) which was sent to approximately 40 000 businesses in the United States in 2008 and for the first time included questions on business innovation. Importantly, the chapter summarizes important results from this survey and stresses differences with its European equivalent, the Community Innovation Survey (CIS). The apparently dissimilar BRDIS/CIS findings raise questions about what is actually being measured and whether the same measurement methods can be utilized across national systems and produce robust statistical results. As a direct result of these concerns, an international effort is now underway led by the OECD to review innovation survey questions used worldwide. The work plan includes cognitive testing of existing questions in English and selected languages, as well as the development and rigorous testing of new questions in areas of growing interest in the study of innovation. The intended result is a set of carefully tested questions and guidelines that can be used to extend the coverage of existing surveys, or for new surveys, to produce internationally and inter-temporally comparable indicators of innovation. This exercise is expected to improve business survey methods and,

consequently, support the public debate of the role of innovation in the national/regional economies.

NOTES

1. Gross Domestic Expenditure on R&D (GERD) in the OECD region has increased sixfold in real terms (current PPP US\$) in these three decades. In 1982 the country members of the OECD spent a cumulative \$178 573.3 million (current PPP US\$). This had risen to \$965 629.1 million in 2008 – the last year with available statistics – and has almost certainly crossed the \$1.1 trillion level by now. The US and EU15 shares of the OECD GERD have fallen from about 45.5 percent and 32 percent in 1982 to about 41 percent and 29 percent respectively in 2008. The Japanese share has also fallen accordingly. Some big spenders are now included in the organization, such as China and Korea, but also rising countries like Mexico. Other rising powers – most notably the BRICS countries, barring China, and Israel – are not included in these numbers. In other words, aggregate global expenditure on R&D has increased by leaps and bounds and its distribution around the world has changed rather dramatically.
2. For extensive surveys of evaluation techniques see Ruegg and Feller (2003), Fahrenkrog et al. (2002) and Georghiou et al. (2002). See also: Bozeman and Melkers (1993) and Georghiou and Roessner (2000).

REFERENCES

- Bozeman, Barry and Julia Melkers (eds) (1993), *Evaluating R&D Impacts: Methods and Practice*, New York: Springer.
- Caves, Richard E. (1982), *Multinational Enterprise and Economic Analysis*, 1st edn, Cambridge: Cambridge University Press.
- Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexander Tubke and Klaus Zinnocker (eds) (2002), 'RTD evaluation toolbox: assessing the socio-economic impact of RTD policies' (EPUB), European Commission, Joint Research Centre.
- Freeman, Chris (1982), *The Economics of Industrial Innovation*, 2nd edn, London: Pinter.
- Georghiou, Luke and Philippe Laredo (2006), 'Evaluation of publicly-funded research: recent trends and perspectives', Report DSTI/STP(2006)7, Paris: OECD.
- Georghiou, Luke and David Roessner (2000), 'Evaluating technology programs: tools and methods', *Research Policy*, **29**, 657–78.
- Georghiou, Luke, John Rigby and Hugh Cameron (eds) (2002), 'Assessing the socio-economic impacts of the framework programme' (ASIF), European Commission, DG Research.
- Kamien, Morton I. and Nancy L. Schwartz (1982), *Market Structure and Innovation*, New York: Cambridge University Press.
- Nelson, Richard R. and Sidney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- Polt, Wolfgang and Nicholas S. Vonortas (2006), 'IST evaluation and monitoring', Final Report, European Commission, DG Information Society and Media.
- Rosenberg, Nathan (1982), *Inside the Black Box*, Cambridge: Cambridge University Press.
- Ruegg, Rosalie and Irwin Feller (2003), 'A toolkit for evaluating public R&D investment', NIST GCR 03-875, Gaithersburg, MD: National Institute of Standards and Technology.
- Stoneman, Paul (1983), *The Economic Analysis of Technological Change*, Oxford: Oxford University Press.

