EVALUATION PRACTICES IN A MODERN CONTEXT FOR RESEARCH: A (RE) VIEW

Emilio Muñoz, Juan Espinosa de los Monteros and **Víctor M. Díaz**
Unidad de Políticas Comparadas, CSIC, Madrid, Spain

Keywords: science policy, accountability, evaluation methodologies, R&D programmes, knowledge production, knowledge impact, knowledge "modes", evaluation models, biology metaphors, functional evaluation, societal evaluation, comparative perspectives, Europe-USA.

Contents

Summary

The increasing relevance of science (and technology) for development has raised the interest in the process of evaluation of research and its impact. The concept of research as well as evaluation practices has been experiencing an evolutive process leading to the merge with other concepts – technology, industrial development, innovation.

The influence of European institutions has been decisive in fostering the application and development of evaluation methods and practices to public policies, and among them to research and innovation policies. Some of these forms of evaluation are presented in relation to the different types of research programs and activities. Some innovative approaches based on new models have been developed to take into account either the role of research laboratories (the 'compass-card' model), the societal quality of research

or the functional performance of research programs (the 'transducing model'). Some case studies are discussed to illustrate these approaches.

The comparative perspective with respect to the situation of research evaluation in the United States reveals the dynamics of that process with respect to the increasing need for accountability of science towards society.

1. Introduction

Science and technology are instruments of increasing strategic value in society for the attainment of wealth and for enabling competitiveness in a globalized world. Changes in the economic context influence the performance of science and technology. Important social changes are occurring in the developed world in the transition from a production to a services society, which open new challenges in several areas of science and technology. An important consequence of this situation is the changes taking place in the design and assessment of science policy. The principle articulated by Vannevar Bush that basic research and its practical benefits 'accrue to society through an apparently unrelated process' is questioned every day. The performers of basic research have to demonstrate user relevance, relationships with industry or the utility of their work.

Evaluation has become a critical factor since government sponsors are calling for greater accountability on the part of researchers and they are looking for funding research areas and projects of strategic relevance. In spite of this trend, evaluation is still showing its limits as the indicators and methods applied to evaluate science and technology outputs and their effects on innovation, are still based mainly on the concept of linearity implied by the principle mentioned by V. Bush. The bibliometric methods used to measure scientific and technological production, i.e. scientific articles and patents, concentrate on productivity. The economic and human resources devoted to science and technology are viewed as inputs into the system. This output/input model of evaluation is leaving aside an assessment of the interactions between science, technology and innovation and their actors. Such a model, inspired by econometrics, does not underscore the role of human players in the process of generation and use of knowledge, as well as the influence of cultural values and the environment on it. However, citations and referencing data have been used in some fields like biotechnology and biosciences to detect the links between science and technology.

In the present review, we describe the trends orienting the (new) science and technology policies and report the efforts undertaken by different countries and institutions to adapt the evaluation practices to this new context.

2. Basic Definitions of Research and Evaluation Methodologies

The concept of research is as old as science, but has experienced an important evolutionary process as it merged with other concepts, such as technology, industrial development and innovation. This combination took place when it was recognized, after the two World Wars, that research and development are essential in an industrial production system.

Therefore, **research and development** (R&D), a concept unheard of until the fifties, became a universal watchword in developed countries. Consequently, a series of indicators providing some grounds for assessing research performance and development needed to be established. Moreover, the efforts to promote R&D activities were shared by different institutions: government, business, and higher education, in such a way, that these three sectors were incorporated into the statistics which accounted for expenditure efforts (economic and human resources) and so were their results (outputs such as scientific production and patents).

The way in which R&D performance was understood and measured, came from an interpretation of the intimate relationship between research and subsequent development and how they intervene in a pipeline leading to industrial production through innovation, starting from a laboratory discovery, through prototype production or manufacturing start-up, to full-scale production and market introduction. However, the difficulties to correlate the R&D efforts in a direct and simple manner with the innovation practices and market successes has led to a shift from the linear interpretation of the R&D influence on industrial production to a more systemic one.

Evaluation practices experienced a similar evolutionary process, starting with the traditional 'peer-review' system applied by the scientific community to assess research projects, to the need to evaluate the strategies of integrated goals through programs usually designed and funded by governments.

In order to be able to approach a review of the different evaluation practices and their influence on the different instruments used to promote research and development it seems logical to point out some definitions and to check them against the evaluation methods currently used for each one.

Innovation might be defined as the application of an invention at any stage of the production process, either in the technical or in the organizational patterns, addressed to a significant market need. Innovation outcomes are difficult to measure and the current trend is to assess the innovation capacity of a firm through a survey which employs quantitative and qualitative methods.

Research is the process of careful, focused inquiry, frequently carried out by trial and error. As a result of the evolutionary process of understanding and measuring the research procedure, several distinctions have emerged throughout the 20th century.

Basic research is defined as the research process carried out by scientists and collaborators who lack a conscious goal, other than the aim to unravel nature's properties and components. Basic research has sometimes been referred to as 'pure' research, but this term misses the fact that a research program's objective might be to address issues on technology, or problems that may be of interest to the government (health, environment) or a given industry (information technologies, pharmaceutical laboratories).

Usually, basic research is funded through calls from the public sector and the funds are allocated following the evaluation of the projects presented by the researchers. This

process of '*ex-ante*' evaluation is carried out by experts, the 'peer-review' system, performed anonymously and free of external influences. The acceptance of the idea that research is influencing economic and social development of the nations has led to some reshaping of the 'peer-review' system. In the European Union research, the peer-review system is applied with some modifications: a) the individual evaluations are carried out independently, but in the Commission headquarters; b) these evaluations are contrasted and discussed jointly with a panel made up of three or four independent evaluators who have previously assessed the project; this leads to a 'consensus' evaluation. In the United States, some agencies, such as the National Health Institutes, are considering including some non-experts into the review process. The debate remains open on the pertinence and qualification of these modifications of the peer-review system.

Basic research has been unable to escape from the increasing demand of accountability. The outcomes of this type of research are scientific publications. Bibliometrics has evolved as the methodology chosen to carry out 'ex-post' evaluation of these research activities (see *Bibliometrics and Institutional Evaluation*).

Applied research addresses the results of basic research to a point where they can be used to meet a specific need or, alternatively, it attempts to solve a specific problem through research. Applied research can be funded in different ways: a) as in the case of basic research, by calls from public agencies which base their funding decisions on the expert evaluation report. These calls may be part of independent research projects or of a research program; b) by calls from public agencies related to the attainment of specific objectives of a research program; c) from industrial requirements through contracts.

'*Ex-ante*' evaluation of applied research projects from public calls follows the same pattern as basic research evaluation, although the need for reshaping the 'peer-review' system appears more obvious for this type of research. As a matter of fact, even critics of this reshaping in the case of basic research, may accept it for this type of research activities. The research carried out by industries may be evaluated by experts, internal or external to the companies, depending on the firms' strategies. The current trend is to apply evaluation procedures in most cases as companies are becoming more familiar with the practices followed by the scientific community.

The '*ex-post*' assessment of the outcomes of applied research may be supported by combining the measures of bibliometric outputs with the information from surveys addressed to the stakeholders whose interests should be met by the research in question. A pressing problem stems from how to include, among the merits of researchers from the public system, the results of the research they have carried out while under contract with industries.

The development stage of R&D refers to the steps required to bring a new or modified product into the market. There are no specific methods to evaluate this part of the R&D process, except in what is related to aspects of dissemination and utilization of the results of the research process. In general, this aspect of the R&D fits the philosophy of a research program better than that of isolated research projects.

Research Program has emerged as a way of gathering the tools necessary to attain specific goals through research. A research program usually reflects an agency's strategy, either public or private, to fulfil the aims of a given policy. There are several ways in which a research program can be launched, and each of these can be evaluated 'ex-ante' by different mechanisms, essentially resorting to experts resources. The strategic value of a *'research program'* and its complexity, as well as its relevance for the implementation of scientific and technological polices, has raised the interest for the 'ex-post' evaluation of research programs and their outcomes. The idea of 'strategic research' has co-evolved, along with the concept of 'research program'. The notion of 'strategic research' was coined in order to qualify research activities dedicated to providing knowledge and/or techniques for solving problems of (social or industrial) relevance.

3. Relationship between Science Policies, Promotion and Management of R&D Activities

At different times throughout the history of science and technology, several types of science policy models, which emphasize different criteria and methods for the evaluation ('ex-ante' and 'ex-post') of research activities.

The different types of science policy present throughout the second half of the twentieth century can be summarized as follows.

3.1 Policy for Science

Policy for science began after the Second World War with support from political and expert leaders of the United States (President Roosevelt and engineer Vannevar Bush). This policy was essentially aimed at fostering research through government funding. This aim was achieved through research projects and public calls. Evaluation was strongly rooted on the 'peer-review' system on the basis of the principle of 'self autonomy' of the scientific community.

-
-
-

Bibliography

General references

Bührer S. and Kuhlmann S. (eds) (1999). *Evaluation of Science and Technology in the New Europe*, Proceedings of an International Conference on 7 and 8 June, 1999, Berlin, 209 pp. German Federal Ministry of Education and Research, European Commission. [Reviews the role and function of evaluation

at various levels of decision making in relation with the changes of the S&T systems in the wider European context.]

Callon M. Laredo P. and Mustar P. (1997). *The Strategic Management of Research and Technology: Evaluation of Programs*, 445 pp. Paris: Economica International. [Contains relevant information provided by top scientists on the links between science, technology and the market and on how to evaluate those links and the resulting hybrid activities.]

Gheorghiu L. Dale A. and Cameron H. (guest-editors) (1995). National systems for evaluation of R&D in the European Union. Special issue of *Research Evaluation* **5**(1), 1–108. [Contains relevant information on the different evaluation systems in European member countries as an outcome of the activities of the Network on European Science and Technology Policy Evaluation.]

Report of the Advanced Science and Technology Policy Planning Network (ASTPPN), authored by Kuhlmann S. Boekholt P. Georghiu L. Guy K. Heraud J. A. Laredo P. Lemola T. Loveridge D. Lukkonen T. Polt W. Rip A. Sanz-Menéndez L. and Smits R. (1999) under the title '*Improving Distributed Intelligence in Complex Innovation Systems*', 92 pp. Karslruhe: Fraunhofer Institute Systems and Innovation Research, ISI. [Discusses the need for a synergy of the combination of the different tools used for strategic decisions in innovation policy: evaluation, technology assessment and technology foresight.]

Specific References
On framework and conceptual aspects

Gibbons M. Limoges C. Nowotny H. Schwartzmann S. Scott P. and Trow M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, 177 pp. London/Thousand Oaks/New Delhi: Sage Publications. [Analyses the modes of knowledge production and identification of the distinctive characteristics of 'mode' 2.]

Gilbert L. (1998). Disciplinary breadth and interdisciplinary knowledge production, *Knowledge, Technology and Policy*, Spring/Summer 1998, **11**(1&2), 5–15. [Introduces definition of disciplinarity and interdisciplinarity in relation to the production of knowledge.]

Laredo P. and Mustar P. (1996). Laboratory configurations: an exploratory approach, 23 pp. Presented in *EASST/4S Conference: Signatures of Knowledge Societies*, Bielefeld 10–13 October. [Aims to delve into the characterisation of laboratory activities in the frame of the Compass Card model.]

Muñoz E. (1994). *Una Visión de la Biotecnología: Principios, Políticas y Problemas*, 145 pp. Madrid: Fondo de Investigación Sanitaria. [Discusses some concepts of the philosophy of biology that may help in the evaluation of the applications of life sciences.]

Sarewitz D. (1996). *Frontiers of Illusion: Science, Technology and the Politics of Progress*, 235 pp. Philadelphia, PA, USA: Temple University Press. [Offers critical insights on the myths that have guided the science policy and presents new challenges that are being confronted by the science and technology enterprise.]

On empirical and case studies

Bell J. I. (1999). Clinical research is dead; long live clinical research, *Nature Medicine*, **5**(5), 477–478. [An analytical exploration of the relations between basic biomedical and clinical research.]

Bloom F. E. (2000). Unseemly Competition. *Science*, **287**(5453), 28 January, p. 589. [Presents US policy to deal with the problems of scientific misconduct and the limits of the evaluation by a peer-review system.]

Espinosa de los Monteros J. Martínez F. Toribio M. A. and Muñoz E. (1994). *El Programa Nacional de Nuevos Materiales en el periodo 1988-1992. Su evaluación mediante una metodología dual.* Documento de Trabajo 94-10, 86 pp. Madrid: IESA, CSIC. [The first of a series of evaluation of Spanish National R&D programs by a blend of quantitative and qualitative methods.]

Espinosa de los Monteros J. Martínez F. Toribio M. A. Muñoz E. and Larraga V. (1995). *El Programa Nacional de Investigación y Desarrollo Farmacéutico durante el periodo 1988–1993. Una evaluación mediante una metodología dual.* Documento de Trabajo 95-08, 74 pp. Madrid: IESA, CSIC. [The application of the hybrid methodology to the evaluation of the Spanish National Programme on Pharmaceutical R&D.]

Espinosa de los Monteros J. Martínez F. Toribio M. A. Muñoz E. and Larraga V. (1995). *El Programa Nacional de Salud durante el periodo 1989–1993. Una evaluación mediante metodología dual.* Documento de Trabajo 95-09, 71 pp. Madrid: IESA; CSIC. [Extension of the methodology to the Spanish National R&D Programme on Health.]

Espinosa de los Monteros J. Larraga V. and Muñoz E. (1996). Lessons from an evaluation of Spanish public – sector biomedical research, *Research Evaluation,* **6**(1), 43–51. [Assesses the inputs and outcomes of the funds devoted to promote research in biomedicine in relation to the objectives of the patient-care system.]

Espinosa de los Monteros J. Díaz V. Toribio M. A. Rodríguez Farré E. Larraga V. Conde J. Claveria L. E. and Muñoz E. (1999). La investigación biomédica en España. (I) Evaluación del Fondo de Investigación Sanitaria (FIS) a través de los proyectos de investigación financiados en el periodo 1989–1995 a centros de investigación a instituciones sanitarias asistenciales (hospitales). (II) Evaluación del Fondo de Investigación Sanitaria (FIS) a través de los proyectos de investigación financiados en el periodo 1989-1995 a centros de investigación, facultades y escuelas', *Medicina Clínica,* **112**(5), 182–197 and (6), 225–235. [The two articles aim to assess the funding activities of the Spanish National Health Research Fund through an analysis of the research projects by surveys addressed to the principal investigators and the managers of the research centres.]

Laredo P. and Schimank U. and Winnes M. (1999). *An Approach to Public Sector Research Through its Research Collectives. Overview, Interim Report B,* 25 pp. plus two Appendices of 123 and 74 pp. Paris/Köln: Armines/CSI and Max-Planck-Institut für Gesellschaft. [Presents a synthesis of the work undertaken for the international comparison of public sector research in Europe and some OECD countries using an exploratory approach based on the quantitative and qualitative analysis of laboratories or research units acting in the field of human genetics selected as a test.]

Muñoz E. (dir) (1999). *Memoria Final Proyecto FIS 96/1803* (mimeo), 517 pp. Madrid: IESA, CSIC. [Contains a series of contributions by the members of the multidisciplinary team that evaluated the different activities-research projects, training schemes, and scientific production-derived from the funding activities of the Spanish National Health Research Fund along the period 1988–1995.]

Muñoz E. in collaboration with Díaz V. M., Espinosa de los Monteros J., and Santesmases, María J. (2000). Targeted research and technological innovation and their relationships in a new socio-political context. Approaches to their evaluation. Presented in the workshop on *Innovation and Diffusion in the Economy*: *The Strategy and Evaluation Perspectives*, Lisbon 24-25 January, Lisbon: CISEP (ISEG, Technical University of Lisbon). [Introduces the comparative view within a European frame of the different evaluation exercises carried out in Spain under the direction of the main author.]

Narin F. and Noma F. (1985). Is technology becoming science? *Scientometrics*, **7**(3–6), 369–381. [In this article the authors apply a quantitative methodology based on publication records to establish the links between science and technology.]

Sanz Menéndez L. (1995). Research actors and the state: research evaluation and evaluation of science and technology policies in Spain. *Research Evaluation,* **5**(1), 79–88. [A review of the development of research evaluation practices in Spain in the last twenty years.]

Snyderman R. and Holmes R. W. (2000). Oversight Mechanisms for Clinical Research. *Science* **287**(5453), 28 January, 595–597. [Debates on the consequences for human subjects of the increasing complexity of clinical trials and on how to assess them and adapt the situation to a changing environment.]

Van der Meulen B. and Rip A. (2000). Evaluation of societal quality of public sector research in the Netherlands. *Research Evaluation,* **8**(1), 11–25. [An important effort to evaluate the relevance and impact of the research activities in the Dutch public sector.]

Biographical Sketches

Emilio Muñoz. Born on January 13, 1937, is Research Professor in Biology and Biomedicine at the Consejo Superior de Investigaciones Científicas (CSIC), now in the Unidad de Políticas Comparadas (Grupo de Ciencia, Tecnología y Sociedad), former Instituto de Estudios Sociales Avanzados, C/ Alfonso XII, 18, 28014 MADRID, Spain. From 1982 to 1991, he served different managerial positions in the Spanish R&D Administration (General Director for Science Policy, General Secretary of the National Plan on R&D, Chairman of the National Research Council, CSIC). He has been member of the CREST European Committee and Chairman of the COST Program. He has published more than 350 articles on biochemistry, molecular biology, science and technology policy as well as public understanding of science with special emphasis in the field of life sciences. He has authored or co-authored several books, among them: *Establecimiento de la bioquímica y biología molecular en España*, Centro de Estudios R. Areces (with María J. Santesmases); *Biotecnología, Industria y Sociedad: El caso español*, Fundación CEFI; *Changing structure, organisation and nature of public research systems. Their dynamics in the cases of Spain and Portugal*, IESA-CSIC (with María J. Santesmases and J. Espinosa de los Monteros). He is now involved in the analysis and evaluation of scientific and technological policies with special emphasis on the problems of R&D Regional Policy and coordination between administrations as well as in the biotechnology field, analyzing the socioeconomical and political implications of the uses of biotechnology. He has been member or associated partner in about a dozen of European projects funded by DG XII on different programs (MONITOR, BRIDGE, STRIDE, FAIR, TSER).

Juan Espinosa De Los Monteros. Born in Madrid (Spain) on June 1, 1938. He holds a Ph.D. in Chemistry granted by the Universidad Complutense of Madrid (Spain) and a postdoctoral research career at the University of Sheffield United Kingdom). He is a Research Associate Professor at the Consejo Superior de Investigaciones Científicas – CSIC (National Research Council of Spain), now in the Unidad de Políticas Comparadas, Grupo Ciencia, Tecnología y Sociedad, Alfonso XII, 18, 28014 Madrid (Spain). From 1982 to 1992, he served different managerial positions in the Spanish R&D Administration as Responsible for Management of several R&D Programs: Acuaculture and New Materials at the Interministerial Committee of Science and Technology (CICYT); and manager in charge of R&D Programs at the Instituto de Salud Carlos III (Ministry of Health and Consumption Affairs of Spain). He is now involved in the analysis and evaluation of scientific and technological policies with special emphasis on the problems of R&D programs and their relationships with regional and international policies. He has published a series of articles on evaluation of public R&D policies and programs in national and international journals (*Science and Public Policy, Research Evaluation, Medicina Clínica*....). He is the author or co-author of a series of books, among them: *Evaluación de las actividades de investigación y desarrollo tecnológico en acuicultura en el periodo 1982-1997*. Tomo I, *Directorio de publicaciones de investigadores españoles en acuicultura*. Tomo II (Ministerio de Agricultura, Pesca y Alimentación, Centro de Publicaciones).

Víctor Díaz. Born on October 22, 1970, holds a degree in Political Sciences and Sociology, speciality of Industrial Sociology, granted by the Salamanca Pontifical University at Madrid. He is performing studies for a doctoral degree in the Department of Sociology, section of Human Ecology and Population, at the Madrid Complutense University. He has carried out research in the Institute of Demography of the Spanish Research Council (CSIC) on the prospective of the evolution of Spanish population until 2025. He possesses diplomas of specialist in several areas related to the evaluation of training programs, human resources and labor law and social security. He is now a member of the Grupo Ciencia, Tecnología y Sociedad in the Unidad de Políticas Comparadas (CSIC), where he is directing and collaborating studies on the evolution of public policies, in particular in the fields of health and biomedicine and biotechnology, being in charge of the monitoring during the last five years of the evaluation of biotechnology sector in Spain. He has published several articles in peer reviewed national and international journals. He has contributed to the first analytical and prospective study on industrial biotechnology in Spain sponsored by the Ministry of Industry and Energy (1999-2000). He has authored or co-uthored several books, among them: *Salud y ancianidad en Segovia* and *Evaluación de las actividades de investigación y desarrollo tecnológico en acuicultura en el periodo 1982/1997*.