MEASURING THE POLITICAL IMPACT OF SCIENCE

A. Gaunand [1,2, µ], L. Colinet[3], M. Matt[4], P.-B. Joly[2]

[1] INRA, Delegation for Evaluation, 147 rue de l'Université, 75338 Paris Cedex 07
[2] LISIS, CNRS, ESIEE Paris, CNRS, UPEM, Université Paris Est, F-77454 Marne-la-Vallée Cedex 2
[3] INRA Directorate College – INRA Department head – 147 rue de l'Université, 75338 Paris Cedex 07
[4] INRA, UMR1215 GAEL, F-38000 Grenoble, France and Univ. Grenoble Alpes, UMR 1215 GAEL, F- 38000 Grenoble, France
µ: corresponding author: Ariane.gaunand@inra.fr

LISIS LABORATOIRE INTERDISCIPLINAIRE
Sciences Innovations Sociétés

Abstract

The production of scientific knowledge is expected to contribute to public policy in a variety of ways; however, methods of assessment of the impact of research on policy are underdeveloped.

This paper proposes an original methodology to assess the political impact of research. It builds a rating scale based on the submission to expert judgment of a set of standardized innovation case studies. Our position regarding political impact assessment is novel: 1) we do not account for the impact of policy implementation; 2) we consider the political impact of all types of research results rather than only policy-oriented research; 3) we adopt a broad definition of political impact which goes beyond instrumental changes. The resulting metric uses a 1 to 5 scale to reveal the intensity of the political impact of research, using generic criteria associated to each rating level. This rating scale can be used to objectify the political impact of self-assessed case studies. We tested the robustness of our rating scale in routine impact assessments of new cases. We believe that the proposed scale can be used to reveal the types of mechanisms involved in the political impact at program portfolio or research organization level based on regular rating of additional case studies.

INTRODUCTION

The hope that the knowledge generated by science will benefit society is the founding principle underpinning the provision of support for public research. Thus, measuring the effectiveness of this research-impact relation is a pressing issue. The literature on the state of the art regarding research impact assessment recommends consideration of a diversity of research impacts (including economic, environmental, and political) on society (Bornmann, 2013; Bozeman and Sarewitz, 2011). Assessing the economic returns to investment in research is not a new issue and has been analyzed since the 1950s. Most analyses of the economic impact of R&D investment concern agriculture, starting with the seminal contribution of Griliches (1958; Cf. Alston, 2010 for a comprehensive survey).

However, there are no robust and reliable methods for measuring the other dimensions of societal impact including the political impact (Bornmann, 2013). On the one hand, Renkow and Byerlee (2014) note that the quantitative assessment methods of CGIAR's and ACIAR's research are relevant only in the context of local political impacts because the broader international impact cannot credibly be attributed to a specific public research organization (PRO). Hazell and Slade (2014) concur with this view and find that quantitative political impact assessment methods are rare: only around 10 among 31 impact studies of policy research commissioned by CGIAR, provide quantitative impact assessments, and they do not cover all the possible dimensions of political impact. On the other hand, qualitative approaches lack objectivity (Bell et al., 2011), and prevent analytical scaling-up from case studies grounded in

non-reproducible theoretical frameworks (Bozeman, 2003; Boaz et al., 2009). However, in the context of assessing political impact, these methods enable description of the contribution of research to the policy process (Boaz et al., 2009). Despite a few initiatives in the health sector (Boaz et al., 2009), there are no well-developed, comprehensive practical frameworks (Almeida and Báscolo, 2006; Cozzens and Snoek, 2010), and no ready-to-use rating scales for measuring political impact (Raitzer and Ryan, 2008) in the literature.

The present research is part of the ASIRPA project which has developed a methodological approach based on standardized case studies to qualify and quantify the socioeconomic impacts of the research results generated by scientists from the PRO INRA (French National Institute for Agricultural Research). Using the ASIRPA methodology (Joly et al., 2015) each case study's ex-post impacts are summarized across five dimensions (economic, political, environmental, social-territorial, and health) corresponding to INRA's missions and aligned to the international literature (Bornmann, 2013). Local impact descriptors for each of these dimensions are collected via desk research and stakeholder interviews. Evidence of impact is reported in a table, and the intensity of the impacts are depicted on a radar diagram and scored on a scale ranging from 1 (negligible impact) to 5 (major impact).

In this paper, we build a rating scale for political impact assessment by combining a quantitative rating scale with qualitative evidence of political impact. We consulted an expert panel to objectify assessment of the intensity of the political impact. Our rating scale is constructed in line with state of the art recommendations related to combining qualitative and quantitative analysis, and aims to be reproducible, comprehensive, and sufficiently general that all possible cases generating political impact can be judged using the same rating scale.

We first review the main works and issues addressed in the literature on political impact assessment and expert panels (section 1). Based on these key conceptual and practical issues, we describe the steps followed by ASIRPA for designing an original and operational methodological tool to assess the political impact of research (section 2). Section 3 discusses how this approach overcomes some of the limitations in earlier works, and provides some insights into the organizational learning facilitated by feedback on its implementation at INRA. Section 4 concludes.

1/ LITERATURE REVIEW: THE FOUNDATIONS OF A PRACTICAL FRAMEWORK FOR ASSESSING THE POLITICAL IMPACT OF RESEARCH

The recent literature on political impact assessment, and expert panels reveals four key elements. First, science and its impact on policy has for long been studied using a linear process. Recent conceptual frameworks to account for social and contextual elements in this complex process are affecting how the contribution of research to political impact should be assessed. Second, very few practical methods derived from these conceptual models of knowledge flows have been implemented to assess the political impact of research programs or institutions. Third, implementations of political impact assessment are fraught with difficulties both methodological (setting a definition of political impact, and developing credible rationales) and technical (collecting data). Fourth, expert panels which traditionally are involved in academic career assessment, can with some caveats, be of great value for assessing political impacts.

1.1 Conceptual frameworks to understand the process of policy-making

To understand the nature of the impact of research on policy, it is important to identify the actors involved and the research result used for policy-making, and the way they are used, as well as the process involved in political impact - the transformations that occur and their timing. There are various conceptual frameworks that explain the possible contribution of knowledge, and help to set the terms for research impact assessment (Davies and Nutley, 2008).

The first conceptual framework proposed to understand the qualitative contribution of research to policy mapped an ideal, linear, policy cycle model based on four sequential stages (Lasswell, 1977; Lasswell and Lerner, 1951; Howlett and Ramesh, 1995): problem identification and agenda setting; policy formulation and adoption; policy implementation; policy evaluation and reformulation. This linear depiction of policy-making assumes that policy-makers are perfectly rational actors, and that if correctly «packed" and «disseminated» the knowledge produced by science can be directly (instantly) used by decision-makers (Caplan, 1979). However, this vision which continues to permeate the way policy research projects are conducted has been challenged.

First, authors such as Weiss (1977, 1979) demonstrated that this seamless linear depiction did not represent reality. Her work which has been taken up by contemporary authors, highlights the importance of context (societal concerns, changes in the economic context) on the timing and flow of policy-making, and the use of research knowledge. More recent evaluation approaches also emphasize the timeline of policy-making (incremental vs. sudden policy changes) which is given less than sufficient weight in linear models (Cozzens and Snoek, 2010; Jones, 2009; Weiss, 1979).

These insights spurred work on the contribution of research to policy changes in light of windows of opportunity to use research results (Kingdon, 1984; Lindquist, 2001). Scientific knowledge is more likely to be used as the basis for policy-making if it matches societal concerns, or sectoral concerns, or political attention (Kingdon, 1984).

Second, there are some models that highlight that the contribution of scientific knowledge to public policy is not primarily a matter of information flow and format but rather is a social process which depends on networks, credibility, and the balance of power in policy-making (Weiss, 1980; Cash et al., 2003; Cozzens and Snoek, 2010). These models suggest that it is necessary to pay attention to the whole process: the conditions in which the knowledge is produced, the contribution made by other sources of knowledge, the interactions between researchers and end-users, and the roles of intermediaries in circulating and mediating the research results.

Third, the proposed models are built on empirical studies which show that scientific knowledge can affect political stakeholders in various ways. The contribution of scientific knowledge to policy is seldom instrumental or clearly identifiable (Almeida and Báscolo, 2006; Weiss, 1980). For example, knowledge can steer policy agendas (Kingdon, 1995), influence the composition of the "issue networks" that provoke and guide the exercise of power (Heclo, 1978) or the strategic positions of actors (Davies and Nutley, 2008), and over time, infiltrate advocacy coalitions (Sabatier and Jenkins-Smith, 1993), change policy paradigms (Weiss, 1980), influence the terms of debate (Davies and Nutley, 2008), affect belief within institutions and networks (Davies and Nutley, 2008; Radaelli, 1995), and change people's knowledge, understanding and attitudes (Davies and Nutley, 2008). Weiss (1979) identifies seven paths for research use in policy, which they cluster under three large categories (Almeida and Báscolo, 2006; Trostle et al., 1999): rational and instrumental use of knowledge to formulate policies or support decision-making; strategic use of knowledge to strengthen or weaken actors' positions; conceptual use of knowledge to enable deeper understanding of issues and policies (a circulation process that is slow and progressive).

This rich conceptual analysis of the policy-making process, accounting for contextual windows of opportunity and a diversity of contributions of scientific knowledge to policy, is the motivation for a comprehensive framework for the analysis of the political impact of research. Our evaluation framework includes the contribution of research results to political agenda-setting and actors' debates and strategic positioning, instrumental use of policy-making, and the long-term effects of the terms of debates. Regarding this last, the enlightenment model developed by Weiss (1977) supports the view that new research results gradually affect the intellectual backgrounds of policy-makers. A consequence of this cumulative and gradual effect is that some political changes may be based on the contribution made by non-policy-oriented research. Therefore, we explore this dimension in the case studies involved in the ASIRPA research project.

1.2 Practical frameworks to evaluate the impact of research on policy

Despite the multiple conceptual frameworks proposed in the literature, recent reviews highlight that the practical frameworks used for impact evaluation usually fail to explore the diversity of the possible contributions of research knowledge to policy-making, and pay an excessive attention to instruments or practices to enhance policy impact (Almeida and Báscolo, 2006; Boaz et al., 2009; Tsui et al., 2014). Some approaches rely on rich explanatory framework, notably in the areas of health and international development. Boaz et al. (2009) refer to three methodologies: the RAPID outcome assessment framework, the research impact framework (RIF), and the Health Economics Research Group payback model.

RAPID is an assessment tool developed by the Overseas Development Institute, and is based on pioneering work conducted by the International Development Research Centre (Earl et al., 2001) to understand the influence of its research on policy. It takes account of factors such as context, and key actors and their behavior (Boaz et al., 2009; Jones, 2009). RAPID includes a timeline tracing the main events affecting the project, and an impact pathway highlighting important changes in the environment, policy, relations among actors, and the links between these aspects. The contribution of the project to policy is demonstrated in information gathered during an intensive workshop with project members and partners, and key informants. The approach involves triangulation of different perspectives regarding the contribution of research results to policy, and the value of their impact for different societal actors. It is less effective for enabling comparison between cases because the analysis is very case-specific; however, cross-analysis of a collection of cases using a common methodology, can be envisioned (see e.g. Carden, 2004; Court and Young, 2006).

The RIF (Kuruvilla et al., 2006) was designed for health research. It features detailed descriptive impact categories to help researchers identify the targets of their research. Policy impact is described under five categories: level of policy-making potentially affected (international, national or subnational level); type of policy affected (practice, service, or governance); the nature of the policy affected (inspired by Weiss' s (1980) typology); changes to policy network (informing policy networks); and changes to political capital (improving the quality of deliberations). This framework provides researchers with a set of categories that structure the case study narratives, and facilitate analysis across projects and time. The RIF remains descriptive, and does not provide tools to objectify assessment of the value of impact (Bornmann, 2013).

The payback model (Hanney et al., 2003; Wooding et al., 2014) developed by the Health Economics Research Group identifies and scores research impact in five categories, including one related to political impact: "informing policy and product development". The approach organizes sets of cases, and the data are collected according to a conceptual framework based on the various stages of a logic model depicting research utilization. The data are then scored under each category, by panels of experts using the Delphi method (expert assessments are shared through repeated consultation rounds). These data are used to compare impact profiles across projects and programs. This model takes account of Weiss's (1979) many uses of research-driven knowledge, and offers a scoring that can be used to compare impact profiles across projects. However, assessment of new cases requires a costly new Delphi method to be implemented, and in addition, the Delphi method does not reveal the judging criteria developed by the experts to score the cases.

There are some practical tools which have been proposed to analyze the impact of research on public policy, and to compare case studies. However, these frameworks do not routinely compare the value of different cases investigated independently, nor do they scale-up learning in relation to the definition of political impact at the level of the organization. The rating scale we build tries to fill these gaps.

1.3 Practical issues regarding data collection

Whatever the overall approach selected to assess political impact, practical issues arise during the collection of data to support the political impact of research. Most of these issues are common to many other dimensions of the socio-economic impact of research. However, the challenges are greater in the case of political impact because of the causality problems identified by many authors (for a review see e.g. Almeida and Báscolo, 2006; Boaz et al., 2009; and Donovan, 2011). Penfield et al. (2014) relate the inability of stakeholders to source the origins of the knowledge they gain in contexts where knowledge is absorbed slowly (what they call 'knowledge creep').

Another important practical issue is related to the major conceptual problems regarding how societal impact (including the impact of policy) should be defined and assessed (Bornmann, 2013; Brewer, 2011; Donovan, 2011). The way political impact is defined affects the data to be collected. Some authors (Hazell and Slade, 2014; Renkow and Byerlee, 2014) distinguish between the contribution of research which results in a policy outcome (often a regulatory change), and the contribution which results in the downstream impact of the policy outcome (changes to societal indicators such as improved health or housing). Trying to piece together the contribution of research in relation to the downstream effect of policy is extremely difficult since the effect of research knowledge is often too dilute to ascertain its contribution (Renkow and Byerlee, 2014; Carden, 2004). Thus, some authors recommend a focus on the impact of research on policy rather than the impact (of the policy) on society (Boaz et al., 2009; Carden, 2004). The lack of consensus among experts about what should be measured as an impact on policy, and how it should be measured, is illustrated in a study by Samuel and Derrick (2015). Drawing on 62 interviews with evaluators from a health-related panel, Samuel and Derrick (2015) found that about a third perceived impact as being achieved only in the presence of some final and positive change on society resulting from the policy implementation. Those evaluators

would for example, discard the inclusion of research results in policy documents as ephemeral measures which provided no long lasting benefits for society. Although the majority of the evaluators considered that "there are different stages of impact", and were willing to consider a broader view of the contribution of research to policy, they were unsure about which stages could be considered as impact, and to what degree they could be scored against each other (Bornmann, 2013; Samuel and Derrick, 2015).

This diversity of views among evaluators needs to be addressed to provide a robust framework for evaluation across different cases. We follow Carden's (2004) view of the assessment of political impact, and report only on the impact of research on the policy process; the effects of policy implementation are reported in another branch of our ASIRPA impact radar. This strategy which is in line with the payback framework (Hanney et al., 2003; Wooding et al., 2014), prevents double-counting. The consequences for data collection include a focus on collecting proof of the often intangible impact of research on the policy-making process, rather than focusing on more easily observable macro-changes related to policy implementation. A fundamental problem related to producing a robust evaluation of the impact of research on policy is defining rules, first about what constitutes acceptable evidence of the impact of research, and second about how to form a judgment of the value of the impact (Penfield et al., 2014).

1.4 Using expert judgment to assess political impact

Expert panels are used commonly to evaluate research projects and programs on the basis that they bring status and credibility to the process (Arnold et al., 2005; Boaz et al., 2009). These methods are most frequent for peer reviews to evaluate the academic quality of research proposals, careers, or papers (Ruegg and Feller, 2003). However, experts are used also to provide credible judgments of the societal impact of research when information is not easily available. For example, panels of researchers, managers, and stakeholders are commonly used to evaluate the societal impact of European Union research programs (Georghiou and Roesner, 2000; Molas-Gallart and Davies, 2006) in the form of what Ruegg and Feller (2003) term "relevance review". Panels to evaluate the socio-economic impact of research tend to adopt the procedures used by peer review panels to evaluate research quality, with little consideration of their specificity; however, they tend to include fewer researchers and a higher representation of stakeholders and end-users (Bozeman and Youtie, 2015).

The literature provides recommendations on panel reviews to ensure the objectivity and diversity of the opinions expressed in the assessment process. Upfront preparation by a secretariat of a synthesis of the available data is important to ensure quality (Ruegg and Feller, 2003), and facilitates interaction among members (Arnold et al., 2005; Boaz et al., 2009). The selection of panelists is also important; the size and the composition of the panel must ensure constructive discussion among members. For example, Arnold et al.

(2005), and Langfeldt (2004) observed that among panelists with little overlaps in competences, the group will follow the lead of a few members with more specific knowledge. This can result in a high division of tasks and little interaction among members which will be detrimental to quality. Thus, Boaz et al. (2009) warn against use of expert panels to evaluate broad thematic areas since this would call for the involvement of too large a number of experts. Bornmann (2013) recommends panels involving stakeholders with experience in the exploitation of research.

Arnold et al. (2005) point out that panelists are often not explicit about how judgments are made, and traditionally act as «the authority». This hinders transparency and consistent evaluation (Langfeldt, 2004; Samuel and Derrick, 2015). Encouraging experts to discuss their evaluation rationale should be promoted.

Building on this experience, our decision to consult an expert panel was adapted to our goal of transparent, objectified judgment. We hypothesize that allowing expert panel to build our scale would reduce variability in assessment which might arise from individual values and perceptions of political impact. Following the recommendations in the literature, we constructed a panel taking account of panel size, member seniority, and overlapping competences.

2/ METHODOLOGY: BUILDING A RATING SCALE

Our objective is to design a rating scale with the following properties:

- To match impact scores based on generic criteria with political impact relevant to a diversity of cases;

- To build a sufficiently explicit scale to allow for objectified self-assessment by the researchers involved in the cases, on the basis of information collected from stakeholders in interviews;

- To build a standalone scale which requires no further mobilization of expert panels for individual cases. We envisage regular external evaluation using expert panels to i) check consistency of use of the scale, and ii) to solve emerging issues not addressed by the scale.

The design of the rating scale involved four steps described below.

C1. Drawing on the literature to design an analytical framework based on the dimensions of potential contribution of the research to policy-making, and its associated impacts.

C2. Using an expert panel to judge a sample of five case studies, using and refining the proposed framework, and clarifying the values on which their judgments are based.

C3. Designing a rating scale, derived from the criteria used by experts to form their judgment.

C4. Testing the rating scale on the 41 cases currently available.

2.1 Building an analytical framework to analyze the political impact of cases

We stylized our analysis by drawing on the rich case study material, and on the literature to design an analytical grid (cf. section 1, and particularly the paths towards use of research in policy, identified by Weiss 1979, and refined by Almeida and Báscolo (2006), the knowledge streams defined by Kingdon (1984).

We used the first five cases[1] with potential political impact as pilots: they constitute the empirical basis for our rating scale design. A report and an executive summary (5-10 pages) were written for each case studied in the ASIRPA project. Following Joly et al.'s methodological guidelines, (2015), these reports provide standardized drafts which include narratives, timelines, impact pathways, and impact radars. The impact pathway highlights the various features of political impact and the mechanisms for their generation. Research outputs (scientific knowledge) are produced by a network of research partners which includes INRA. This scientific knowledge can be absorbed and used by a diversity of intermediaries and end-users (sectoral professionals, decision-makers, media, members of local administrations, members of parliament). Its exploitation can have various first-level, political impacts for a first sphere of end-users: it might affect the steps involved in the policy-making process (agenda-setting, negotiation, formulation, implementation, evaluation) at different levels of policy (local, national, international); it may be used in sectoral and general public debate. The use of research outputs can also have second-level impacts, that is, affecting a more removed sphere of users, resulting from a scaling-up or scaling-out of the first-level impacts. These second level political impacts will likely result from a slow percolation of new knowledge and ideas which will materialize as long-term, wide ranging effects. This research-to-impact pathway will be affected by contextual events (societal concerns, disputed policy regulation, political and media focus) which provide windows of opportunity for the use of knowledge in policy, and which affect the societal importance of the focal policy domain.

For each case, impact-generating mechanisms and impact data will be collected via desk research and interviews with stakeholders. This will allow qualitative and quantitative local descriptors to be collected for each impact dimension (economic, political, environmental, etc.).

The grid includes 4 dimensions and 13 sub-dimensions. The first three dimensions are related to the political impact, the fourth deals with the absolute importance of the affected policy domain (see table 1).

---

1 Fire Paradox: integrated European project on the management of forest fires, Collective scientific advice on Pesticides; Supporting conservation policies for Atlantic salmon: catch quotas; Alert on Bisphenol A; A genetic approach to fight against scrapie in sheep. For summary of these 30 pages report case studies, please visit:

http://www6.inra.fr/asirpa_eng/Method-and-Cases/Case-studies

Table 1: Dimensions and sub dimensions of the ASIRPA analytical framework for political impact

Dimensions of political impact	Sub dimensions
Use in public debate and policy negotiation *inspired by Almeida and Báscolo's (2006) strategic use of knowledge to influence actors' positions*	Quality and strength of research messages conveyed *(Cozzens and Snoek, 2010; Weiss, 1979)* Timeliness of debate and political agenda-setting *(Cozzens and Snoek, 2010; Kingdon, 1995)* Intensity and quality of media coverage *(insights from intermediary activities reported in the pilot cases)* Intensity and quality of debate *(Cash et al., 2003; Davies and Nutley, 2008)*
Use for policy-making *inspired by Almeida and Báscolo's (2006) rationale and instrumental use of knowledge to formulate policies or support decision-making*	Stages of the policy cycle affected: agenda-setting, and formulation, implementation, and evaluation of policies *(Cozzens and Snoek, 2010)* Territorial scales of policies *(insights from territorial embeddedness reported in cases)* Relevance and novelty of the solution provided for policy *(Cash et al., 2003, and insights from the research outputs depicted in the pilot cases)*
Long-term percolation of ideas *inspired by Almeida and Báscolo's (2006) conceptual use of knowledge to deepen the understanding of issues or policies*	Importance of knowledge in the debates *(Cash et al., 2003; Davies and Nutley, 2008; Radaelli, 1995)* Circulation of ideas in later studies/debates and broader spheres *(Almeida and Báscolo, 2006; Sabatier and Jenkins-Smith, 1993; Weiss, 1980; and insights from cases on scaling-out of impacts)* Long-term relevance of ideas and non-distortion of messages *(Weiss, 1980, and insights from scaling-out and scaling-up effects reported in second-level impacts of pilot cases)*
Societal importance of the policy domain at stakes *inspired by Kingdon's (1984) problem, policy and political streams*	Potential gravity and systemic aspects of stake *(Renkow and Byerlee, 2014, and insights from the pilot case on the sheep scrapie sanitary crisis)* Magnitude of the policy and affected population *(Renkow and Byerlee, 2014)* Societal concern *(Kingdon, 1984; Lindquist, 2001; and insights from the pilot case on the sheep scrapie sanitary crisis)*

2.2 Composing and consulting an expert panel and preliminary work

The consultation was carried out in two steps: experts were asked first remotely and individually to rate the pilot cases before meeting up to discuss views.

Composition of the expert panel

We selected five experts on the basis of their experience in assessment of the impact of research on policy-making. All the experts were French nationals with research backgrounds. We chose a small panel to promote interaction and consensus but taking into consideration overlapping competences. Three experts came from the three ministries that were potential users of the research knowledge produced in the case studies; they were appointed from divisions dealing with knowledge to public policy, and two experts were involved in studying the societal impact of research. With the exception of one, all of our experts were knowledgeable about the agricultural and environmental policies implemented in previous decades; the one exception was involved in evaluation research. All the experts were expected to take similar precedence in discussions since all were familiar with policy-making processes, had similar same seniority, and were not linked within a hierarchical relation.

Providing panelists with comprehensive, easily exploitable data

The ASIRPA team reviewed the five pilot cases using the conceptual framework presented in table 1. Evidence of impact extracted from the case reports was sorted to match the corresponding analytical dimension. Experts were provided with executive summaries and political impact analytical tables (=table 1) prefilled with the evidence collected for the five pilot cases.

Individual experts' preparatory work: remote rating

Each expert was asked, for each case, to rate each of the four dimensions of the analytical framework on a 1 to 5 scale. They were asked to provide an argument for each of their judgments. We predicted that discussing the rating of real cases would provide more information than if experts were to discuss desirable evaluation criteria. Initially, our experts were reluctant to act as authorities in rating the political impact of cases, since they were skeptical about the quality of the mediated evidence delivered by INRA. However, since the rating task was designed to obtain a rationale for judging impact on each of the dimensions in order ultimately to design a rating scale, they agreed to the task.

Expert meeting and definition of criteria

The objective of the meeting was not to obtain a consensus on a single mark for each case but to elicit each expert's rationale for the rating awarded. The experts were invited to a one-day meeting. One served as meeting chair to facilitate discussion and enlarge the debate. The experts were invited to comment on the marks assigned to each case study, dimension by dimension, and compare the merits of cases. The size and composition of the panel proved successful for promoting expression of a variety of opinions in a frank and dynamic exchange. The experts declared themselves to be satisfied with the dialogue, and in-depth analysis of the political impacts.

The experts validated the subdivision of political impact into the four dimensions presented in the initial analytical framework (table 1). The arguments used by the experts to justify how they rated each case helped us to refine, complete, and organize the list of sub-dimensions to be assessed for each dimension. During the panel meeting, experts shared their rationales for their individual ratings for each dimension of political impact, based on the available evidence of impact provided by the cases. Their individual arguments were debated and challenged, resulting in individual explanation and reformulation to develop more robust and diverse arguments. This collective screening of judgment arguments based on the sample of cases was carried out for each dimension and sub-dimension of political impact. The experts proposed 1) a rating on a 1 to 5 scale assigned by each expert to each dimension of political impact for each case, and 2) the arguments associated to each rating, and related to each sub dimension of political impact.

2.3 Processing data and designing an evaluation rating scale

After the meeting, we analyzed the dispersion of the expert ratings, and their arguments associated to each sub-dimension. After collective discussion and rerating, the ratings awarded by the different experts to specific cases along each political dimension differed. However, there was no overall incoherence in the ratings since the experts tended to agree on assignment of a high or low score, justified by similar arguments.
We processed the data first by identifying generic evaluation criteria, and second by hierarchical ordering of these criteria.

Experts did not offer isolated arguments to justify their ratings for each sub dimension; rather they combined several arguments related to a given sub-dimension. This constitutes a set of evidence on which the rating is based (see table 2). The cases highlight the pathways to political impact, and disentangle the mechanisms that generate a processual and overall political impact. For example experts considered that while dissemination of research knowledge in the media would be considered positive for policy debate, it would be less important if the message became distorted.

Table 2: Examples of the arguments proposed by the experts on three cases for the dimension "Use in public debate and policy negotiation"

Experts	Mark (/5)	Arguments related to each sub-dimension of "Use in public debate and policy negotiation":			
		Quality and strength of research messages conveyed	Timeliness of debate, political agenda-setting	Intensity and quality of media coverage	Intensity and quality of debate
Case *Alert on the dangers of Bisphenol A(BPA)*					
A	4	Research raised new questions for the agenda		Very intense mainstream media coverage	
B	3	Some weaknesses in the message affect credibility (data and research design)			
C	3	INRA is not the main scientific referee, affects strength of messages			
D	5		Direct effect on a political window of opportunity	Intense media coverage to a large audience (policy, politic, citizens, private sector). Little distortion of messages	Large contribution of research to the debate beyond sectoral policy, at the national level (parliament)
Case *Scientific Public Expertise on Pesticides (Pesticides expertise)*					
A	4			Very intense media coverage in the technical sphere, less intense for the public sphere	Intense use in policy debate
B	4	Strong credibility of messages, supported by political scientific referee			Large contribution of research in shaping debate. Messages understood and relayed.
C	4	Some policy recommendations are new	Mostly national sectoral policy windows. Missed some windows of opportunity	Intense media coverage. A few messages were not relayed	
D	3	Presentation of scientific state of the art. Nothing really new		Large intensity of media coverage	Policy debate largely used information, however a few messages were not used, or distorted
Case *Genetic fight against scrapie*					
A	3			Mostly technical media coverage, little mainstream coverage towards general public	Contribution of research reduced tension in debate for sectoral policies
B	3		Sectoral windows of opportunity prompted by SBE crisis		Strong contribution to a local debate. No national debate
C	2	Strong credibility of messages but some research results did not spread			
D	3			Information given to stakeholders of the sectoral policy only	Strong contribution to sectoral debate, but not opened to national debate

Based on collective challenging of the individual arguments (table 2), we derived generic evaluation criteria related to each sub-dimension (see table 3 for an example for the dimension "Use in public debate and policy negotiation"). Given our previous remark related to combining arguments on the rationale for judging a complex policy process, evaluation criteria were not considered in isolation.

Table 3: Evaluation criteria for the analytical dimension "Use in public debate and policy negotiation"

Sub dimension	Criteria for evaluation
Quality and strength of research messages conveyed	• Excellence and scientific reputation • Adaptation of media to the diversity of audiences (technical level, ambiguity, etc. • Bringing new knowledge to stakeholders • Legitimacy of knowledge beyond the scientific arena, based on PRO reputation • Originality of research results, contrast with existing positioning of stakeholders
Timeliness of the debate and political agenda-setting	• Relevance of the issue to the political sectoral, societal agendas, and the scientific controversies
Intensity and quality of media coverage	• Amount of media coverage • Efficiency of media coverage: diversity of media, their contribution and targets • Distortion of messages : selection of fragmented messages, controlled use and distortion of messages
Intensity and quality of debate	• Relevance of the debate triggered compared to targeted stakeholders or policies • Diversity of audiences involved in the debate (sectoral stakeholders, general audience, etc. • Matching debate level and decision-level

We used the expert ratings to build a rating scale for each sub=dimension, for each expert associating the particular argument used to refer to specific evaluation criteria based on the individual expert's judgment about the intensity of the political impact (see table 2).

Hierarchy of evaluation criteria through a Condorcet method

To order the evaluation criteria we use a Condorcet method to collectively rank the cases based on the individual rankings of each expert. The Condorcet method elects the candidate that would win according to majority rule in all pairings against the other candidates. Since the experts were invited to compare each case study to the other case studies for each impact dimension, we considered that the expert ratings expressed their order of preference. For each case we conducted a series of pairwise comparisons with the other four cases (involving 10 pairwise hypothetical elections per sub-dimension). The case identified individually by a majority of the experts to upgrade each of the other cases (in a pairing comparison) was ranked collectively higher, and vice-versa for lower ratings. Since the number of votes was limited, we chose to consider a difference of one vote in favor of a candidate to be a tie. If the Condorcet method did not separate two cases, we awarded them the same marks.

This collective consensus re-ranking procedure was performed for each of the four political impact dimensions. Since each expert ranking was associated with the expert's

arguments this collective re-ranking of cases led to a hierarchy of evidence regarding the evaluation criteria.

The rating scale

This provided a five point rating scale for each dimension of political impact with each ranking associated with related evidence regarding the evaluation criteria. The scale is presented in table 4. The table had a few empty cells where there was no information derived from the case studies. The resulting scale and methodological conclusions were sent to the panel and refined further based on their comments in order to complete the empty cells.

Computing the total score for political impact

The overall political impact score for each case is calculated as the arithmetic mean of its rankings for each of the four dimensions. Weights are assigned as follows: a factor of 1 was assigned to each of the first three dimensions related to the contribution of INRA to political impact, and a factor of 3 was assigned to the dimension related to the importance of the policy domain at stake. We chose to balance the intensity of the contribution with the importance of the policy, in order to avoid bias which would reward an important contribution to a minor local policy, and discourage a small contribution to a global policy challenge. The possibility of this bias emerged from discussion among our panel of experts and review of the literature (see e.g. Renkow and Byerlee, 2014).

Table 4: Rating scale for each Analytical Dimension

Table 4.1 Dimension Use in public debate and policy negotiation

Mark	Quality and strength of research messages conveyed	Timeliness of debate and political agenda-setting	Intensity and quality of media coverage	Intensity and quality of debate
5	Original messages, easily traceable in the public debate Strong credibility related to PRO renown	Agenda-setting of new questions	Large media coverage to general public and stakeholders involved. Messages properly conveyed.	Large public debate The debate involves the whole spatially relevant political sphere
4	Messages easily traceable in the public debate Original messages but arising from a state of the art rather than new research results OR Original knowledge but moderate credibility	Knowledge produced during a political or societal window of opportunity	Large media coverage to general public and stakeholders involved. Messages properly conveyed, although with some slight cherry-picking	Large public debate The debate partially involves the spatially relevant political sphere
3	Messages easily traceable in the public debate But weaknesses for some reasons : technical, legitimacy, ambiguity	Knowledge produced during a sectoral window of opportunity	Media coverage to sectoral stakeholders only. Messages properly conveyed	Broad sectoral debate at relevant spatial level, but no public debate
2	Messages poorly traceable in the public debate	Knowledge mediated during a sectoral or societal window of opportunity	Incomplete media coverage to sectoral stakeholders	Scattered debates with no territorial or sectoral relevance
1	Messages are not new	Knowledge produced and mediated out of any agenda	No or very small media coverage	Restricted debate, no public debate

Table 4.2. Dimension Use for policy-making

Mark	Stages of the policy cycle affected: agenda-setting, formulation, implementation and evaluation of policies	Territorial scales of policies affected	Relevance and novelty of the solution provided for the policy
5	Use at all stages of the national policy cycle (agenda-setting, formulation, implementation and evaluation of policies)	Important use at all the spatial scales relevant for implementing the whole policy cycle (local, national, international)	A new political solution, largely inspired by the scientific knowledge produced. The political solution is so relevant that it is used simultaneous by public and private sectors (in the last case, it is a technical solution to comply with public policy)
4	Important use at some but not all stages of the national policy cycle	Use at some but not all spatial scales concerned with policy (e.g. tools enabling the local monitoring of international commitments, but no national effect)	A new political solution, largely inspired by the scientific knowledge produced. The political solution is relevant and is used by public and private actors (in the last case, it is a technical solution to comply with public policy)
3	Contribution to the design of local implementation of a national policy	Use is uneven but concerns a variety of national, sectoral, and local stakeholders	A new political solution but just one among existing others. The solution is inspired partly by scientific knowledge but credibility related to PRO renown facilitates use of the political solution. No effect on the private sector
2	Contribution to the design of local implementation of local policy	Use is uneven. It concerns some local sectoral stakeholders	One political solution among others, partly inspired by scientific knowledge, but poorly used.
1	Minor use of knowledge in the policy cycle	Very few uses, even for local policies	One political solution among others, largely unexploited

Table 4.3. Dimension Long-term percolation of ideas

Mark	Importance of knowledge in debate	Circulation of ideas in later debates/studies, and broader spheres	Long-term relevance of ideas and non-distortion of messages
5	Likely contribution to shifting the terms of important debate. Reduced tension among major stakeholders in debate. Changes to some professional identities or positions assumed by stakeholder institutions	Rapid and sustained percolation of ideas to broad sectoral, scientific, and global audiences at the national and international levels	Very low risk of messages being distorted over the long-term thanks to: the PRO's established reputation as a scientific referent; Strong traceability of ideas affiliated to the PRO; Sustained involvement of the PRO and individual researchers in research and intermediation Strong relevance/resilience of ideas over the long run
4	Probable contribution to changing the terms of important debate. Few changes to professional identities or positions assumed by institutions but developing internal debate over disputed questions	Sustained but slower and more uneven percolation of ideas. Sectoral, scientific and general audience, and the national and international levels are not affected simultaneously	Low risk of messages being distorted over the long-term thanks to the PRO's reputation as a scientific referent; Sustained involvement of the PRO and individual researchers in research and intermediation; ideas poorly traceable due to their origins in an individual researcher rather than the PRO. Rapid obsolescence of ideas (e.g. : scientific advice)
3	Probable short-term contribution to debate with limited scope. Development of some internal debates and heterodox positions. Existing positions are reinforced by the scientific status of knowledge	Sustained but slow and very incomplete percolation: only one type of audience is affected (scientific, or sectoral, or general), at the national level only	Moderate risk of messages being distorted over the long-term: Sustained involvement of the PRO and individual researchers in research and intermediation; PRO has no scientific excellence in the domain at stake; ideas poorly traceable, due to their origins in an individual researcher rather than the PRO. Rapid obsolescence of ideas
2	Little change expected to the terms of debate (ideas already outdated) but existing positions reinforced by the scientific status of knowledge. Knowledge may be used in other policy areas	Slow but quite sustained percolation of ideas in scientific arenas	High risk of messages being distorted over the long-term: Sustained involvement of the individual researchers in intermediation activities but no sustained research funded by the PRO; the PRO is not scientifically renowned in the domain at stake; ideas poorly traceable due to their origin in an individual researcher rather than the PRO. Rapid obsolescence of ideas
1	No changes expected in the terms of debate (ideas already outdated), No possibilities for knowledge to be used in other policy areas	Punctual, opportunistic percolation of ideas	Very high risk of messages being distorted over the long-term: Punctual involvement of the PRO and its researchers in research and intermediation activities; PRO is not scientifically renowned in the domain at stake; ideas poorly traceable due to origins in an individual researcher rather than the PRO. The relevance of ideas is currently being contested

Table 4.4. Dimension Societal importance of the policy domain at stakes

Mark	Potential gravity and systemic aspects at stake	Magnitude of the affected population and policy	Societal concerns
5	Public policy addressing an issue with multiple aspects (e.g. sanitarian, environmental and economic) of critical importance (e.g. threat to human lives)	The whole national and/or an important share of international population is affected. National policies are affected.	Of huge societal and political concern. Regular crisis covered in the media. No societal consensus on the issue to be tackled.
4	Public policy addressing an issue with several intertwined and important aspects which are less crucial (e.g: non-lethal toxicity)	Almost the whole national population is affected. National policies are affected.	Great societal and political concern conveyed regularly in the media. Societal consensus on the nature of the issue but not its technical solution Public policy addressing important isolated social, territorial, or environmental issues
3	Public policy addressing important isolated social, territorial, or environmental issues	The affected populations are confined to one or a small number of areas or species or the policies affected are local.	Societal disorder (in time and space), irregularly covered by the media
2	Importance limited to one or two issues (e.g. territorial with cultural or heritage features but not economic or environmental ones)	The affected population is limited to one or two areas or species; the affected policies are local	Low societal concern but the topic may gain awareness
1	Low potential gravity of the issues at stake	The population affected is limited and the policies are local	Low level of societal concern currently and likely in the future

2.4 Testing the rating scale

The rating scale was included in the standard methodological guidelines for conducting ASIRPA case studies and is now implemented on a routine basis by INRA. Political impact is investigated systematically in all case studies. In 2016, 41 standardized cases were available, 23 of which had some political impact. The rating scale was used successfully on these 23 cases.

Table 5 : Marks of political impact (/5) of all 23 cases available in 2016 and scored in the rating scale

	Average	Min	Max	Nb Cases >1/5
Global political impact	3,0	1,5	4,8	23
Use in public debate (X1)	2,6	1	5	16
Use in policy-making (x1)	3,6	2	5	23
Long-term percolation of ideas (X1)	2,4	1	4	17
Societal importance of political domain at stakes (X3)	3,1	1	5	21

Political impact ratings for the 23 cases ranged from 2 to 5. The scale enables reasonable discrimination along the four dimensions of political impact. It proved relevant to judge a diversity of cases ranging from the approval of biocontrol products, to the design of non-toxic food packages. It enabled good discrimination of the type and intensity of political impact.

3/ DISCUSSION: THE IMPACT OF MEASURING POLITICAL IMPACT

Implementation of our rating scale for objectifying the political impact of science yielded good results in relation to improving self-assessment practices, and sharing value systems in INRA. However, some additional improvements and further work are planned.

3.1 Aspects of the rating scale allowing self-assessment

Using the rating scale to self-assess the political impact of case studies, based on evidence collected from stakeholders, provides a framework for judgment independent of the individual variability of the specific expert panel, and increases transparency and objectivity related to the judgments made. This rating scale can be implemented for self-assessment of the political impact for a diversity of cases. It has several advantages: it saves on cost since it avoids systematic expert consultation; it reduces delays by providing a proxy for impact to allow additional cases to be assessed without consulting an expert panel; it increases robustness since it avoids judgment bias related to different panel composition.

So far, implementation of the rating scale to evaluate the political impact of the cases produced using the ASIRPA approach demonstrates robustness for a range of different cases. However, this scale should not be considered to be definitive: regular assessment of new cases will provide opportunities for improvements. After application to 50 cases consultation with another expert panel will be used

to revise and enrich the scale if necessary, and to check the consistency and robustness of the evaluation. This second expert consultation will confirm whether the rating scale varies, or is the same, and whether the present scale would be accepted by another panel.

3.2 Using the scale to learn first lessons regarding the contribution of a PRO to public policies

Investigation of ASIRPA cases routinely involves teams of researchers, engineers, and technicians from INRA units participating in the research examined in the case study. These team members are responsible for collecting evidence of external impact from relevant stakeholders. Among the majority of researchers, the definition of political impact initially was limited to the direct contribution to policy-making. Table 5 shows that the average score for the dimension « use in policy-making », is higher than the scores for the other three dimensions, and for all cases reporting a overall political impact the minimum rating for this dimension was 2. This reveals some level of bias in the search for evidence of political impact. Table 5 shows also that the frequency and intensity of the long-term percolation of ideas seems to be lower (on average and maximum) among the 23 cases than all other dimensions of political impact. This might suggest bias related to the smaller effort expended to investigate that dimension due to problems related to collecting evidence of the contribution of research from distant stakeholders, or lack of interest in the indirect, second-level impact of INRA's research where its contribution is more dilute.

However, applying this rating scale to the 41 cases forced the evaluators to consider political impact along the four analytical dimensions described in the methodology. They found that the research could promote unexpected political impact. This broadening of the concept and definition of political impact is of interest for researchers involved in self-assessment.

Analysis of the political impact of the 41 case studies produced to date, provides lessons regarding the impact-generating mechanisms at the level of the institute. "Where, how, and by whom" bridges are built between research knowledge and policies (Almeida and Báscolo,

2006), and the nature of the contribution of INRA to public policy are better characterized. INRA's research contribution to political impact (rankings for the first three dimensions) seems to be larger if the scientific investment in the underlying research themes is long-standing, and is acknowledged by the actors. These conditions facilitate dissemination and preservation of the identity of scientific messages. These conditions are associated also with the fact that researchers bring expertise and participate in and sometimes frame political debate. The critical mechanisms linking research contributions to political impact depend on whether the work was commissioned by the public sector, or is the result of independent research. Generally, the critical points in the translation of scientific knowledge occur in the intermediary stages when dissemination of research results exploits windows of opportunity provided by political agendas, the territorial scale of the policy, or strategic distortion of scientific messages by relevant parties. The cases that result from public sector commissions have a relatively straightforward impact on the policy they are expected to affect. Even work not commissioned by the public sector but reflective of major societal concerns can have a direct impact if its results are published in a referred scientific journal (e.g., an article published in Science in 2012 on the disorientation of bees).

3.3 Features of our definition of political impact

Recall that our definition of political impact can be considered irrelevant for a minority of experts. The dissenting view is that political impact is an intermediary stage (Samuel and Derrick, 2015) before environmental, health, economic, social, or territorial, or "real-world impacts" (Cohen et al., 2015) are generated. In that respect policies and public opinion are considered to be barriers that must be overcome, or catalysts of the adoption of innovation and related massive societal changes. Following the usual practice, we consider that INRA's impact on policy does not include the impact of the policy on society (which latter is considered along other dimensions in our radar diagram, see Joly et al., 2015). We propose a definition of political impact which comprises one dimension related to the importance of the policy domain, and three dimensions related to the path from research results to political impact.

The score for political impact constitutes a snapshot of the results of a case study at a given time. Impact can evolve and the rating might increase (or decrease) over time for reasons beyond the control of the PRO.

3.4 Explaining the value assigned to political impact

The experts manifested an interest or belief in the relative value to society of the research results; their beliefs reflect the experts' personal values. For instance, there was lively debate among the experts on the importance of the focal policy domain and the significance of the research contribution to that policy. Opinions diverged among the experts, reflecting their different values. Given that different stakeholders have different interests in and expectations of

research, it is virtually impossible for their rankings of societal impact to coincide (Spaapen and Van Drooge, 2011). This clearly supports our objective of building a standalone rating scale that is independent of panel composition and free from rating bias due to individual members' interests. Discussion among panelists in consensus meetings helps to reduce rating inconsistency resulting from misinformation or misinterpretation of facts, while retaining divergence due to the experts' personal values (Wooding et al., 2014).

The method developed in this paper based on consulting a panel guided by an initial analytical framework contributes to a more transparent discussion of which items should be considered evaluation criteria, and quantified, and which items are secondary and need not be considered. The complex value system embedded in the evaluation process, and grounded on a combination of dependent judging criteria, is explained in our rating scale. The rating scale designed in collaboration with the experts reveals a set of consistent values attached to the measures of political impact. The experts scored the following aspects high:

- Integrity (accuracy, completeness, topicality) and correct sourcing/affiliation of research results and messages mediated and conveyed over time in public policies and debates. These values are facilitated by the consistent involvement of the same research actors which protects the integrity of the scientific message from possible tactical distortion. Supporting research maintained over the long-run facilitates impact by improving political credibility. It requires frequent updating of research priorities to take account of external changes and prevent decreasing relevance of results; permanence of support also facilitates the crucial involvement of researchers in knowledge mediation and downstream valorization.

- Support for the whole political cycle. Long-run research is likely to affect different steps in policy-making, from agenda-setting to assessment of policy effects, including policy formulation, decision-making, and policy implementation.

- Political relevance of the research contribution compared to the level of policy decisions. This includes the PRO's positioning and credibility in the policy domain, and the relevance of the circulation of knowledge to policy agenda setting and spatial and hierarchical decision levels. The impact on policy is higher if mediated scientific knowledge is available when decisions are being taken. Almeida and Bascolo (2006, p. 9) describe this as the "entry points for research in the policy process". A key success factor is the credibility and strength of the research messages, and the diversity of stakeholders targeted (general audience, sectoral policy-makers, political arenas, etc.

More generally, the value systems of experts which are embedded in the rating scale, raises questions regarding ideals and desirable goals for research. For example, the value attached by experts to the position of the research

in the policy cycle, and the timeliness of the research contribution, raises questions about the desirable role of a PRO: should the research or the researcher be reactive or proactive? Should the research stimulate, lead, or mitigate debate? Depending on the PRO's position (basic research or mission-oriented such as INRA), the thematic areas covered by the research, and an independent referee's perceived need for the research, experts may attach different values to the proximity between research and policy.

4/ CONCLUSION

We designed what has proved to be a successful rating scale to evaluate the contribution of research to public policy. This grid enables comparison of cases and scaling-up learning from individual cases to the level of the PRO. The rating scale have been proven to be robust, generic, standalone, and operational. It will require consolidation through regular reviews by panels of external experts. The procedure involved in the design of the scale included: 1. Designing an analytical framework based on the literature, and examining pilot cases within the framework; 2. Composing and consulting an expert panel to refine the framework, to obtain their judgment rationales, and to obtain ratings based on explicit judgment criteria; 3. Processing the resulting data to design a rating scale through the inter-ranking of scores based on related judgment criteria; 4. Expert validation of the rating scale and testing on a series of diverse cases. This procedure for building a rating scale proved efficient and generic: the ASIRPA team will consider adapting it and using it to design a similar scale for judging environmental impacts.

5/ ACKNOWLEDGEMENTS

6/ REFERENCES

Almeida, C., Báscolo, E., 2006. Use of research results in policy decision-making, formulation, and implementation: a review of the literature. Cadernos de Saúde Pública 22, S7–S19.

Alston, J., 2010. The Benefits from Agricultural Research and Development, Innovation, and Productivity Growth. OECD Food, Agriculture and Fisheries Papers, OECD Publishing 31. doi:http://dx.doi.org/10.1787/5km91nfsnkwg-en

Arnold, E., Clark, J., Muscio, A., 2005. What the evaluation record tells us about European Union Framework Programme performance. Science and Public Policy 32, 385–397.

Bell, S., Shaw, B., Boaz, A., 2011. Real-world approaches to assessing the impact of environmental research on policy. Research evaluation 20, 227–237.

Boaz, A., Fitzpatrick, S., Shaw, B., 2009. Assessing the impact of research on policy: a literature review. Science and Public Policy 36, 255–270.

Bornmann, L., 2013. What is societal impact of research and how can it be assessed? a literature survey. Journal of the American Society for Information Science and Technology 64, 217–233.

Bozeman, B., 2003. Public Value Mapping of Science Outcomes: Theory and Method, in: Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development. D. Sarewitz, Et. Al., 2.

Bozeman, B., Sarewitz, D., 2011. Public Value Mapping and Science Policy Evaluation. Minerva 49, 1–23.

Bozeman, B., Youtie, J., 2015. Socio-ÐEconomic Impacts and Public Value of Government-ÐFunded Research: Lessons From Four US National Science Foundation Initiatives. Presented at the ImpAR Conference, INRA, Paris, France.

Brewer, J.D., 2011. The impact of impact. Research Evaluation 20, 255–256.

Caplan, N., 1979. The two-communities theory and knowledge utilization. The American Behavioral Scientist 22, 459.

Carden, F., 2004. Issues in assessing the policy influence of research. International Social Science Journal 56, 135–151.

Cash, D., Clark, W.C., Alcock, F., Dickson, N., Eckley, N., Jäger, J., 2003. Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making. SSRN Electronic Journal. doi:10.2139/ssrn.372280

Cohen, G., Schroeder, J., Newson, R., King, L., Rychetnik, L., Milat, A., Bauman, A., Redman, S., Chapman, S., 2015.

Does health intervention research have real world policy and practice impacts: testing a new impact assessment tool. Health Research Policy and Systems 13, 3. doi:10.1186/1478-4505-13-3

Court, J., Young, J., 2006. Bridging research and policy: insights from 50 case studies. Evidence & Policy: A Journal of Research, Debate and Practice 2, 439–462. doi:10.1332/174426406778881764

Cozzens, S., Snoek, M., 2010. Knowledge to Policy Contributing to the Measurement of Social, Health, and Environmental Benefits. Presented at the Workshop on the Science of Science Measurement, Washington, DC.

Davies, H.T., Nutley, S., 2008. Learning more about how research-based knowledge gets used: Guidance in the development of new empirical research. W T Grant Foundation, New-York.

Donovan, C., 2011. State of the art of assessing research impact: introduction to a special issue, Research Evaluation. Research Evaluation 20, 175–179.

Earl, S., Carden, F., Smutylo, T., 2001. Outcome mapping : building learning and reflection into development programs. International Development Research Centre (IDRC), Ottawa, Ont, Canada.

Georghiou, L., Roesner, D., 2000. Evaluating technology programs: tools and methods. Research Policy 29, 657–678.

Griliches, Z., 1958. Research costs and social returns: hybrid corn and related innovations. The Journal of Political Econnomy 66, 419–431.

Hanney, S.R., Gonzalez-Block, M.A., Buxton, M.J., Kogan, M., 2003. The utilisation of health research in policy-making: Concepts, examples and methods of assessment. Health Research Policy and Systems 1, 2.

Hazell, P., Slade, R., 2014. Policy Research: The Search for Impact, in: Workshop on Best Practice Methods for Assessing the Impact of Policy-Oriented Research: Summary and Recommendations for the CGIAR, Independant Impact Assessment Report. Washington D.C.

Heclo, H., 1978. Issue networks and the executive establishment. Public administration: Concepts and cases 413.

Howlett, M., Ramesh, M., 1995. Studying public policy: Policy cycles and policy subsystems. Oxford University Press.

Joly, P.-B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., Matt, M., 2015. ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. Research Evaluation 24, 1–14. doi:10.1093/reseval/rvv015

Jones, H., 2009. Policy-making as discourse: a review of recent knowledge-to-policy literature. ODI-IKM Working Papers 37.

Kingdon, J.., 1984. Bridging Research and Policy: Agendas, Alternatives, and Public Policies, New-York: Harper Collins. ed. New-York: Longman.

Kingdon, J.W., 1995. Agenda setting. Public policy: The essential readings 105–113.

Kuruvilla, S., Mays, N., Pleasant, A., Walt, G., 2006. Describing the impact of health research: a Research Impact Framework. BMC health services research 6, 134.

Langfeldt, L., 2004. Expert panels evaluating research: decision-making and sources of bias. Research Evaluation 13, 51–62.

Lasswell, H.D., 1977. Political socialization as a policy science, in: Handbook of Political Socialization. SA Renhson, pp. 445–467.

Lasswell, H.D., Lerner, D., 1951. The policy sciences. Stanford University Press.

Lindquist, E.., 2001. Discerning Policy Influence: Framework for a Strategic Evaluation of IDRC-Supported Research. Presented at the Cases, Concepts and Connections : the Influence of Research on Public Policy; Evaluation Workshop, School of Public Administration University of Victoria, Ottawa, ON, CA.

Molas-Gallart, J., Davies, A., 2006. Toward Theory-Led Evaluation: The Experience of European Science, Technology, and Innovation Policies. American Journal of Evaluation 27, 64–82.

Penfield, T., Baker, M.J., Scoble, R., Wykes, M.C., 2014. Assessment, evaluations, and definitions of research impact: A review. Research Evaluation 23, 21–32.

Radaelli, C.M., 1995. The role of knowledge in the policy process. Journal of European public policy 2, 159–183.

Raitzer, D.A., Ryan, J.G., 2008. State of the art in impact assessment of policy-oriented international agricultural research. Evidence & Policy: A Journal of Research, Debate and Practice 4, 5–30.

Renkow, M., Byerlee, D., 2014. Assessing the Impact of Policy-Oriented Research: A Stocktaking, in: Workshop on Best Practice Methods for Assessing the Impact of Policy-Oriented Research: Summary and Recommendations for the CGIAR, Independant Impact Assessment Report. Washington D.C.

Ruegg, R., Feller, I., 2003. A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade. (Grant/Contract Report). National Institute of Standards and Technology, Gaithersburg.

Sabatier, P.A., Jenkins-Smith, H.C., 1993. Policy change over a decade or more, Westview Press. ed. CO Boulder.

Samuel, G.N., Derrick, G.E., 2015. Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014. Research evaluation rvv007.

Spaapen, J.M., Van Drooge, L., 2011. Introducing "productive interactions" in social assessment. Research Evaluation 20, 211–218.

Trostle, J., Bronfman, M., Langer, A., 1999. How do researchers influence decision-makers? Case studies of Mexican policies. Health policy and planning 14, 103–114.

Tsui, J., Hearn, S., Young, J., 2014. Monitoring and evaluation of policy influence and advocacy (Working paper No. 395). Overseas Development Institute, London.

Weiss, C.H., 1980. Knowledge creep and decision accretion. Science Communication 1, 381–404.

Weiss, C.H., 1979. The Many Meanings of Research Utilization. Public Administration Review 39, 426–431. doi:10.2307/3109916

Weiss, C.H., 1977. Research for policy's sake: The enlightenment function of social research. Policy analysis 531–545.

Wooding, S., Hanney, S.R., Pollitt, A., Grant, J., Buxton, M.J., 2014. Understanding factors associated with the translation of cardiovascular research: a multinational case study approach. Implementation Science 9, 47. doi:10.1186/1748-5908-9-47