

# Counting what really counts? Assessing the political impact of science

**A. Gaunand, L. Colinet, P.-B. Joly & M. Matt**

**The Journal of Technology Transfer**

ISSN 0892-9912

J Technol Transf

DOI 10.1007/s10961-017-9605-9



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Counting what really counts? Assessing the political impact of science

A. Gaunand<sup>1,2</sup> · L. Colinet<sup>3</sup> · P.-B. Joly<sup>2</sup> · M. Matt<sup>4</sup>

© Springer Science+Business Media, LLC 2017

**Abstract** The production of scientific knowledge is expected to benefit society in a variety of ways. However, and despite the many theoretical models available in the literature, there are few practical frameworks for assessing the dimensions of the effect of research on society—including its political impacts. As part of the ASIRPA approach, in this paper we propose an ordinal rating scale to allow assessment of the political impact of research, based on a review of the literature, qualitative evidence of political impact gleaned from a collection of case studies, and an expert panel. The resulting metric uses a 1–5 scale to evaluate the intensity of the political impact of research according to generic criteria associated to each rating level. Routine application of this scale in case study research is increasing, and is allowing robust, simple and consistent self-assessments of the political impacts across cases, to complement qualitative case study descriptions. The methodology used to design the rating scale prompted the panel experts to reveal their evaluation rationales and justify their judgments, increasing the transparency of the assessments. We believe that the benefits of assigning an ordinal measure to the political impact of research outweigh the risks of misuse of an impact number. The advantages include influencing political agenda-setting by showing what really matters, the opportunities it provides for scaling-up analyses of multidimensional impacts and identifying impact-generating mechanisms, and learning about and promoting discussion of the value systems reflected in the assessment.

---

This title makes an implicit reference to the famous quote (unduly) attributed to Albert Einstein: “Not everything that can be counted counts, and not everything that counts can be counted.”

---

✉ A. Gaunand  
Ariane.gaunand@inra.fr

<sup>1</sup> INRA, Delegation for Evaluation, 147 rue de l'Université, 75338 Paris Cedex 07, France

<sup>2</sup> LISIS, ESIEE Paris, CNRS, UPEM, Université Paris Est, 77454 Marne-la-Vallée Cedex 2, France

<sup>3</sup> Directorate College, INRA, 147 rue de l'Université, 75338 Paris Cedex 07, France

<sup>4</sup> INRA, CNRS, Grenoble INP, Univ. Grenoble-Alpes, GAEL, 38000 Grenoble, France

**Keywords** Ex-post evaluation · Political impact of science · Metric · Panel experts · ASIRPA

**JEL Classification** H43 · H83 · O33 · O38 · A13

## 1 Introduction

In recent years, research on impact assessment (RIA) has burgeoned, and a progressively growing interest in assessment of the societal impacts of research has emerged. This has resulted in a number of academic and non-academic publications focusing on the one hand on the need to understand impact-generating mechanisms as complex processes involving productive interactions, and, on the other hand the need to go beyond scientific and economic impacts and to consider the diversity of research impacts (environmental, social, political, etc.) seriously (Bornmann 2013; Bozeman and Sarewitz 2011). Overall, it would not be an overstatement to say that the recent focus on societal impact is a paradigm change. These recent approaches of societal impact attempt to enter the black box of impact generating mechanisms in order to learn from the assessment rather than just to account for external control. The integration of impact case studies in the 2014 Research Excellence Framework (REF) exercise is a landmark example of this shift (Hill 2016). The approaches to impact have been discussed by major European organizations including Science Europe (the association of European research performing organizations and research funding agencies) and League of European Research Universities (LERU) which have published Position Papers on the topic.

In this changing context, one issue that arises is the extent to which it is necessary to quantify impacts. On the one hand, there could be concern that putting numbers on complex issues such as political, social or environmental impact could lead to a narrow and reductionist approach. On the other hand, since decision makers trust numbers (Porter 1996), failure to provide any type of quantification could result in these impacts being overlooked. In this paper, we are concerned more about the latter possibility. Having acknowledged that there are sound reasons to conduct quantification of impact, the fundamental issue is how to design an impact metric which rather than leading to a reductionist approach provides an enriching heuristic. Quantification technologies have different social and political properties (Power 1994), and it is possible to design assessment systems that reduce the risk of their misuse (Paradeise 2012; Stirling 1997).

This paper proposes a metrology framework to quantify the impact of research on policy making. This study is part of the ASIRPA project which aims to design an approach to the ex post assessment of the societal impact of research (Cf. Insert 1). This case-study based approach has three main objectives: (1) to improve understanding of the mechanisms that generate research impact, in order to foster learning; (2) to adapt impact assessment to various institutional scales (research team, research division, research organization) through cross-cutting case analysis; and (3) to account for the multidimensionality of impact. In this context, the introduction of a metric complements qualitative case study description, and allows comparison of impact intensities across cases thereby enriching the qualitative case study data. This opens a new direction for learning via transversal or typological analysis of impact pathways in the cases examined

in ASIRPA. To achieve this requires all the cases to be rated using the same metric and at reasonable cost.

---

### Insert 1 The ASIRPA approach

---

The ASIRPA project developed a methodological approach based on standardized case studies to qualify and quantify the ex-post socioeconomic impacts of the research results generated by scientists in the French National Institute for Agricultural Research (INRA). A report and an executive summary (5–10 pages) were written for each case studied. The standardization of case studies combines standardized narrative description of the different steps in an impact pathway (productive configuration, outputs, intermediaries and impacts) with three analytical tools (chronology, impact pathways and vector of impacts). The analysis is based on a theory of impact inspired by innovation studies, and more particularly, actor network theory (Callon 1986). The ASIRPA team was meticulous in its design of the approach which is now implemented routinely by INRA research departments (41 case studies had been completed by mid-2017).

A first objective of the ASIRPA methodology (Joly et al. 2015) is related to learning. To understand the mechanisms that generate impact, each case describes the actor network that is mobilized, the contribution of each actor, the diversity of the impacts produced and the critical points. Standardization of a sufficient number of case studies allows systematic codification of each case study variables, and the building of a typology of impact pathways (Matt et al. 2017) which highlights generic lessons.

ASIRPA aims also at a fine description and robust assessment of the impacts—and notably their non-economic dimensions. We considered five categories of impact (economic, political, environmental, social–territorial and health) most commonly studied in the literature (Bornmann 2013) and which are significant in the context of INRA's missions. All five categories are explored systematically in each of the case studies in order to capture unexpected impacts. Thus, political impacts are envisioned even for non-policy-oriented research. Local impact descriptors for each of the categories were collected via desk research and stakeholder interviews. Evidence of impact is reported in a table which is accompanied by a radar diagram depicting impact intensity along the five categories which are scored on a scale from 1 (negligible impact) to 5 (major impact).

---

The development of this assessment system responds to some of the shortcomings in the current literature. There are no robust and reliable methods for assessing other dimensions of societal impact including political impact (Bornmann 2013). On the one hand, Renkow and Byerlee (2014) note that the CGIAR and ACIAR methods for quantitative assessment 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) agree, suggesting that quantitative political impact assessment methods are scarce: only around 10 of 31 impact studies of policy research commissioned by CGIAR, provide quantitative impact assessments, and these do not cover all the possible dimensions of political impact. On the other hand, qualitative case study based approaches provide rich descriptions of the contribution of research to the policy process (Boaz et al. 2009) but often do not allow analytical scaling-up since generally, studies are grounded in individual, one-off theoretical frameworks.

Conceptual developments in knowledge exchange have proposed frameworks that attempt to organize what is known about the ways research can contribute to policy. Davies et al. (2015) reported 71 substantial literature reviews on this topic across the areas of health, social care and education. However, on this same topic (Cozzens and Snoek 2010) comment that “the literature on knowledge to policy is long on models and short on measurements”. The literature suggests that to assess the full impact of research, notably the informative role of knowledge, it is important to examine how the policy evolved over a 10-year period (Cozzens and Snoek 2010). In relation to the implementation of the

existing models, performance reports on policy-oriented research programs tend only to identify instrumental uses of research information in a limited range of application over a short time period. This is because program evaluations usually do not have the time and resources for depth analysis, as they are conducted soon after a program is completed. Boaz et al. (2009) and Greenhalgh et al. (2016) agree; apart from a few initiatives, notably in the health sector (Boaz et al. 2009), the literature provides no well-developed, comprehensive practical frameworks (Almeida and Bascolo 2006; Cozzens and Snoek 2010), and no ready-to-use rating scales for assessing political impact (Raitzer and Ryan 2008). A systematic review of the literature (including gray literature) on impact frameworks published between 2005 and 2014 (Greenhalgh et al. 2016) found that only four approaches<sup>1</sup> had proven sufficiently flexible and robust to become established in empirical studies. However, none of them combines all the attributes we are interested in: consistent assessment of the impact of a growing sum of activities across time, rich description of the iterative process linking research and policy development, a non-labor intensive framework whose costs would not prohibit its application to all case studies.

In this paper, we build an ordinal rating scale for the assessment of political impact based on a review of the literature, qualitative evidence of political impact gleaned from a collection of case studies, and an expert panel, that enables objectification of assessment of the intensity of political impact. Our aim was to build a rating scale that is consistent, comprehensive and sufficiently general to allow it to be applied to evaluate all cases that might possibly generate political impact. For building this rating scale, we drew on the judgment of experts who intervened in an iterative process comprising of several individual and collective work.

We first review the main works and the issues addressed in the literature on political impact assessment (Sect. 2). On the basis of these key conceptual and practical issues, we describe the steps taken in the ASIRPA approach to design an original and operational methodological tool to assess the political impact of research (Sect. 3). Section 4 discusses how this approach overcomes some of the limitations in earlier works, and provides some insights into the organizational learning gained from feedback on its implementation in INRA. Section 5 concludes.

## 2 A practical framework for assessing the political impact of research

Two key issues emerge from the recent literature on political impact assessment. First, conceptual delineation of the scope of political impact is the subject of debate. Second, science and its impact on policy have for long been studied using a linear process. Recent conceptual frameworks to account for the complexity of the process are affecting how the contribution of research to political impact should be assessed. Yet very few practical methods derived from recent conceptual models of knowledge flows have been implemented to assess the political impact of research programs or institutions.

### 2.1 Delineating the political impact of research

It is important to start by providing a conceptual definition of the political impact of research (Bornmann 2013; Brewer 2011; Donovan 2011). Some authors (Hazell and Slade

---

<sup>1</sup> Payback, Research Impact Framework, Canadian Academy of Health Sciences, Monetization. The author later added the REF, which is used widely but postdated the review.



2014; Renkow and Byerlee 2014) distinguish between the contribution of research which results in a policy outcome (often a regulatory change), and the contribution of research which results in the downstream impact of policy outcome (changes to societal indicators such as improved health or housing). It is difficult to piece together the contributions of research in relation to the downstream effect of policy since the effect of research knowledge is often too diluted 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 constitutes an impact on policy and how it should be assessed, is illustrated by 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 of the evaluators perceived impact as being achieved only in the presence of some final and positive change on society resulting from policy implementation. For example, those evaluators would discard the inclusion of research results in policy documents as ephemeral measures that provide no long lasting benefits for society. Although the majority of the 62 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, and to guide the type of impact data which should be collected. Following Carden (2004) view of the assessment of political impact, we consider that INRA’s impact on policy does not include the impact of the policy on society. This view is in line with the payback framework (Hanney et al. 2003; Wooding et al. 2014), and with Samuel and Derrick (2015) who consider that political impact is an intermediary stage before environmental, health, economic, social, or territorial, or “real-world impacts” (Cohen et al. 2015) are generated. In that respect, policies and public opinion can be considered barriers which must be overcome, or catalysts of the adoption of innovation and related massive societal changes. This strategy prevents double-counting of the effects of policy implementation reported in another branch of the ASIRPA impact radar (see Joly et al. 2015).

## 2.2 Taking stock of the 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, the research results used for policy-making and how they are used, and the process involved in political impact—the transformations that occur and their timing (Davies and Nutley 2008). There are different typologies that describe the historical development of models of the interaction of knowledge and policy. In general, they begin with one-way, linear concepts and move to more interactive networked and systems views (Cozzens and Snoek 2010).

The first conceptual framework proposed to understand the qualitative contribution of research to policy as an ideal, linear, policy cycle model based on four sequential stages (Howlett and Ramesh 1995; Lasswell 1977; Lasswell and Lerner 1951): problem identification and agenda setting, policy formulation and adoption, policy implementation, and policy evaluation and reformulation. This linear depiction of policy-making assumes that policy-makers are perfectly rational actors, and that the knowledge produced by science if correctly “packed” and “disseminated” can be used directly (instantly) by decision-

makers (Caplan 1979). However, this vision which often continues to permeate the way policy research projects are conducted, has been challenged from many angles.

A first development considers the intermittent attention of policy makers. Weiss's (1977, 1979) contribution which has been taken up by contemporary authors, highlights the importance of context (societal concerns, changes to the economic context) on the timing and flow of policy-making, and the use of research knowledge. More recent evaluation approaches also emphasize the policy-making timeline (incremental vs. sudden policy changes) which is given insufficient weight in linear models (Cozzens and Snoek 2010; Jones 2009; Weiss 1979). These insights spurred work on the contribution of research to policy changes from the perspective of the "windows of opportunity" to use research results (Kingdon 1984; Lindquist 2001). Scientific knowledge is more likely to become the basis for policy-making if it matches societal or sectoral concerns, or political preoccupations (Kingdon 1984).

Another development posits 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 (Cash et al. 2003; Cozzens and Snoek 2010; Weiss 1980). The flow of knowledge into policy depends on the explicit efforts of various actors and on existing power structures. These models suggest that it is necessary to pay attention to the whole process: the conditions in which the knowledge is produced, the contributions made by other actors and sources of knowledge, the interactions between researchers and end-users, and the roles of intermediaries in circulating and mediating the research results.

A third development considers the use of knowledge by political stakeholders. Empirical studies show that 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 (Hecló 1978) or the strategic positions of actors (Davies and Nutley 2008), and over time, can 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 to the use of research in policy. These paths have been clustered into 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; and conceptual use of knowledge to enable a deeper understanding of issues and policies. According to Greenhalgh et al. (2016) citing the results of Amara et al. (2004), research evidence is more often used conceptually (for general enlightenment) or strategically (to justify a chosen course of action) than instrumentally (to feed a particular policy decision). This is because research is more likely to reveal the complexity of a phenomenon than to provide a simple solution that can be fed directly into policy.

Despite the multiple conceptual frameworks proposed in the literature, recent reviews highlight that the practical frameworks used for impact evaluation usually do not rely on rich analysis of the policy processes related to the three developments reported above; With the exception of a few cases in the areas of health and international development (Boaz et al. 2009), the frameworks in the literature fail to explore the diversity of the possible contribution made by research knowledge to policy-making, and instead pay excessive attention to short term instrumental impacts on policy (Almeida and Báscolo 2006; Boaz et al. 2009; Tsui et al. 2014). A recent literature review (Greenhalgh et al.



2016) shows that very few of these frameworks have been widely used. Also, they do not allow routine comparison of different cases investigated independently, nor allow learning regarding generic impact generating mechanisms at the level of the organization. The rating scale we build tries to account for these shortcomings.

### 3 The methodology developed

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

- To match impact scores to generic criteria which are relevant to the diversity of political impacts reported in a range of cases;
- To build a sufficiently explicit scale to allow objectified self-assessment by the researchers involved in the cases on the basis of information collected from interviews with stakeholders;
- To build a scale which does not require further involvement of expert panels in order to assess individual cases. Regular external evaluation using expert panels may still be needed to (1) check consistency of use of the scale, and (2) solve emerging issues not addressed by the scale.

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

1. Drawing on the literature to design an analytical framework based on the dimensions of the potential contributions of the research to policy-making, and the associated impacts.
2. 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 were based.
3. Processing the data to derive the criteria used by experts to form their judgments, and to hierarchize them collectively.

#### 3.1 Building an analytical framework to analyze the political impact of cases

The literature posits that the research–policy nexus involves complex processes that defy simple analysis (e.g. Cozzens and Snoek 2010; Greenhalgh et al. 2016). However, to be feasible, an evaluation framework necessarily requires some structure and simplification. The design of our analytical grid drew on the literature (Sect. 2.2) and on inductive analysis based on our case studies.

In order to take comprehensive account of the different ways that knowledge may impact on policy making, we consider three main dimensions that are in line with the research exploitation paths identified by Almeida and Báscolo (2006): (1) contribution of research results to policy negotiation and actors' positioning; (2) instrumental use in policy-making; (3) long-term effects on the terms of debate. We add a fourth dimension which is the societal importance of the policy domain as highlighted in Kingdon (1984). Each of these dimensions was detailed in sub-dimensions presented in Table 1.

We used as pilots the five case studies reporting potential political impact, available at the time of the research. These five pilot cases<sup>2</sup> constitute the empirical basis for our rating

<sup>2</sup> 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 fighting scrapie in sheep. A summary of these 30 page case study reports is available at: [http://www6.inra.fr/asirpa\\_eng/Method-and-Cases/Case-studies](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 <i>inspired by Almeida and Báscolo (2006) strategic use of knowledge to influence actors' positions</i>	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 ( <i>insights from intermediary activities reported in the pilot cases</i> ) Intensity and quality of debate (Cash et al. 2003; Davies and Nutley 2008)
Use for policy-making <i>inspired by Almeida and Báscolo (2006) rationale and instrumental use of knowledge to formulate policies or support decision-making</i>	Stages of the policy cycle affected: agenda-setting, and formulation, implementation, and evaluation of policies (Cozzens and Snoek 2010) Territorial scale of policies ( <i>insights from territorial embeddedness reported in cases</i> ) Relevance and novelty of the solution provided for policy (Cash et al. 2003, and <i>insights from the research outputs depicted in the pilot cases</i> )
Long-term percolation of ideas <i>inspired by Almeida and Báscolo (2006) conceptual use of knowledge to deepen the understanding of issues or policies</i>	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; Weiss 1980; and <i>insights from cases on scaling-out of impacts</i> ) Long-term relevance of ideas and non-distortion of messages (Weiss 1980, and <i>insights from scaling-out and scaling-up effects reported in second-level impacts of pilot cases</i> )
Societal importance of the policy domain at stakes <i>inspired by Kingdon's (1984) problem, policy and political streams</i>	Potential gravity and systemic aspects of stake (Renkow and Byerlee 2014, and <i>insights from the pilot case on the sheep scrapie sanitary crisis</i> ) Magnitude of the policy and affected population (Renkow and Byerlee 2014) Societal concern (Kingdon 1984; Lindquist 2001; and <i>insights from the pilot case on the sheep scrapie sanitary crisis</i> )

scale design. We formatted the pilot cases using the analytical grid: evidence of impact (i.e. local qualitative and quantitative descriptors of impact collected through interviews with stakeholders and desk research) was extracted from the case reports, and sorted to match the corresponding analytical dimensions. This step confirmed the sub dimensions identified in the literature, but led to some additional ones (e.g. Intensity and quality of media coverage, Territorial scale of policies).

This broad analytical framework covers the broad policy-making process literature, and complements the complex innovation theories that underpin research impacts by focusing on actors' contributions, productive configurations, knowledge translation and impact-generating mechanisms (Joly et al. 2015).

## 3.2 Using expert judgments to assess the political impact of research

### 3.2.1 Rationale for using expert judgment

Expert panels traditionally are involved in academic career assessment and with some caveats can be very useful for assessing political impacts. Expert panels are used commonly to evaluate research projects and programs on the premise that they provide the process with status and credibility (Arnold et al. 2005; Boaz et al. 2009). These methods are most frequent in the context of peer review 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 in contexts where information is not easily available. For example, panels of researchers, managers and stakeholders are used frequently to evaluate the societal impact of European Union research programs (Georghiou and Roesner 2000; Molas-Gallart and Davies 2006); they take the form of what Ruegg and Feller (2003) call a “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 higher representation of stakeholders and end-users (Bozeman and Youtie 2015).

The literature provides recommendations on panel review to promote objectivity and diversity in the opinions expressed in the assessment process. Prior 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). Also important is the selection of panelists; the size and composition of the panel must ensure constructive discussion among members. For example, Arnold et al. (2005) and Langfeldt (2004) observed that in panels with little overlap in competences, the group will follow the lead of a few panelists 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 require 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 often are not explicit about how their judgments are made, and traditionally act as “the authority”. This hinders transparency and consistent evaluation (Langfeldt 2004; Samuel and Derrick 2015). Experts should be encouraged to discuss their evaluation rationales.

Building on this experience, our decision to consult an expert panel was adapted to our goal of transparent, consistent judgment. Following the recommendations in the literature, we constructed a panel taking account of panel size, seniority of panelists, and overlapping competences.

### 3.2.2 Implementing the expert judgment: the meeting of the expert panel and the building of a rating scale

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 their ratings.

We selected five experts on the basis of their experience in assessment of the impact of research on policy-making. All were French nationals with a research background. We chose a small panel to promote strong interactions and intense deliberations but account for overlapping competences. Three experts were from the three ministries that were the potential users of the research knowledge produced in the case studies; they were appointed from

divisions dealing with knowledge to public policy. The other two experts were involved in studying the societal impact of research. All except one 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, were of similar seniority, and were not linked within a hierarchical relation.

Experts were provided with comprehensive and easily exploitable data, namely executive summaries and political impact analytical tables (=Table 1) prefilled with the evidence collected for the five pilot cases.

Each expert committed to remote individual preparatory work. For each case, each expert was asked to rate each of the four dimensions of the analytical framework on a 1–5 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 the experts were to discuss desirable evaluation criteria.

At first sight, some of the experts were reluctant to act as authorities in rating the political impact of cases they had not assessed. However, since case studies were only a supporting material, and since the objective of the rating task was to obtain a rationale for judging impact and, ultimately design a rating scale, they agreed to undertake the work.

The objective of the meeting was not to achieve consensus on a single mark for each case but to elicit each expert's rationale for the rating he awarded. The experts were invited to a 1-day meeting. They were invited to comment on the marks assigned to each case study, dimension by dimension, and compare the merits of case. The size and composition of the panel proved successful for promoting expression of a variety of opinions in a direct and dynamic exchange. The experts validated the subdivision of political impact along the four dimensions in our initial analytical framework (Table 1). The arguments put forward 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. This was an iterative process in which the experts alternatively consulted and referred to the case and the grid. The individual arguments were debated and challenged, resulting in individual explanations and reformulations 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 declared themselves satisfied with the dialogue and in-depth analysis of the political impacts. The meeting resulted in (1) a rating on a 1–5 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.

### 3.3 Processing data and designing an evaluation rating scale

After the meeting, we analyzed the dispersion of the expert ratings, and the 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, the overall ratings were coherent since the experts tended to agree on the assignment of a high or low score, and justified it using 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 constituted a set of evidence on which the rating was based (see Table 2).

**Table 2** Examples of the arguments proposed by the experts on three cases (out of the five pilot 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
<i>Case alert on the dangers of Bisphenol A (BPA)</i>					
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)
<i>Case scientific public expertise on pesticides (pesticides)</i>					
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

**Table 2** continued

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
<i>Case genetic fight against scrapie (scrapie)</i>					
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

Table 2 reports a score and matching arguments for each of the three pilot cases exemplified, for each expert, and for each sub-dimension. Based on collective challenging of these individual arguments, we derived generic evaluation criteria related to each sub-dimension (the criteria can be seen in the final rating scale in Tables 3, 4, 5, and 6). These generic criteria are derived from the experts’ combined arguments.

To order the evaluation criteria we used a Condorcet method to crank the cases collectively based on each expert’s individual rankings.<sup>3</sup> A 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). Insert 2 reports the implementation of the Condorcet method on the three cases which are those presented in Table 2. 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 distinguish between two cases, we awarded them the same mark.

<sup>3</sup> Another method called order 1 statistical dominance was used to triangulate the Condorcet method. It yielded initially coherent information for the case study ordering.



**Insert 2** Implementation of the Condorcet collective re-ranking method on the experts individual scores of three cases on the dimension “use in public debate and policy negotiation”

Expert	BPA	Pesticides	Scrapie
A	4	4	3
B	3	4	3
C	3	4	2
D	5	3	3

Then

For expert A: Scrapie < BPA = Pesticides

For B Scrapie = BPA < Pesticides

For C: Scrapie < BPA < Pesticides

For D: Pesticides = Scrapie < BPA

The Condorcet pairing process identified for the dimension “use in public debate and policy negotiation”, Scrapie < Pesticides = BPA

This collective consensus re-ranking procedure was performed for each of the four political impact dimensions. Since each expert score was associated with the expert’s arguments this Condorcet collective re-ranking of cases led to a hierarchy of evidence regarding the evaluation criteria.

## 4 Analysis: the rating scale released

This process of criteria explication, sorting and hierarchizing resulted in a five-point rating scale for each dimension of political impact with each ranking associated with related evidence regarding the evaluation criteria (Tables 3, 4, 5, and 6).

The score for impact on each dimension is the average of its sub dimension scores. If no information was available in the case report for a given sub-dimension, the score was awarded 1/5.

Finally, 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 overall contribution with the importance of the policy, in order to avoid a bias that would reward an important contribution to a minor local policy and discourage a small contribution to a global policy challenge.

### 4.1 Testing the rating scale

The rating scale was sent to the experts for validation. This external validation completed the iterative consultation process comprising: preparation of pilot cases by the ASIRPA team; remote individual ranking work by experts on the basis of the material provided; a meeting allowing experts to exchange arguments; interpretation and analysis by the ASIRPA team; final email validation by the experts.

**Table 3** Dimension Use in public debate and policy negotiation

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

**Table 4** Dimension Use for policy-making

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

**Table 5** Dimension Long-term percolation of ideas

	5	4	3	2	1	Total
Importance of knowledge in debate	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	Probable contribution to changing the terms of important debate Few changes to professional identities or positions of institutions but developing internal debate over disputed questions	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	Little change expected to the terms of debate (ideas already outdated) but existing positions reinforced by scientific status Knowledge may be used in other policy areas	No changes expected in the terms of debate (ideas already outdated) No possibilities for knowledge to be used in other policy areas	/5
Circulation of ideas in later debates/studies, and broader spheres	Rapid and sustained percolation of ideas to broad sectoral, scientific, and global audiences at the national and international levels	Sustained but slower and more uneven percolation of ideas. Sectoral, scientific and general audience, and the national and international levels are alternatively affected	Sustained but slow and very incomplete percolation: affecting only one audience (scientific/sectoral/general), at the national level	Slow but quite sustained percolation of ideas in scientific arenas	Punctual, opportunistic percolation of ideas	/5

Table 5 continued

	5	4	3	2	1	Total
Long-term relevance of ideas and non-distortion of messages	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 activities	Low risk of distortion over the long-term thanks to the PRO's and individual researchers' reputation, sustained involvement in research and intermediation; but ideas poorly traceable due to their origins in an individual researcher rather than the PRO	Moderate risk of distortion over the long-term: PRO's and individual researchers sustained involvement in research and intermediation; but PRO has no scientific excellence in the domain at stake and ideas poorly traceable	High risk of distortion over the long-term: Sustained involvement of individual researchers in intermediation but no sustained funding of research; the PRO is not scientifically renowned in the domain; ideas poorly traceable	Very high risk of distortion over the long-term: Punctual involvement of the PRO and its researchers in research and intermediation; PRO is not scientifically renowned in the domain; ideas poorly traceable. The relevance of ideas already contested	/5
	Strong relevance/resilience of ideas over the long run	Rapid obsolescence of ideas	Rapid obsolescence of ideas	Rapid obsolescence of ideas		Total average/5

**Table 6** Dimension Societal importance of the policy domain at stakes

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



**Table 7** 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

The rating scale was included in the standard methodological guidelines for conducting ASIRPA case studies and is being implemented routinely by INRA. Political impact is investigated systematically in all case studies. In 2016, 41 cases were available, 23 of which had some political impact. The rating scale was applied successfully to all these 23 cases.

Table 7 shows that the political impact ratings for the 23 cases range from 1.5 to 4.8. The scale enables reasonable discrimination along the four dimensions of political impact. It proved relevant to judge a diversity of cases ranging from approval of biocontrol products to design of non-toxic food packaging. It enabled good discrimination of the type and intensity of political impact.

## 5 Discussion: the impact of assessing political impact

Despite the many instances of misuse of quantification methods illustrated by different cases (including the use of scientometric for assessment of research quality) we believe that the benefits provided by our rating scale, which was constructed mainly to increase learning about the research–policy nexus, outweigh any risks of its being used inappropriately. History shows that the power of numbers is high (Porter 1996), and that counting what counts (e.g. biodiversity, climate change, etc.) is instrumental to setting agendas and triggering/designing action. Robust quantification of societal impacts allows case comparison, and analysis of their diversity and multidimensionality. In addition, this ordinal measure allows political impact to be analyzed as a dependent variable of the impact pathway, opening the way for an elaborated understanding of the systematic patterns of impact pathways (see Matt et al. 2017 for a preliminary work in this area). Quantifying societal impact can serve learning purposes and promote an open debate of the value systems reflected in the assessment.

In this discussion, we argue that the risks of misuse are limited by the intrinsic properties of the rating scale. We argue also that the extrinsic features, namely the overall design of the research assessment as a learning tool, are essential for producing the expected benefits.

### 5.1 The intrinsic properties of the rating scale

Since different stakeholders have different commitments and interests, it is likely that their rankings of societal impact of research will not coincide (Spaapen and Van Drooge 2011). In this paper, the design of a rating scale for policy impact was aimed at providing a robust assessment that produces consistent results, but does not claim to be universal. The

assessment is consistent since application of the tool by different trained assessors yielded the same results. However, we would not claim that it produced an objective assessment considered as a *view from nowhere* (Nagel). Another rating scale (based on another theory of impact, other weighing criteria, etc.) would probably yield other scores, and it is likely that the proposed rating scale will—similar to other systems of assessment—evolve over time. Instead of providing a very precise measure whose achievement would be very costly, we chose a simple ordinal measure. This choice reflects both the modest objectives of the scale, and the pragmatic constraints on its use.

The process that was used to design this rating scale prompted the experts to reveal their evaluation rationale and to explain the founding values of their judgments. The initial individual remote scoring revealed how each expert interpreted the analytical framework and the local descriptors of impact collected for each case. The argumentation each expert offered to support her/his impact scores, and the debate that ensued, contributed to a clear explanation of the rationale for the choice of items to be taken into account.

The divergences in the expert opinions were particularly meaningful in terms of the values conveyed. All the experts showed a real interest in the relative value of the research results to society; diverging views reflected the experts' personal values. For instance, there was lively debate among the experts about the importance of the policy domain considered, and the significance of the research contribution to that policy. The discussion among the panelists helped to reduce rating inconsistency resulting from misinformation or misinterpretation of facts (Wooding et al. 2014).

The complex value system embedded in the evaluation process, and grounded in a combination of dependent judging criteria, is outlined in our rating scale. The experts valued the following aspects:

- Integrity (accuracy, completeness, topicality) and correct sourcing/affiliation of research knowledge combined when mediated and conveyed over time in public policies and debates;
- Support for the complete policy cycle;
- Political relevance of the research contribution compared to the level of policy decision. A key success factor is the credibility and strength of the research message, and the diversity of the stakeholder audience targeted (general audience, sectoral policy-makers, political arenas, etc.).

Such a rating scale 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 the further involvement of an expert panel; it increases robustness since it avoids judgment bias related to different panel composition. The rating scale is intended to be improved over time. Regular assessment of new cases, and changes in global political concerns will present opportunities to revise and enrich the rating scale. Hence, we envisage further consultation of experts to review the rating scale and suggest improvements.

## 5.2 Counting as one element of impact assessment as a learning tool

The rating scale presented in this paper is not meant as a stand-alone assessment. It is part of the ASIRPA approach that is framed by a general philosophy of action. Power (1994) distinguishes two evaluation ideal-types. The first is audit-like: it is framed by accountability objectives and aims at external control, and is based on quantitative unidimensional measures. The second is assessment-like: it is oriented toward learning objectives,

proceeds through self-evaluation and is based on multidimensional qualitative and quantitative assessment. ASIRPA belongs to this second family of evaluation approaches. This philosophy of evaluation is reflected in the ways the rating scale is used, as illustrated by a brief account of our experience.

The investigation of ASIRPA cases involved researchers, engineers and technicians from INRA units who had contributed to the research examined in the case study. They were responsible for collecting evidence of external impact from relevant stakeholders. Initially, most of the researchers considered political impact only through the direct contribution of knowledge to instrumental policy-making. Application of the ASIRPA rating scale to the 41 cases prompted them to consider political impact along the four analytical dimensions described in the methodology. They found that research can promote unexpected political impact. This broadening of the concept and definition of political impact ascribes a performative property to our research: it opened new perspectives for the researchers involved in the self-assessment. This can be seen as a first step towards changes to research practices.

Analysis of the political impact of the 41 case studies conducted by INRA 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 INRA’s contribution to public policy are better characterized. INRA’s research contribution to political impact (rankings for the first three dimensions) seems to be larger when 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 into policy occur when dissemination of research results confront windows of opportunity provided by political agendas, relevant territorial scale of the policy, or strategic distortion of scientific messages by relevant parties. The cases that result from public sector commissions can have a relatively straightforward impact on the policy they are expected to affect. However, work not commissioned by the public sector but reflective of major societal concerns can also have a direct impact if its results are published in high impact scientific journals (e.g. an article published in *Science* in 2012 on the disorientation of bees) and prompt media and political attention.

These are some of the lessons that can be drawn from the assessment at the different organizational levels (Researcher involved, Research Unit, Research division, Research organization) and which contribute to learning processes and to the promotion of an impact culture.

**Acknowledgements** The analysis was made possible by financial support from INRA (France) through the ASIRPA Project (Socio-economic Analysis of the diversity of Impacts of Public Agricultural Research).

## 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.

- Amara, N., Ouimet, M., & Landry, R. (2004). New evidence on instrumental, conceptual, and symbolic utilization of university research in government agencies. *Science Communication*. doi:[10.1177/1075547004267491](https://doi.org/10.1177/1075547004267491).
- Arnold, E., Clark, J., & Muscio, A. (2005). What the evaluation record tells us about European Union Framework Programme performance. *Science and Public Policy*, 32(5), 385–397.
- Boaz, A., Fitzpatrick, S., & Shaw, B. (2009). Assessing the impact of research on policy: A literature review. *Science and Public Policy*, 36(4), 255–270.
- Bormmann, 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(2), 217–233.
- Bozeman, B., & Sarewitz, D. (2011). Public value mapping and science policy evaluation. *Minerva*, 49, 1–23.
- Bozeman, B., & Youtie, J. (2015). Socio-economic impacts and public value of government—Funded research: Lessons from four US National Science Foundation Initiatives. In *Présenté à ImpAR conference*. Paris: INRA. <https://colloque.inra.fr/impar/Program-Material>
- Brewer, J. D. (2011). The impact of impact. *Research Evaluation*, 20(3), 255–256.
- Callon, M. (1986). The sociology of an actor-network. In M. Callon, J. Law, & A. Rip (Eds.), *Mapping the dynamics of science and technology*. London: Macmillan.
- Caplan, N. (1979). The two-communities theory and knowledge utilization. *The American Behavioral Scientist*, 22(3), 459.
- Carden, F. (2004). Issues in assessing the policy influence of research. *International Social Science Journal*, 56(179), 135–151.
- Cash, D., Clark, W. C., Alcock, F., Dickson, N., Eckley, N., & Jäger, J. (2003). *Saliency*. SSRN Electronic Journal: Linking research, assessment and decision making, credibility, legitimacy and boundaries. doi:[10.2139/ssrn.372280](https://doi.org/10.2139/ssrn.372280).
- Cohen, G., Schroeder, J., Newson, R., King, L., Rychetnik, L., Milat, A., et al. (2015). Does health intervention research have real world policy and practice impacts: Testing a new impact assessment tool. *Health Research Policy and Systems*, 13(1), 3. doi:[10.1186/1478-4505-13-3](https://doi.org/10.1186/1478-4505-13-3).
- Cozzens, S., & Snoek, M. (2010). Knowledge to policy contributing to the measurement of social, health, and environmental benefits. Présenté à 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*. New-York: W T Grant Foundation.
- Davies, H. T., Powell, A. E., & Nutley, S. M. (2015). Mobilising knowledge to improve UK health care: Learning from other countries and other sectors—A multimethod mapping study. *Health Services and Delivery Research*, 3(27), 17. doi:[10.3310/hsdr03270](https://doi.org/10.3310/hsdr03270).
- Donovan, C. (2011). State of the art of assessing research impact: Introduction to a special issue, research evaluation. *Research Evaluation*, 20(3), 175–179.
- Georghiou, L., & Roesner, D. (2000). Evaluating technology programs: Tools and methods. *Research Policy*, 29, 657–678.
- Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: A narrative review. *BMC Medicine*, 14(1), 78. doi:[10.1186/s12916-016-0620-8](https://doi.org/10.1186/s12916-016-0620-8).
- 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*, Washington, DC.
- Hecl, H. (1978). Issue networks and the executive establishment. In *Public administration: Concepts and cases* (Vol. 413, pp. 46–57).
- Hill, S. (2016). Assessing (for) impact: Future assessment of the societal impact of research. *Palgrave Communications*. doi:[10.1057/palcomms.2016.73](https://doi.org/10.1057/palcomms.2016.73).
- Howlett, M., & Ramesh, M. (1995). *Studying public policy: Policy cycles and policy subsystems* (p. 239). 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(4), 1–14. doi:[10.1093/reseval/rvv015](https://doi.org/10.1093/reseval/rvv015).
- Jones, H. (2009). *Policy-making as discourse: A review of recent knowledge-to-policy literature* (no 5, p. 37). ODI-IKM working papers.
- Kingdon, J. W. (1984). *Bridging research and policy: Agendas, alternatives, and public policies*. New-York: Harper Collins.

- Kingdon, J. W. (1995). Agenda setting. In M. A. Cahn & S. Z. Theodoulou (Eds.), *Public policy: The essential readings* (1st ed., pp. 105–113).
- Langfeldt, L. (2004). Expert panels evaluating research: Decision-making and sources of bias. *Research Evaluation*, 13(1), 51–62.
- Lasswell, H. D. (1977). Political socialization as a policy science. In S. A. Renhson (Ed.), *Handbook of political socialization* (pp. 445–467). New York: Free Press.
- Lasswell, H. D., & Lerner, D. (1951). *The policy sciences*. Redwood City: Stanford University Press.
- Lindquist, E. (2001). Discerning policy influence: Framework for a strategic evaluation of IDRC-supported research. Présenté à cases, concepts and connections: The influence of research on public policy; evaluation workshop. Ottawa, ON: School of Public Administration University of Victoria.
- Matt, M., Gaumand, A., Joly, P.-B., & Colinet, L. (2017). Opening the black box of impact—Ideal-type impact pathways in a public agricultural research organization. *Research Policy*, 46(1), 207–218. doi:10.1016/j.respol.2016.09.016.
- Molas-Gallart, J., & Davies, A. (2006). Toward theory-led evaluation: The experience of European science, technology, and innovation policies. *American Journal of Evaluation*, 27(1), 64–82.
- Paradeise, C. (2012). Le sens de la mesure: la gestion par les indicateurs est-elle gage d'efficacité? Présenté à 9ème conférence de l'AFD/EUDN, Paris.
- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Power, M. (1994). *The audit explosion*. London: Demos.
- Radaelli, C. M. (1995). The role of knowledge in the policy process. *Journal of European Public Policy*, 2(2), 159–183.
- Raitzer, D. A., & Ryan, J. G. (2008). State of the art in impact assessment of policy-oriented international agricultural research. *Evidence and Policy: A Journal of Research, Debate and Practice*, 4(1), 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*, Washington, DC.
- 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. Gaithersburg: National Institute of Standards and Technology.
- Sabatier, P. A., & Jenkins-Smith, H. C. (1993). *Policy change over a decade or more*. Boulder, CO: Westview Press.
- Samuel, G. N., & Derrick, G. E. (2015). Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014. *Research Evaluation*, 24, 229–241.
- Spaapen, J. M., & Van Drooge, L. (2011). Introducing 'productive interactions' in social assessment. *Research Evaluation*, 20(3), 211–218.
- Stirling, A. (1997). Multicriteria mapping: Mitigating the problems of environmental valuation? In J. Foster (Ed.), *Valuing nature? Economics, ethics and environment*. London: Routledge.
- Trostle, J., Bronfman, M., & Langer, A. (1999). How do researchers influence decision-makers? Case studies of Mexican policies. *Health Policy and Planning*, 14(2), 103–114.
- Tsui, J., Hearn, S., & Young, J. (2014). *Monitoring and evaluation of policy influence and advocacy* (p. 90). Working paper no. 395. London: Overseas Development Institute.
- Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy analysis*, 3, 531–545.
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426–431. doi:10.2307/3109916.
- Weiss, C. H. (1980). Knowledge creep and decision accretion. *Science Communication*, 1(3), 381–404.
- 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(1), 47. doi:10.1186/1748-5908-9-47.