3rd International Conference

on Public Policy (ICPP3)

June 28-30, 2017 – Singapore

Panel T07P06 Session 1

Evidence-based Policy Making and Policy Evaluation

Title of the paper

Can evaluations really contribute to evidence-based policy making at government level? - the case of the French Government Modernisation Evaluations

Author(s)

Delahais Thomas, Quadrant Conseil, France, tdelahais@quadrant-conseil.fr

Lacouette-Fougère Clément, French General Secrétariat for Government Modernisation, Prime minister's office, France, clement.lacouette-fougere@modernisation.gouv.fr clement.lacouette@sciencespo.fr

Date of presentation

Thursday, June 29th

V2 – June, 17th 2017

Introduction

It is a common-sense observation that policy making does not always rely on relevant data. Of course, political choices do not need to follow what evidence apparently dictates. Yet the continuing debate also proves that even the most consequential decisions may not be 'evidence aware' (Nutley & Webb, 2000).

Evaluators are often concerned about the use of evaluation in policy making, as the amount of empirical research on this topic tends to suggest (Henry, 2004, p. 35). This is because evaluation, which can be defined as 'the systematic collection of information about the activities, characteristics, and outcomes of programmes to make judgements about the programme, improve programme effectiveness, and/or inform decisions about future programming' (Patton, 2008), is a source of evidence that is supposedly directly aimed at informing decision-making.

In this paper, we build upon an evaluation commissioned in 2016 by the French Secretary General for Modernisation of Public Action, SGMAP (KPMGQuadrant Conseil, 2017), which included *inter alia* a metaevaluation aimed at assessing the quality of 65 evaluations launched between 2013 and 2016 by the French Government, and an 'evaluation of evaluation' using a contribution analysis approach, which focused on 8 cases of these evaluations. The metaevaluation part was based on a comprehensive documentary analysis, including previously undisclosed administrative decision-making documents related to each of the 65 evaluation. The evaluation of evaluation relied on a case-study approach, including additional documentary analysis and 59 semi-structured interviews with evaluation stakeholders, within and outside the French State administration. One of the authors was involved in delivering the evaluation services while the other steered this evaluation as an SGMAP officer.

The approach we follow here consists in not only assess evaluation use, but also the mechanisms (Pawson & Tilley, 1997) through which and the reasons why evaluation has been used or not, taking into consideration individual and interpersonal reasons as well as contextual and institutional reasons.

We will draw these explanations from a general theoretical framework of public policy analysis inspired both by the tradition of sociology of decision and the theory of neo-institutionalism, that are generally called to explain observed changes in policy making:

- The first tradition, developed at a microsociological scale, insists on the key role of the actors (individually and collectively) and of the political will on decisional behaviours (Halpern, 2004). Since Herbert Simon's criticism of bounded rationality (Simon, 1945), a wide-ranging corpus of works has built an iterative conception of decision, full of pragmatism and negotiation between the actors and groups of actors involved in policies (Cohen, March, & Olsen, 1991). Charles Lindblom's complete opus is thus dedicated to the process of decision and non-decision. His theory of Incrementalism (Lindblom, 1959), once adapted to evaluation research, was seminal in thinking evaluation utilisation from the 1980s. Patton's and Weiss's empirical researches, for instance, follow and extend this approach.

- The second tradition analyses less the decision makers than the institutional contexts that make the change possible as a general and organised process. According to this sociological school of thoughts, inspired by the writings of neo-institutionalists (Hall & Taylor, 1996; Mayntz & Scharpf, 1995), political structures influence governments and the way they lead their actions. Then, what is at stake is to comprehend observed changes throughout changes in the institutional configuration. In this respect, works in terms of institutionalisation of the evaluation have proven relevant in better understanding the weight of institutional choices in the development of evaluation systems, all over the world (Dente & Kjellberg, 1988; Furubo, Rist, & Sandahl, 2002; Jacob, Speer, & Furubo, 2015; Stockmann, Meyer, & Nolte, 2017).

In this paper, we will first present the (incomplete) institutionalisation process of evaluation in France, up to the *Modernisation de l'action publique* (Modernisation of Public Action, MAP) evaluation programme, and consider how it should have affected the role of evaluations in decision-making (part 1). Then, using the metaevaluation we will identify, the attributes of the evaluation process and report that should, as per our theoretical framework, lead to stronger use in decision-making, and assess the 65 MAP evaluations accordingly (part 2). In a 3rd section, we will try to assess the actual contribution of these evaluations to decision-making, zooming on a smaller set of 8 case studies and using a set of empirical tests (part 3). In the final section, we will focus on the main identified mechanisms explaining a contribution, or a lack thereof, both at the micro and macro-sociological level, using Mark & Henry's Theory of Influence (Henry, 2004) as a bridge between these two levels (part 4).

We mainly expect to provide empirical evidence of evaluation use in the French context; contribute to the methodological debate about the actual use of evaluation; and play a part in opening the black box on mechanisms leading to evaluation use in circumstances of incomplete institutionalisation.

From 1989 to 2012: a Tortuous Path Towards Institutionalisation

Evaluation officially arose in the 1980s in France and followed an initial process of institutionalisation in the 1990s that led to a specific evaluation "à la française". At State level specifically, evaluation and modernisation have had an intimate yet complicated relation; the current MAP is the culmination of this relation.

Initial Institutionalisation of Evaluation in France

It has been evidenced before that national evaluative practices follow specific pathways which depend on the peculiarities of the political and administrative systems in which they grow (Bemelmans-Videc, Eriksen, & Goldenberg, 1994) and can reach different stages of maturity both in terms of formalism (laws, regulations, organisations) and practices (Monnier & Fontaine, 2002, pp. 63-75).

Jacob and Varone have identified 4 variables to assess the degree of institutionalisation of evaluation in 19 OECD countries (Jacob & Varone, 2004):

1. Effective practice of evaluation (number of evaluations ordered, performed, number of evaluators, budget dedicated to the evaluation…)
2. Evaluation clauses (national evaluation regulations, law…)
3. Organisational structures (isolated structures or inter-organisational structures)
4. Development of an epistemic evaluation and research community (clubs, networks, national evaluation societies, quality standards, scientific reviews, etc.)

In France specifically, evaluation remains quite a recent endeavour. At the beginning of the 1980, "evaluation in France is neither a school of thought nor an organised administrative practice" (Nioche, 1982).

In Jacob and Varone's typology, though, France was already in 2001 considered quite mature when it comes to evaluation like, among others, Australia, Canada, Denmark, Finland, Switzerland or the United States. Since then, its evaluation culture has been improving, according to Jacob's update of its initial work (Jacob et al., 2015, pp. 6-31).

Between the beginning of the 1980s and 2001, indeed, evaluation came of age in France (Barbier & Hawkins, 2012). This culture has grown on the complexification of public action, constant struggles in its management and justification, and the professionalisation of public policy analysis, especially in the private sector (Monnier et al., 1992, p. 54). But it is also the product of two driving forces that contributed to the institutionalisation of evaluation on the formal ("evaluation clauses") and structural aspects ("organisational structures"):

- First, in the 1980s emerged in France different levels of government, at supra (European Union), and infra-national (local authorities) levels. For these authorities, evaluation has been a resource of power and an instrument of legitimation (Monnier & Fontaine, 2002), and therefore has progressed in the context of vertical counter-power affirmation from the European Union and from French Regional authorities (Jacob & Varone, 2007);
- At the same time, political change brought at the end of the 1980s new actors to the fore, who joined forces on an agenda of State "modernisation" (Pollitt, 1995). Evaluation, because it focuses on the consequences of public action rather than on the administrative culture of rules and procedures (Silvent, Dureau, & Guillaume, 2002), and as a way to provide information and knowledge to decision-making processes (Thoenig, 2000), quickly made consensus among politicians and some senior State officials as an instrument of modernisation (Bezes, 2009; Duran, 2010; Leca, 1993).

This, in turn, led to further institutionalisation by the development of a market for evaluation services, while in 1999, evaluation professionals created the French Society of Evaluation.

At the beginning of the 2000s, an evaluation "à la française" as emerged, in which "emphasis is placed on the 'plurality principle', […] anchored in the [2003] French *Charte de l'Évaluation*"

(Baron & Monnier, 2003; Jacob et al., 2015). A pluralist evaluation not only associates to its process the decision maker, but also the stakeholders that are operating, are targeted or benefit from the evaluated intervention (Guba & Lincoln, 1989).

Institutionalisation at State-level: a long relation between evaluation and modernisation

At State level, evaluation is from the beginning a part of an endeavour by a number of politicians, senior State officials and consultants to "modernise" the State.

The first act of institutionalisation is a 1990 Decree which emphasised the need to respect democracy and inform citizens, but also stressed the internal purpose of evaluation as a tool for improving the effectiveness of the State (Monnier & Fontaine, 2002). It led to the creation of different structures, among which the most prominent was the Interministerial Evaluation committee (CIME), which members met periodically to validate, under the Prime Minister's authority, evaluation projects requiring the co-operation of several ministries. A National Fund for the Development of Evaluation was also created to support evaluation at ministry level.

30 interdisciplinary, large-scale evaluations, related to topics such as energy control, the fight against aids or actions towards the youth, were commissioned and led during its 10 years of existence. However, the CIME stopped being gathered in 2001. Though it deserved credit for launching a recurring evaluation practice at State level, it progressively received negative press by the ministries, who were seeing it as a complex, heavy and elitist structure, not useful given their timeframe and research function (Jacob, 2005). Besides, these evaluations had to face an ancient opposition of the French administration's *grands corps*[1], who were claiming a monopoly on State expertise and were hostile to the plurality principle (Lacouette-Fougère & Lascoumes, 2013).

In 2007, the "modernisation through evaluation theme" was relaunched in two contrasting ways.

On the one hand, as in Switzerland, France rooted the principle of evaluation into its Constitution. The constitutional reform of 2008 especially gave to the French Court of Auditors a role in the evaluation of the State public policies (Barbier, 2010). Also, a new alliance of senior State officials and politicians promoted the use of social experimentation for new public policies, accompanied by evaluation (Devaux-Spatarakis, 2014). For instance, the reform of the *Revenu de Solidarité Active* (RSA), a welfare benefit aimed at people not covered by unemployment benefit and which then concerned more than 1 million persons, started by an evaluated experimentation.

[1] The *Grands corps* are institutions within the French State, especially in charge of control and audit functions, e.g. inspection bodies, whose members are senior officials usually sharing a similar social origin, a similar academic curriculum, etc. Marie-Christine Kessler, a French specialist of the *Grands corps*, has stated that they can be "agents of reform, of stagnation, or of regression" (Kessler, 1986).

On the other hand, though, with President Nicolas Sarkozy, evaluation almost disappeared from the "modernisation" theme, replaced by management practices embodied by the General Review of Public Policies (RGPP) and implemented by the ministry of Budget. For the first time, management consultants gained access to the higher level of the State at the expense of the inspection bodies belonging to the ministries (Matyjasik, 2010). In this respect, RGPP could be seen as is a regime of "administrative exception" (Gélédan, 2012, pp. 4-5). Indeed, at the end of 2007, audits covering the main areas of State activities had been launched. What was new was that these audits were part of a binding decision process which would necessarily lead to political decisions at a fast step. Reduced spending quickly became the main criterion for decision and the related disciplines used for that (organisational audit, management control…) somehow replaced evaluation at the interministerial level at least (Barbier, 2010). But these audits, notably taking out the Parliament as well as the public policies services and operators failed in providing conditions of mobilisation for the implementation of the reforms.

The Modernisation of Public Action

That is why, in 2012, with President Hollande, the pendulum swung back. The *grands corps*, who had for some of them initially supported the RGPP, published their own evaluation of the RGPP (Bondaz, 2012). In a nutshell, the report stated that in the 2007–2012 phase the modernisation ambition was marred by inadequate method and implementation, and called for a new approach that would still aim at renewing public action, but this time with the full participation of the administration itself. At the same moment, the freshly elected left-wing government had announced a programme of "policy evaluations done in partnership with all concerned actors (State, local authorities, social security bodies and operators) to build a collective vision of the challenges, aims and implementation details for each [of the State's] public policies" (KPMGQuadrant Conseil, 2017), the *Modernisation de l'Action publique* (MAP).

The use of evaluation as the main instrument of reform has been considered as the truly original part of the MAP (Le Clainche, 2013). And indeed, the initial objective of the MAP was to evaluate all the State's public policies before 2017. For specialists of institutional systems of evaluation, such a level of ambition can be considered as an exception among countries involved in evaluation processes (Jacob et al., 2015), though in practice, this objective quickly turned into each ministry having to launch at least one evaluation within its field during the mandate.

A blueprint of the method to be used was proposed in a guide prepared by the three main inspection bodies (Finance, Administration and Social Affairs) and validated in December 2012 by the Prime Minister during the first Interministerial Modernisation of Public Action Committee (CIMAP).

The guide calls for evaluations that "do not exactly follow the evaluation model promoted by the French Evaluation Society, as they would be too complex, too human-resource intensive and would need too much time when compared to what can be done in the framework of Public

Action modernisation" (Destais, Marigeaud, Battesti, & Bondaz, 2012). However, the MAP enforces principles that are: feeding decision; favour consultation (*concertation*) at all steps of the evaluation, as well as transparency; go beyond a measurement of effectiveness and efficiency, and also consider relevance and social betterment brought by the evaluated policies; and ensuring independence. For instance, the principle of consultation leads to obligations such as each evaluation has to be steered by a committee which includes external stakeholders to the commissioning ministry, and stakeholders must be consulted during the data collection phase. In that sense, the MAP evaluations are a continuation of the *évaluation à la française* principles and, from an institutional point of view, of the 1990s CIME evaluations: it re-engages with Prime Minister Steering (instead of the ministry of Finances) and with inter-ministerial governance.

But the MAP is also a continuation of the RGPP in the requirement that evaluations should be part of a "fast-track" decision process. This was done in several ways: at political level, the Prime Minister was to sign a "letter of assignment" addressed to the evaluator, to give political weight to their work; the CIMAP would also meet to supervise the progress following each evaluation. At the institutional level, an administrative vehicle, the Secretariat General to the Modernisation of Public Action (SGMAP), was created to monitor and support this endeavour, including an "Evaluation Department" with a staff of 4 persons. At the individual-evaluation level, each study is to follow two phases: a data-collection and analysis phase (*diagnostic*), and a recommendation phased centred around *transformation scenarios*. This second part was an innovation and was aimed at ensuring that the evaluations would not only be retrospective, but would also spend enough time and resources on the changes that could be brought to the evaluated public policies.

Expectations towards the MAP evaluations were very wide, including *organisational transformations* (for instance an evaluation that would recommend merging several existing operators on a public policy), *strategic transformations* (i.e. redefinition of the global objectives of a policy, such as generic medicines or family policy) or more *spread-out transformations* (progressive evolution of a work modality in employment policy) or reshaping conditions for the grant of aid in a specific sector (Picavet, 2015).

Another innovation was that the inspection bodies actually stepped into the MAP as the sole evaluators of the State policies[2], as it was decided during the first CIMAP (Picavet, 2015). This move can be seen as a continuation of the RGPP, in which they had been involved, as well as a "revenge" against consultants who had breached into their expertise monopoly at the higher level of the State. It is also a true change given their traditional hostility to evaluation (as well as to plurality). But they perceived evaluation as a resource of power to extend and legitimate their role of inspection and control within the State (Boissier, 2015).

[2] Inspection bodies could rely on external advisory services, e.g. to design and administer surveys or to hold workshops with stakeholders.

Finally, following the theories of institutionalisation, France after this long pathway of development proved to reach in 2017 quite a high level of institutionalisation not only in terms of formalism but also in terms of real practices and professionalisation.

Summary: how does the MAP a priori affect evaluation influence?

From an institutionalisation point of view, France (and especially the State) was reaching for a higher level of maturity with the MAP, at least in terms of clauses and structures. With 80 evaluations performed between December 2012 and the end of President Hollande's mandate in May 2017 (85% of which having been released on the SGMAP website), the MAP also contributed to an increase in evaluation practice in France.

From this institutional context, what could be expected in terms of influence of the MAP evaluations?

On the one hand, these evaluations should directly contribute to the decision-making process; they should do so by providing a more open view than the RGPP style audits, especially by considering the point of view of external stakeholders; they are supposed to provide realist and relevant recommendations; inspection bodies know the internal functioning of the administration very well and are likely to fit the recommendations accordingly; and there is (at least at first) a strong political backing.

On the other hand, evaluation is now the byword for any transformation process, but some ministries may not know how to formulate an evaluation request – or they may need another type of service altogether; also, inspection bodies have stepped in, but they may not have evaluation skills and be able to play a role in ensuring that evaluation is useful (Patton, 2008).

In the following sections, we first verify whether the attributes of the performed evaluations make them likely to contribute to decision-making; then, we look at the actual contributions of these evaluations; and finally, on the mechanisms that explain these decisions, taking into consideration the initial expectations and institutional context presented above.

Attributes of the Evaluations and Consequences for Their Use in Decision-making

In this section we present the main attributes of 65 MAP evaluations (out of 80) that were terminated at the time of our research. Evaluations that were just finished and which had no possibility yet to have an influence over decision-making were also removed from our sample.

Method

In the late 1970 and early 1980s the basis for theory of evaluation use was developed around three types of usage: instrumental, conceptual and political/symbolic use (King & Pechman, 1984; Leviton & Hughes, 1981):

- *Instrumental use* refers to the direct use of evaluation findings for decision-making
- *Conceptual use* refers to the contribution of information, ideas and perspective that an evaluation can convey in the arenas in which decisions are made (Weiss, 1999)
- *Symbolic use* refers to examples when a person uses "the mere existence of the evaluation, rather than any aspect of its results, to persuade or to convince" (Johnson et al., 2009)

Michael Q. Patton popularised a fourth use called *process use* and that refers to the fact that not only evaluation results but also process of participating in an evaluation can produce outcomes or lead to learning effects (Patton, 2008).

As Carol H. Weiss has pointed out, evaluation is a "rational enterprise": "the assumption is that by providing 'the facts', evaluation assists decision-makers to make wise choices among future courses of action. Careful and unbiased data on the consequences of programmes should improve decision-making" (Weiss, 1993). Yet the circumstances under which the evaluation is decided, commissioned, performed and received are crucial to their actual use.

The literature identifies several attributes of evaluations that can be associated, under the right conditions, to their use in decision-making. Evaluation can theoretically be *tuned* for better chances of use. For instance, the object of the evaluation should be delimited upstream to correspond to an opportunity for decision (Weiss, 1983); the evaluation should be designed to be useful for the commissioner and to the other stakeholders – these should be sufficiently involved in the process (Patton, 2008); also the recommendations should be realistic and mutual adjustments between stakeholders should be sought (Lindblom, 1965).

Turning these principles into actual quality criteria is a challenge in each metaevaluation, which usually needs to be tailored to the expected use and corpus of evaluations (Coosky & Caracelli, 2009). First, there is no global standards in evaluation, but either professional standards (e.g. the American Evaluation Association Programme Evaluation standards, see (Shulha, Caruthers, & Hopson, 2010)), or organisation standards (e.g. the OECD Development Assistance Criteria). In France, specifically, the French Evaluation Society has adopted a set of principles, the *Charte de l'évaluation*, instead of quality criteria. In the end, we developed an *ad hoc* set of evaluation quality criteria corresponding to our needs, based on the American standards for the most part.

Also, in our context, we had two main constraints: the metaevaluation had to be done using desk research only (documentation included the report and paperwork pertaining to the evaluation and decision-making process); and given the sensitivity of dealing with the quality of reports written by members of the *grands corps*, it had to be unassailable. We came with a purely factual approach, which was relevant for some criteria but less for others (for instance, it was not possible to assess the evaluation competencies of the evaluators' team except by identifying whether they had been involved in previous evaluations). This is a limitation of our study. The set of indicators used builds on previous in-house experiences of metaevaluation (e.g. (Euréval, European Commission, Directorate-General for the Information Society, Media, 2011) as well as on a recent empirical research on evaluation attributes (Lloyd & Schatz, 2015).

The full results are available in the study's report (KPMGQuadrant Conseil, 2017); in the section below, we focus on these attributes most likely to influence the evaluation's contribution to decision-making, using 3 categories: evaluation approach (how the evaluation was designed for use); evaluator, its competences and its role in the evaluation and decision-making process; and evaluation outputs, including the credibility of the evaluation content (Appleton, 2012).

Results

Evaluation approach

In terms of rationale for the evaluation, we found that 95% of the evaluations were prompted by specific, well-enunciated purposes. Besides organisational purposes (63% of the motivated evaluations) and reduce spending (45%), 40% explicitly mention a future decision to be taken (e.g. a new law or a new programme adoption). This is due to a systematic initial process in which the demands of commissioners are discussed and then written down in a letter of assignment signed by the Prime Minister.

Are the evaluations focused enough on specific information needs? We found that 40% of the evaluations included evaluation questions, but this rate increased from 30% in 2013 to 90% in 2015, thanks especially to a stronger methodological involvement of the SGMAP team in the inception phase of the evaluations. As it was expected, the evaluations answered a wider than before range of evaluation questions, including relevance and external coherence in 42% of the cases each. However, the actual contribution of the questioning to focusing evaluation work should be balanced: in many cases we found several dozens of questions in the Terms of Reference (ToR), making it impossible to answer each of them with the proper depth of analysis. Besides, within the MAP framework commissioners had to launch evaluation, but this type of service may not answer their needs: we found that three out of four evaluations included non-evaluative questions, either prospective, or related to cost, implementation or transformational issues.

When it comes to the expectations of other stakeholders, we evidenced that a steering committee was convened in 88% of the evaluations, and that it was open to at least one representative of the beneficiaries of the evaluated policy in 62% of the total number of cases. However, we found that once the commissioner has exchanged views over the ToR with the evaluator, the evaluation demand is merely presented to the associated stakeholders, not discussed. In reality, they neither have had their say over the demand, nor over the actual content or conclusions and recommendations of the evaluation.

As for the way the evaluations were designed, we found that the structuring part was extremely weak. We found almost no instances of a description of the criteria that the evaluators would use to guide the evaluation process and assess the evaluated intervention in the end; we found no cases where the intervention logic of the evaluated policy was reconstructed; actually, the stakeholders targeted by the evaluated intervention were explicitly identified in 55% of the evaluations' ToR or inception note; the expected consequences on them, in only 38%. As for the questioning, this tended to improve over time though.

Evaluators

Regarding the evaluators, the metaevaluation appears quite weak because of the approach followed, which only depended on documentary analysis. The MAP evaluation could rely on an impressive workforce. Some evaluation teams included more than 10 members of the inspection bodies; also, they could count on external advisory or data collection services to help them.

When it comes to the skills of the team and their role in the evaluation process, the findings are more balanced. Members of the Inspection bodies are used to work collaboratively, and despite a few counter-examples, the evaluation team could usually deliver on its assignment. They are also accustomed to work on a variety of topics and, in some cases, they relied on thematic expertise to provide a fresh look on the evaluated policy. However, it is only by exception that they called for evaluative expertise. This was an issue given that most evaluators had no prior expertise of evaluation, and this may have led to teams conducting the evaluation work just as they would have done a classical inspection assignment.

Besides, there is evidence of a learning curve among evaluators and commissioners alike, which can be seen in the increasing quality of the ToR over time.

Evaluation outputs

Evaluations usually arrived on time, especially when they were bound to contribute to a specific decision that had been announced in advance: the evaluator would adapt the ambition of the evaluations to the date, or provide specific content when asked so for a certain need (e.g. to support or refute an amendment in a legislative process).

Evaluators often proved committed to consulting stakeholders beyond the implementation sphere. They would perform several dozen interviews, and up to 200 in some cases. This was in their view an important aspect of the credibility of their work. However, the sheer number of interviews and the lack of strategies to, for instance, use different data collection tools for exploratory and confirmatory phases meant that the synthesis of information was in many cases extremely strenuous. We relate that to a lack of evaluation skills, which was also evidenced by difficulties in taking proper advantage of surveys to collect information.

In the end, we found that the collected data was not always traceable. Conclusions and recommendations are usually based on an analysis, but in 60% of the cases, this analysis does not explicitly present how the different sources of information were used to reach the conclusion. Also, in half the evaluation reports, it was difficult in the analysis to identify in some instances what was a personal point of view of the evaluator, and what was a factual piece of information.

However, when it comes to the credibility of the outputs, it should be noted that the *grands corps* have a credibility of their own that concerns their productions too, including evaluations. Despite spotted issues, it is extremely rare that the evaluation stakeholders call the quality of the evaluation in question, even when they would not agree with the results.

Overview

The metaevaluation shows that the evaluation process was conceived to take into consideration the commissioners' needs and embed the evaluation in a decision process. Evaluators mostly did a serious job, on time and being attentive to the commissioners' requests. However, their lack of evaluation skills may have largely hampered the ability of these services to feed the decision-making *in an "evaluative" way*, e.g. by questioning the rationale of its policy.

In the next section, we will then see whether these evaluations did contribute to the decision-making process.

Contribution to decision-making and public action

The criteria above give a view of the attributes of the evaluation activities (process and report), but what about the actual contribution of these evaluations to decision-making? In this section, we identify the changes that followed the 65 studied evaluations and then focus on 8 evaluations and assess the probability that they have had an impact on the evolution of the evaluated policy.

It should be noted that, given the timeframe of the evaluation, we focus mainly on short-term influence. However, we had the opportunity to identify some more complicated pathways towards influence.

Method used

In line with the contribution analysis approach, our aim at this step was to point out first whether the expected changes in terms of policy change could be spotted, and then in a second time examine the evidence of an actual contribution of the evaluation to these changes.

To verify whether changes had been identified within the domain of each evaluation, we had recourse to the policy monitoring done by the SGMAP following each evaluation. We were confident that this monitoring was comprehensive, as the SGMAP was put under political pressures to show results. In spite of that, the monitoring was limited to public decisions (or to what the commissioners would actually divulge of the internal decision-making processes at stake); also, the decision-making processes in the administration usually abide to their own calendar (or the Parliamentary calendar – or the opinion calendar, see below) and some changes following the most recent evaluations may take some time before being observed.

On the other hand, one of the objectives of this study was to learn about the process and mechanisms of contribution. Consequently, we chose 8 evaluations for in-depth appraisal that were considered by the SGMAP as having had an impact on decision and representative of the diversity of evaluation themes. The evaluations related to: the education policy in disadvantaged neighbourhoods (henceforth "*education*"); support to training and capacity building for farmers (*agriculture*); fight against the use of drugs (*drugs*); road safety policy (*road*); support to the development of same-day surgical care (*surgery*); organisation of political elections (*elections*); local management of household waste (*waste*); pooling of resources at intercommunal level (*pooling*).

Some of the evaluated policies are managed in partnership by one or several ministries and external stakeholders, e.g. professional Unions or local authorities: this is especially the case for Agriculture, Education and Waste. Besides, in almost all cases, the Ministry may decide orientations, but the success of the policy ultimately depends on the uptake of these orientations by the stakeholders targeted by these orientations (for instance, the Ministry of Health may decide to disseminate good practices in terms of same-day surgical care, but patients will only see the benefits of such practices if hospitals decide to follow them). Evaluation use in decision-making should therefore be considered not only at the commissioner level.

This is why, for each case study, we interviewed 5 to 13 stakeholders of the evaluation process, in the evaluation team; on the commissioner side; and among the external stakeholders associated to the process (especially members of the Steering Committee); we studied the corpus of documents associated with the evaluation, especially the documents that were elaborated by the administration as a consequence of the evaluation process. We could also rely on the metaevaluation output related to this specific evaluation process.

To assess how likely it was that the evaluation actually contributed to these observed changes, we built on Process Tracing (Beach & Pedersen, 2013) as well as on our own efforts to create operational tests supporting causal inference. In Process Tracing (PT), the evaluator derives from the theory of change facts that should be observed if the theory is true. The PT framework provides 4 types of empirical tests; passing them may be necessary or sufficient for affirming causal inference (Collier, 2011).

Our aim was to develop "real-world" indicators that would match the PT empirical tests in the case of the contribution of evaluations to decision making. In a previous attempt related to the evaluation of the contribution of research to systemic changes in the area of forestry (Delahais & Toulemonde, 2017), we had found retrospectively that our causal arguments could mainly be related to 4 types of evidence: *consistent chronology* (the changes arrive after the intervention); *authoritative source* (when a contribution has already been proved in a robust study, e.g. a peer-reviewed scientific article); *signature* (when the intervention leaves a direct trace in the changes observed) and *convergent triangulated sources*.

In this case, we aimed at developing such tests *ex ante* so as to verify contributions. We identified 10 such tests that we combined in answer to three sub-questions: is the contribution of the evaluation to the observed changes *plausible*? *Possible*? Or *strong*?

Table 1 below presents these tests.

Table 1: Evidence of contribution: empirical tests

Contribution	Name	Definition
Plausible?	1. Chronology	Changes in the realm of the evaluated policy were observed after, or in some cases in the final part of the evaluation.
	2. Identity	The observed changes at least partially match the evaluation recommendations.
Probable?	3. Action plan	The commissioner elaborates an action plan following the evaluation where s/he takes position over the recommendations and announces that changes reflecting the evaluation will be engaged.
	4. Mention	The legal or administrative base of the observed changes (e.g. a new law, administration instructions) mentions the evaluation as an inspiration for the changes.
	5. Report	Reports laying the ground for the observed change (e.g. a parliamentary report) mention the evaluation as a source of information or an inspiration for the changes.
	6. Consensus	Interviewed stakeholders commonly agree that the evaluation contributed to changes either at the commissioner level or among stakeholders targeted by the evaluation policy, and can give concrete examples of such contribution.
Strong?	7. Innovation	The observed changes are innovative and follow innovative recommendations of the reports. "Innovative" means that we could not find an example where the decision maker was already supporting such a change or a mention in a previous report handed to the commissioner.
	8. Consultation	The commissioner engages in a consultation process over the results or recommendations of the evaluation with the targeted stakeholders (either as a follow-up of the evaluation or sometimes within the process).
	9. Guidance	The commissioner (sometimes the evaluation team) elaborated guidance documents aimed at the targeted public of the evaluated policy, in a view of dissemination.
	10. Other	The evaluation recommendations served as a basis for consultation with stakeholders; members of the evaluation team were asked to contribute to a Parliamentary report; the evaluation report creates momentum in the administration; a member of the evaluation team becomes Head of the evaluated policy.

The answer to the plausibility question (tests 1 and 2) is akin to a "hoop test" in the PT framework: succeeding in the test does not mean that there is a contribution, but not passing it makes contribution not impossible, but unlikely. Also, these tests are quite straightforward and do not necessitate massive data collection, therefore they are useful as a first step of the analysis.

The best way to answer the probability question is through a "smoking-gun test", i.e. obtaining sufficient evidence of a contribution. Tests 3 and 4 would do so by highlighting the "signature" of the evaluation in the observed changed. Still, an evaluation may contribute to a decision-making process without an actual action plan being drafted, or even without being explicitly mentioned. This is why it was possible to fall back to two weaker, "straw-in-the-wind tests" (tests 5 and 6): None of these pieces of evidence is enough to confirm or refute the theory, but taken together in a triangulation process, they increase the confidence in an actual contribution (Mathison, 2016). The interviews are also used to identify changes at the level of targeted stakeholders.

Finally, the strength of a contribution can be characterised in many ways: in this instance, we chose to consider the number or magnitude of the changes that can be related to the evaluation; their innovativeness; and the ability to touch the stakeholders targeted by the evaluated policy. Here, no test is decisive by itself, but several straws in the wind may in the end converge towards a weaker or stronger contribution.

Results

In this section, we present the results of our analysis, starting with the change monitoring performed by the SGMAP. At the end of 2015, changes posterior to the evaluation process could be observed in the realm of the evaluated policy for 38 out of 65 evaluations (58%). Older evaluations displayed a higher rate (e.g. 66% for evaluations finalised one year before). Many different types of changes were observed, including legislative or regulatory processes; evolution of the content of the spending programmes managed by the appointing Ministries or other stakeholders (State agencies, local authorities, hospitals, schools…); and other evolutions, related to the management, monitoring or governance of the commissioning Ministries' interventions.

As said previously, the mere existence of changes in the field of the evaluated policy does not mean that they are due to the evaluation process. An in-depth analysis showed that in 9 evaluations (14%), the main change in the evaluated policy did not match any recommendation or conclusion of the evaluation report.

Therefore, we concluded that a contribution of the evaluation to the changes in the evaluated policy was possible in 29 out of 65 evaluation processes (44%), notwithstanding not-yet-happened changes or unknown internal changes.

To identify whether the contribution was plausible and strong, a process which needs a more comprehensive data collection, we turn to the 8 case studies. Table 2 displays the aggregated results of the empirical tests of contribution. Please note that the case studies were all chosen among the evaluations for which a contribution was at least plausible.

Table 2: Assessed contribution of 8 evaluations

	Probable?				Strong?			
	Action p.	Mention	Report	Consensus	Innovation	Guidance	Concertation	Others
Education	X	X	X	X	X	X	X	X
Agriculture								X
Drugs	X				X			
Road		X		X	X			
Surgery		X		X		X	X	
Election			X	X	X			X
Waste	X			X	X			
Pooling			X	X		X		

First, out of 8 evaluations for which a contribution was plausible, we ended up identifying 1 for which a contribution was unlikely. The Agriculture evaluation report was among those reaching a higher standard of quality, but a parallel exercise of reflection within the Ministry, better connected with the decision-making process, led to similar conclusions. Besides, in a context of tense atmosphere between Farmers Unions and the Ministry, agriculture stakeholders would not agree to engage in a collective evaluation process. It was evidenced though that the evaluation report was symbolically used to support the parallel process of decision. Also, the publication of the evaluation report helped create momentum.

The Drug evaluation is also equivocal on that regard: An Action plan was developed, following the evaluation report on some key points and followed by the actual implementation of some measures. Interviewed stakeholders on the commissioner side claim that these measures had already been identified before and that they merely used the evaluation in a symbolic way. However, the animosity between the evaluation team and the commissioner, rooted in previous experiences and which soared in the evaluation process, makes it almost impossible to obtain any consensus on what happened.

Contribution is confirmed, however, in the 6 other evaluations, by order of magnitude in Education (major parts of the recommendations have been implemented), then Surgery and Waste (key aspects of the recommendations), and finally Road, Pooling and Elections (specific contributions). Also, in the Education, Surgery and Pooling cases, there is evidence that the targeted stakeholders (e.g. secondary education establishments, hospitals, local authorities) may have decided to change their own interventions as a consequence of the evaluation process.

Overview

We have evidenced in that section how a set of empirical tests could be used to increase the confidence in the contribution of evaluation to decision processes. We also found that even when there is no evidence of an instrumental use, there is at least a symbolic use.

We present in the next section the processes through which these contributions were observed and mechanisms explaining these contributions.

What makes the evaluations useful and why?

Uncovering the delivery process

We relied on Contribution analysis to explain how the evaluations have contributed to the decision process. Contribution analysis can be presented as a "pragmatic approach to applying the principles of theory-based evaluation", which is particularly adapted to complex settings such as decision-making processes (Delahais & Toulemonde, 2012; 2017; Mayne, 2000; 2012).

Contribution analysis relies on a theory of change (ToC) to guide the evaluation process. A ToC is a causal model displaying the links between one or several interventions (here, the evaluation process and report) and their expected consequences (Weiss, 1997), which can then be used to guide the evaluation. In our case, this theory was reconstructed at the beginning of the evaluation with the SGMAP evaluation team.

The ToC displays two main causal chains: the first one is about influencing the decision process at the State level (taking into consideration the administrative and legislative processes); the second is about influencing the decision processes at the level of stakeholders targeted by the evaluated intervention.

Figure 1 on the next page summarises this ToC by displaying its main causal assumptions.

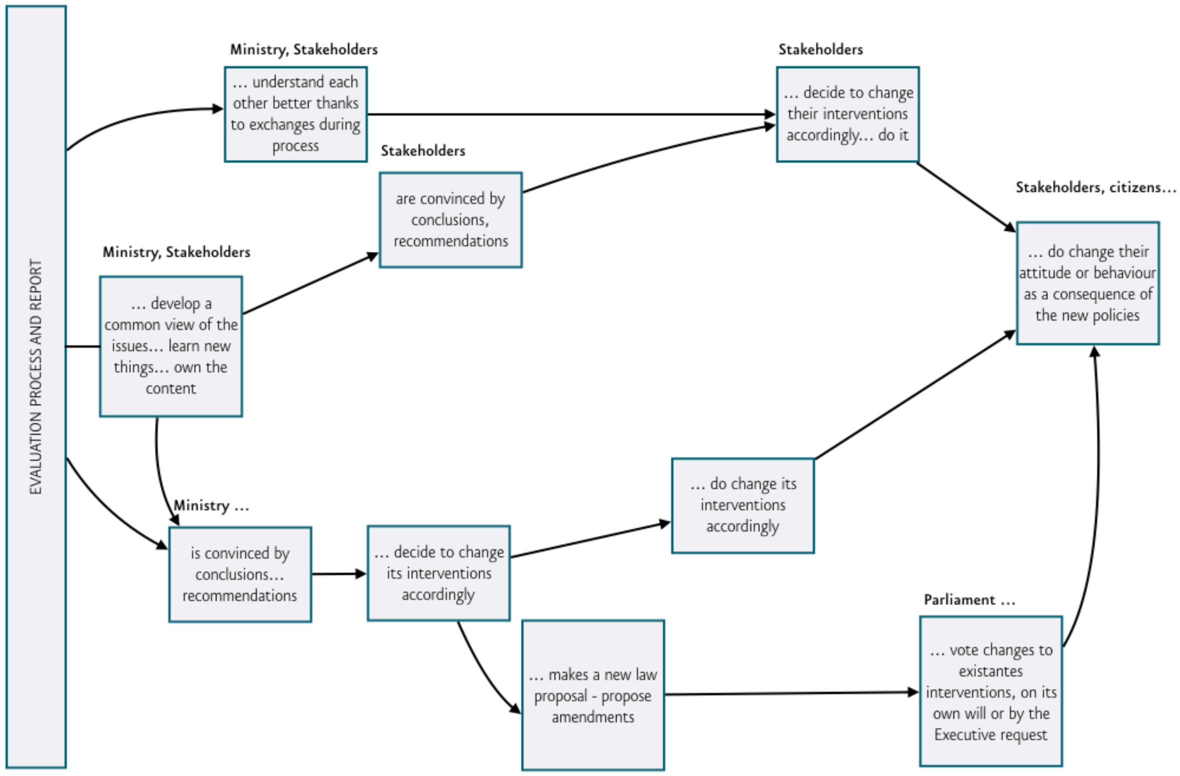
Following the Contribution Analysis approach, evidence was systematically collected at each step of the Theory of Change, aiming at confirming or refuting each potential explanations of the observed changes (including, but not limited to the evaluation).

Our study found that the delivery process of the contribution was, in the 6 cases where a contribution had been identified, broadly following the expected pattern when it comes to the instrumental use of the evaluation within the administration, but with many alternative processes or alternative explanations stepping in.

Usually contribution to decision would take place in the immediate aftermath of the evaluation, but in the Waste case, the commissioner asked the evaluator for findings that would support some upcoming decision before the end of the evaluation. On the Road case, the evaluation was left aside until road casualties began to soar again a few months later. The ministry wanted to show they were committed to fighting this and found some new measures to implement in the report.

The unexpected yet clearest link between the evaluation and decision was in the Education case, where the evaluation team leader became the Head of the "Éducation prioritaire" Department, effectively being in charge of implementing the recommendations of the report. Another example of the evaluation team being later involved in the decision process was the Election case, in which the authors were involved in further reports on the same topic.

Figure 1 : Diagram summarising the theory of change of the MAP at individual-evaluation level



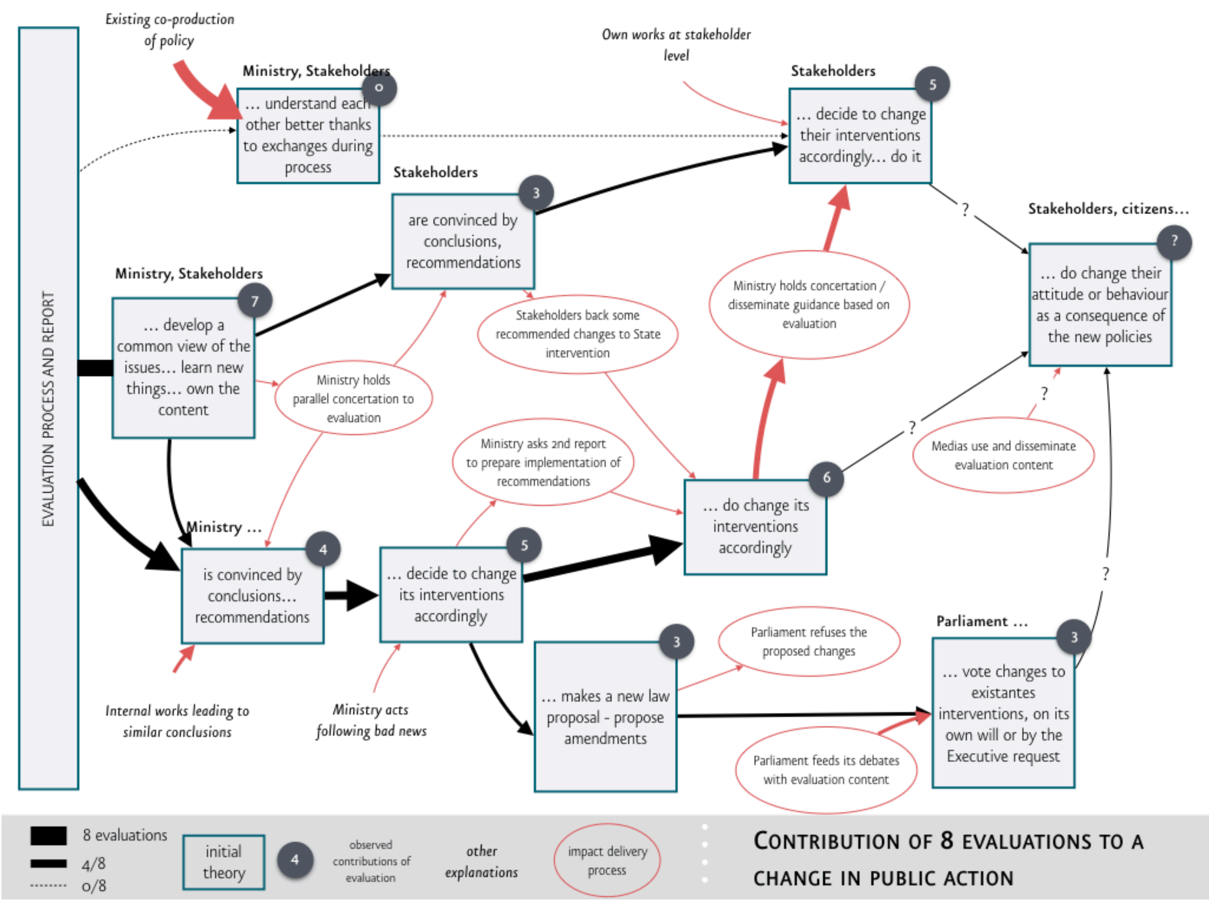
The Parliament introduces another element of unexpectedness. In the Election case, the National Assembly refused the changes proposed by the Government. The national association of communal authorities, the AMF, supported amendments reflecting a recommendation of the Pooling report against the advice of the administration. A negotiation led to a halfway solution.

The opposite was true when it comes to the stakeholders targeted by the evaluated policy. We found that the process use of the evaluation in terms of better common understanding was nowhere to be seen in our 8 case studies, and that stakeholders were convinced by conclusions and recommendations in 3 out of 8 cases. This was a surprise as previous meta-evaluations in France had concluded to a lack of instrumental use, but a frequent process use leading to better co-operation between the evaluating authorities and other stakeholders (Epstein, 2009; Toulemonde, Genard, Jacob, & Varone, 2006).

A first reason though is that on many evaluated topics, a formal, sometimes institutionalised dialogue already existed between the State and concerned stakeholders, e.g. on Education, Waste, Agriculture or Road. The value added of an evaluation process in which external stakeholders had a lot less possibilities to influence decision was not evident to them. Besides, we found that commissioners were (in these admittedly "best-case scenarios") very aware that targeted stakeholders needed to be convinced by their new political orientations if they were to be a success. However, they would bypass the MAP evaluation process to do so.

Figure 2 below summarises the observed process deliveries.

Figure 2 : Diagram summarising the identified patwhays to the expected changes of the MAP



For instance, in the Education case, they organised a parallel concertation to ensure that the unions of teachers and directors would agree with the proposals. In the Pooling, the evaluators were involved in elaborating a guide for local authorities summarising the evaluation lessons; in the Surgery case, tools to support the development of same-day surgery.

Explaining the delivery processes

Explaining how the evaluations have contributed to the decision process can be done in two ways: first, by uncovering the processes through which these contributions have been delivered, as above; and then by identifying the mechanisms explaining these contributions.

Building on original conceptualisations described above, more recent research have proposed a shift in terminology from *evaluation use* to *evaluation influence* (Kirkhart, 2000). Kirkhart defines influence as “the capacity or power of persons or things to produce effects on others by intangible or indirect means” (Kirkhart, 2000, p. 7).The main argument behind this shift is that the term influence is less mechanical than the term use and better refers to the unintentional, plethoric and long-term consequences and *unaware/unintended* impacts an evaluation could or could not have (Alkin & Taut, 2002).

Mark and Henry pushed this theory of influence further in 2003 and developed a "comprehensive theory of evaluation influence" that offers a "more specific framework and typology of influence". In their model of alternative mechanisms that may mediate evaluation influence, they present a complete set of influence mechanisms combining four types of process or outcome (general influence, cognitive and affective, motivational, and behavioural types) and three levels of influence (individual, interpersonal, and collective levels). One stimulus to the model's development is the idea that the traditional end-state uses did no trace the sometime long and complex path through which evaluation comes to have its influence.

Since then, few studies attempted to establish empirical basis and evidence for the practice of evaluation influence[3]. Compared with *evaluation use*, empirical studies of *evaluation influence* are still limited and relatively little is known about how evaluation influence may impact on decision makers' attitudes and actions.

In our case, due to the short time since the delivery of the evaluation, it was unlikely that we could identify the long chains of mechanisms and domino effect described by Mark & Henry. The chains are shorter and mainly deal with the general influence and behavioural dimensions rather than on the cognitive dimensions. However, as mentioned above, the pathways to deliveries were more diverse than initially expected.

Below we identify some of the most prominent mechanisms we have observed, at macro- and microsociological levels.

Mechanisms at macrosociological level

Although outcomes of institutionalization mechanisms are not always easy to gauge, "authors argue that institutionalization contributes to results utilization and quality improvements" (Jacob and al., 2015, p.18). Besides the number of evaluations done, we think two contributions are particularly interesting here:

- Prime minister impetus as an evaluation operational clause

This new clause in the MAP evaluation process had a strong effect in so far as evaluation and decision are embedded in a clear will to build transformation or reform decision on the evaluation findings. Evaluation is part of the decision-making process and we observed that ministries, in almost every case, waited for the evaluation results to make a final decision.

[3] These studies include those by (Weiss, 2005), Christie (2007), and (Gildemyn, 2014). The first two studies reported that all three types of evaluation information (including large-scale evaluation study data, case study evaluation data, and anecdotes) "influence decision makers' decisions" (Christie, 2007: 22), and "evaluation evidence travelled to influence decisions about D.A.R.E" (Weiss, 2005, p. 27).

This may also have lead to symbolic use, where ministries would only pretend using evaluation to feed their decisions to formally abide to the Prime Minister's will. Also, ministries who had an important evaluation activity, such as the Ministry of Overseas, would keep away as much as possible from the MAP process, to keep its liberty to use it as it wished to.

- Structuring an Evaluation community within the State

Inspections bodies first played a quantitative role: more than 300 members of the Inspection bodies were involved in MAP evaluations. By way of comparison, the French society of Evaluation (SFE) counts a similar number of evaluation professionals for entire France.

Members of the Inspection bodies are used to be part of the decision processes and they could use their own credibility to create a climate of trust between the commissioners and them. They also had a direct access to the top officials of the administration and even to ministries or their head of office.

Mechanisms at the microsociological level

At the micro level, it was possible to recognise many of the traditional findings related to decision-making.

A MAP key attribute towards utilisation is the attention towards potential use, by defining precisely the subject of the evaluation, and making sure that the results would be available at the right time for decision-making.

We also could observe how decision-makers would make their choice among recommendations, based on their perceived relevance (does it fit their own vision? Is the argument sound enough?); whether the changes were deemed acceptable to the targeted stakeholders; technical feasibility and opportunity (can a current legislative process or policy design process be used to implement the recommendations?). In the Waste case, the "political window" for some changes in that area was passed; but the commissioner used the opportunity of two upcoming laws to use the evaluation in the process of drafting legislation.

Another important aspect was that in many cases, decision was not only about commissioners "deciding" to follow the recommendations. They also had to convince others: their own administration, other involved ministries, and the Budget ministry if the changes had financial consequences. They would therefore use evaluation as a justification tool in the decision-making process, within or outside the space of negotiation provided by the MAP process. In an instance, we found that external stakeholders could also use the evaluation to weigh on the decision (Pooling).

It does not seem however that the evaluation would persuade the politicians themselves. For instance, the Education evaluation report did not change the personal aspiration and commitment of the Ministry of Education, which was personally involved in the evaluation discussion and methodological debates and its follow up, but contributed to (a) solidify interpersonal judgement about program effectiveness as well as (b) policy-oriented learning, which in turn (c) fostered a collaborative change in elaboration practices throughout the redaction of an action plan (d) widely reported on the media, that finally led to (e) a major policy reform (policy change).

Evaluation could also be used to transform the changes of external stakeholders but rather through dissemination of the results than through the actual evaluation process. For instance, whereas few studies had dealt until then with the concrete implementation of same-day surgery in France, the evaluation work provided (a) a technical justification for ambitious development objectives at State level. The ministry used this justification to (b) emit a national instruction supporting the development of same-day surgery. One of the output of the evaluation was the provision to local health actors of (c) much needed tools to define their own development objectives accordingly.

Overview

We have presented here how the processes of decision making reflects some of the mechanisms and explanations provided in the literature. The institutionalisation processes are more unique though and may have some unexpected consequences on the future development of evaluation.

Conclusion

In this article, we have achieved two objectives related to the international agenda for evaluation research : providing an empirical account of a large-scale endeavour to use evaluation in decision-making; and contributing to the debate on how to analyse the contribution of evaluation to decision-making.

Was our own metaevaluation useful? Time only can tell. In a way, the MAP has taken the exact opposite approach to previous attempts of evaluation institutionalisation. In the 1990s, ambitious, up-to-the-standards evaluations had not been used in decision-making. With the MAP, "below-standards evaluations" have been used in a variety of ways, but not always in "evaluative ways". We found that in this case, credibility brought by the unique position of the evaluators in the State, as well as by the institutionalised evaluation of the MAP were enough to "replace" methodological credibility. The question is therefore whether the members of inspection bodies will want to keep this role in evaluation and whether they can get better at doing them, so as to continue the path towards institutionalisation of evaluation in France. This was at the core of this study's findings and recommendations.

The fact though is that institutionalisation is a stop-and-go process, with moves forward and backward effects. It is not known yet whether the new government will actually launch a new cycle of national evaluations. Audit-like processes akin to the RGPP may be more seductive in a context of fast reforming strategies. Evaluation may also be focused on the current or expected experimentations that the government wants to launch.

Bibliography

Alkin, M. C., & Taut, S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, *29*(1), 1–12. http://doi.org/10.1016/S0191-491X(03)90001-0

Appleton, S. (2012). *Understanding the Mechanisms and Outcomes of Evaluation Influence Within Population Health Partnerships*. ResearchSpace@ Auckland.

Barbier, J.-C. (2010). Elements for a sociology of public policy evaluation in FranceAbstract. *Revue Française Des Affaires Sociales*, (1), 25–49.

Barbier, J.-C., & Hawkins, P. (2012). Evaluation Cultures: Sense-making in Complex Times. Transaction Publishers.

Baron, G., & Monnier, É. (2003). Une approche pluraliste et participative. *Informations Sociales*.

Beach, D., & Pedersen, R. B. (2013). Process-Tracing Methods. University of Michigan Press.

Bemelmans-Videc, M. L., Eriksen, B., & Goldenberg, E. N. (1994). Facilitating organizational learning : human resource management and program evaluation. In F. L. Leeuw, R. C. Rist, & R. C. Sonnichsen (Eds.), *Can Governments Learn* (pp. 145–188). New Brunswick, NJ.

Bezes, P. (2009). Réinventer l'État. Presses Universitaires de France. http://doi.org/10.3917/puf.bezes.2009.01

Boissier, P. (2015). L'évolution des compétences et des métiers des corps d'inspection : l'exemple de l'inspection générale des affaires sociales. *Revue Française D'administration Publique*, *N° 155*(3), 679–684. http://doi.org/10.3917/rfap.155.0679

Bondaz, M. (2012). Bilan de la RGPP et conditions de reussite d"une nouvelle politique de reforme de l"État. Paris.

Cohen, M. D., March, J. G., & Olsen, J. P. (1991). Le modèle du «garbage can» dans les anarchies organisées. In *Décisions et organisations*. Paris: Décisions et organisations.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science \& Politics*, *44*(04), 823–830. http://doi.org/10.1017/S1049096511001429

Coosky, L., & Caracelli, V. (2009). Metaevaluation in practice. Journal of multidisciplinary Evaluation.

Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of practice. *Evaluation*, *18*(3), 281–293. http://doi.org/10.1177/1356389012450810

Delahais, T., & Toulemonde, J. (2017). Has Research contributed to Sustainable Forest Management? Making causal claims without counterfactuals. *Evaluation*.

Dente, B., & Kjellberg, F. (1988). The dynamics of institutional change: Local government reorganization in Western democracies.

Destais, N., Marigeaud, M., Battesti, J. P., & Bondaz, M. (2012). *Cadrage méthodologique de l'évaluation des politiques publiques partenariales*. La Documentation française. Retrieved from http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/124000683.pdf

Devaux-Spatarakis, A. (2014). *La méthode expérimentale par assignation aléatoire: un instrument de recomposition de l'interaction entre sciences sociales et action publique en France?* Centre Emile Durkheim–science politique et sociologie comparatives, UMR5116.

Duran, P. (2010). L'évaluation des politiques publiques : une résistible obligation. *Revue Française Des Affaires Sociales*, (1), 5–24.

Epstein, R. (2009). À quoi sert l'évaluation? Les leçons de la politique de la ville. *Tracés*, (3), 187–197.

Euréval, European Commission, Directorate-General for the Information Society, Media. (2011). Evaluating legislation and non-spending interventions in the area of information society and media. Luxembourg: EUR-OP. http://doi.org/10.2759/47388

Furubo, J.-E., Rist, R. C., & Sandahl, R. (2002). International Atlas of Evaluation. Transaction publishers.

Gélédan, F. (2012). Le management au secours de services publics ? Modernisation de l'État et régimes de domination à l'heure de la Révision générale des politiques publiques (RGPP) – 2007-2012. *Siècles*, (35-36).

Gildemyn, M. (2014). Understanding the Influence of Independent Civil Society Monitoring and Evaluation at the District Level. *American Journal of Evaluation*, *35*(4), 507–524. http://doi.org/10.1177/1098214014525257

Guba, E. G., & Lincoln, Y. S. (1989). Fourth Generation Evaluation. SAGE Publications Ltd.

Hall, P. A., & Taylor, R. C. R. (1996). Political Science and the Three New Institutionalisms. *Political Studies*, *44*(5), 936–957. http://doi.org/10.1111/j.1467-9248.1996.tb00343.x

Halpern, C. (2004). Entrée "Décision." In L. Boussaguet, S. Jacquot, & P. Ravinet (Eds.), *Dictionnaire des politiques publiques*.

Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*(1), 35–57.

Jacob, S. (2005). La volonté des acteurs et le poids des structures dans l'institutionnalisation de l'évaluation des politiques publiques. *Revue Française De Science Politique*, *Vol. 55*(5), 835–864.

Jacob, S., & Varone, F. (2004). Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970-2003). *Politiques Et Management Public*, *22*(2), 135–152. http://doi.org/10.3406/pomap.2004.2844

Jacob, S., & Varone, F. (2007). L'évaluation des politiques au niveau régional.

Jacob, S., Speer, S., & Furubo, J.-E. (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, *21*(1), 6–31. http://doi.org/10.1177/1356389014564248

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on Evaluation Use. *American Journal of Evaluation*. http://doi.org/10.1177/1098214009341660

Kessler, M.-C. (1986). Les grands corps de l'État. *La Revue Administrative*, *39*(231), 221–228. http://doi.org/10.2307/40779956

King, J. A., & Pechman, E. M. (1984). Pinning a Wave to the Shore: Conceptualizing Evaluation Use in School Systems. *Educational Evaluation and Policy Analysis*, *6*(3), 241. http://doi.org/10.2307/1163870

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, *2000*(88), 5–23. http://doi.org/10.1002/ev.1188

KPMG, Quadrant Conseil. (2017). *Évaluation de la démarche globale d'évaluation des politiques publiques menée dans le cadre de la modernisation de l'action publique*. Retrieved from http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-evaluant-ses-politiques-publiques/levaluation-des-politiques-publiques-evaluee-a-son-tour

Lacouette-Fougère, C., & Lascoumes, P. (2013). Les scènes multiples de l'évaluation-les problèmes récurrents de son institutionnalisation. *SciencesPo, LIEPP*.

Le Clainche, M. (2013). La Modernisation de l'action publique (MAP) n'est pas la poursuite de la RGPP. *Revue Française D'administration Publique*, *n° 145*(1), 203–208. http://doi.org/10.3917/rfap.145.0203

Leca, J. (1993). Sur le rôle de la connaissance dans la modernisation de l"État et le statut de

l"évaluation. Revue française d'administration publique.
Leviton, L. C., & Hughes, E. F. X. (1981). Research On the Utilization of Evaluations. *Evaluation Review*, *5*(4), 525–548. http://doi.org/10.1177/0193841X8100500405
Lindblom, C. E. (1959). The Science of "Muddling Through." *Public Administration Review*, *19*(2).
Lindblom, C. E. (1965). The Intelligence of Democracy. New York: The Free Press.
Lloyd, R., & Schatz, F. (2015). Improving Quality: Current Evidence on What Affects the Quality of Commissioned Evaluations. Centre for Development Impact.
Mathison, S. (2016). Why Triangulate? *Educational Researcher*, *17*(2), 13–17. http://doi.org/10.3102/0013189X017002013
Matyjasik, N. (2010). *L'évaluation des politiques publiques dans une France décentralisée. Institutions, marché et professionnels*. Université de Bordeaux; Université Montesquieu-Bordeaux IV; Institut d'études politiques de Bordeaux; SPIRIT.
Mayne, J. (2000). Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly. *ResearchGate*, *16*.
Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, *18*(3), 270–280.
Mayntz, R., & Scharpf, F. W. (1995). Der Ansatz des akteurzentrierten Institutionalismus. In *Schriften aus dem Max-Planck-Institut für Gesellschaftsforschung* (pp. 39–72). Campus Verlag.
Monnier, E., Conan, M., Allen, B., Duran, P., Spenlehauer, V., & Toulemonde, J. (1992). Evaluations de l'action des pouvoirs publics. *Collection CPE-Economica*.
Monnier, É., & Fontaine, C. (2002). Evaluation in France. In J.-E. Furubo, R. C. Rist, & R. Sandahl (Eds.), *International Atlas of Evaluation*.
Nioche, J.-P. (1982). De l"évaluation à l"analyse des politiques publiques. *Revue Française De Science Politique*, *32*(1), 32–61. http://doi.org/10.3406/rfsp.1982.393996
Nutley, S., & Webb, J. (2000). Evidence and the policy process. *What Works*.
Patton, M. (2008). Utilization-focused evaluation. London: Sage.
Pawson, R., & Tilley, N. (1997). Realistic evaluation. Sage.
Picavet, O. (2015). Les caractéristiques des évaluations réalisées par les corps d'inspection dans le cadre de la Modernisation de l'action publique. *Revue Française D'administration Publique*, *N° 155*(3), 713–722. http://doi.org/10.3917/rfap.155.0713
Pollitt, C. (1995). Justification by Works or by Faith? *Evaluation*, *1*(2), 133–154. http://doi.org/10.1177/135638909500100202
Shulha, L. M., Caruthers, F. A., & Hopson, R. K. (2010). The Program Evaluation Standards. (SAGE, Ed.) (3rd ed.).
Silvent, F., Dureau, G., & Guillaume, H. (2002). Gestion publique. L'État et la performance. Presses de Sciences Po et Editions Dalloz.
Simon, H. A. (1945). Theory of Games and Economic Behavior.John Von Neumann , Oskar Morgenstern. *American Journal of Sociology*, *50*(6), 558–560. http://doi.org/10.1086/219699
Stockmann, R., Meyer, W., & Nolte, C. (2017). Evaluation globe – compendium on the world-wide institutionalization of evaluation.
Thoenig, J.-C. (2000). L'évaluation, source de connaissances applicables aux reformes de la gestion publique. *Revue Française d'Administration Publique*, (92), 681–693.
Toulemonde, J., Genard, J.-L., Jacob, S. B., & Varone, F. (2006). Appropriation des résultats de l'évaluation: leçons de la pratique en Région Limousin. L'évaluation au niveau regional.
Weiss, C. H. (1983). Ideology, Interests, and Information. In *Ethics, The Social Sciences, and Policy Analysis* (pp. 213–245). Boston, MA: Springer US. http://doi.org/10.1007/978-1-

4684-7015-4_9
Weiss, C. H. (1993). Where politics and evaluation research meet. *Evaluation Practice*, *14*(1), 93–106. http://doi.org/10.1016/0886-1633(93)90046-R
Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, *1997*(76), 41–55.
Weiss, C. H. (1999). The Interface between Evaluation and Public Policy. *Evaluation*, *5*(4), 468–486. http://doi.org/10.1177/13563890990050040 8
Weiss, C. H. (2005). Murphy-Graham, Erin og Birkeland, Sarah (2005):"An Alternative Route to Policy Influence. How Evaluations Affect DARE" American Journal of ….