Evaluating Evaluations of Innovation Policy: Exploring Reliability, Methods, and Conflicts of Interest

Elias Collin, Christian Sandström, and Karl Wennberg

Abstract Expansions of innovation policies have been paralleled with an increase in the evaluations of such policies. Yet, there are few systematic evaluations of how such evaluations are conducted, by whom, and their overall conclusions. We analyze 110 evaluations of innovation policy in Sweden from 2005 to 2019. Our findings show that the majority of these evaluations are positive, about one-third are neutral in their conclusions, and very few are negative. The majority of evaluations were conducted by consulting firms, close to one-third by expert government agencies, and around 10% by university researchers or as self-evaluations by the governmental agencies responsible for the policy themselves. Few evaluations employed causal methods to assess the potential effects of policies. We discuss conflicts of interest and question the reliability of evaluations of innovation policy.

Keywords Evaluation · Consultants · Evaluating practice · Meta-Evaluation · Sweden · Innovation policy

E. Collin (✉)
School of Economics and Management, Lund University, Lund, Sweden
e-mail: elias.collin@fek.lu.se

C. Sandström
Jönköping International Business School, Jönköping, Sweden

The Ratio Institute, Stockholm, Sweden
e-mail: christian.sandstrom@ju.se

K. Wennberg
Stockholm School of Economics, Stockholm, Sweden
e-mail: karl.wennberg@hhs.se

1   Introduction

Innovation policies have become increasingly popular over recent decades. The European Union, national governments, and regional agencies provide a collection of support systems and structures. Firms can apply for innovation grants, collaborative projects, training, and other types of public support. Billions of euros are made available for developing certain ex-ante chosen technologies, such as hydrogen gas or Artificial Intelligence applications. Inspired by scholars such as Mariana Mazzucato (2012, 2018), such public sector initiatives have grown in size and scope in recent years.

This expansion of interventionist innovation policies has been followed by an equal growth in the number of evaluations of innovation policies. Little is yet known regarding these evaluations: Who performs these evaluations? What methods are employed in order to make evaluations? What conclusions are generally drawn? How are results, methods, and the kind of evaluator interrelated? Are evaluations reliable?

With this chapter, we add a piece to the puzzle of innovation policy by analyzing a set of policy evaluations. Drawing upon a random sample of 110 innovation policy evaluations in Sweden from 2005 to 2019, we provide descriptive and multivariate statistics to answer the aforementioned questions. Our results show that the majority of evaluations are positive, many are neutral, and very few are negative. We also show that evaluations are often performed by private consulting firms. Based upon our results, we discuss issues concerning evaluators' independence and potential conflicts of interest.

Our study makes three contributions. First, our empirical analysis provides insights of relevance to both the innovation studies and program evaluation literatures by showing that policy evaluations may differ across different types of evaluative actors and across the distinct methods employed in their evaluations. Second, our focus on a whole body or corpus of evaluations in a specific policy domain provides a novel approach to studying evaluations in that previous studies have often offered commendable evaluations of specific policies or reforms, or meta-evaluations—assessing the quality of certain evaluative projects—but a holistic approach to the evaluation area has, to our knowledge, hitherto been lacking within the fields of evaluation research and innovation policy. By examining connections between different types of evaluative actors, their methods, and their conclusions, the study facilitates a more in-depth understanding of how evaluating actors' and their methods are related to results and recommendations from such evaluations. Third, our discussion regarding evaluators' independence and potential conflicts of interest provides insights of broader relevance to academic and policy discussions about the role of evaluations in public policy.

The remainder of this chapter is organized as follows: The next section provides an overview of innovation policy evaluations and literature on evaluations. Next, we present and discuss our empirical data. Latterly, a concluding remark is provided.

2 Background: Evaluation as a Practice

Evaluating public policy is a somewhat difficult operation. Any society will likely succumb to public waste without any evaluative elements making sure that public resources are not wasted or misused (Furubo et al., 2002). Yet, it is easy to imagine how too close and frequent control of public servants or policy quickly becomes absurd. Having a grade school teacher being monitored in detail during daily classes or having every agency's decision double-checked by another auditing agency would not only prove costly but also, most likely, quite futile. Hence, a balance between the two is necessary—societies need both trust and evaluation in order to work.

The term evaluation is often used in a rather general and arbitrary way. In a broader sense, evaluation is distinguished from similar practices like auditing or reviewing through the fact it features judgment. An evaluation is not just a display of numbers or opinions but includes some sort of judgment of the studied practice in relation to a predesignated norm or goal (Scriven, 1991; Pollitt, 2003; Knill & Tosun, 2012).

Based on this definition, a multitude of evaluation practices exist. Among these no specific practice can be distinguished as superior to the others. Different practices rather serve different purposes. As with scientific methodology in general, the choice of evaluation method and practice depends on the value or goals of interest to the evaluator.

The trend toward the large-scale evaluations we see today started in the United States in the 1950s and 1960s. Great hope was then invested in various social and political scientists, who, with the help of quantitative and objective methods, were to scientifically find the best ways to govern society. Subsequent evaluators would question this evaluation practice in favor of what can be described as a more constructivist approach. Greater emphasis was put on experiences from public officials and the people targeted by the studied political intervention. Today, both traditions live on and are present in many Western countries (Dahler-Larsen, 2007; Bovens et al., 2008).

Since the late 1900s, evaluation activities in society have increased exponentially, noted not least by Power (1997) in *The Audit Society*. The huge increase in public scrutiny can be attributed to an expanded public interest in such activities, an increased focus on goal and result management and several of the various governance practices that are referred to under the name *New Public Management* (NPM)—in part replacing the preceding Weberian public servant model predominating in Western democracies throughout the twentieth century.[1]

Other factors driving the trend toward more evaluation are organizations such as the European Union and the World Bank putting external pressure on countries to

---

[1] Named after the German sociologist Max Weber, *Weberian bureaucracy* is seen as a system of public administration in which bureaucrats in hierarchal organizations executed political decisions in accordance with predetermined and exact rules and equality before the law.

further their evaluative commitments, often as a condition for financial support or other benefits (Furubo et al., 2002).

2.1 Different Evaluators

Evaluations are conducted by a variety of different actors ranging from researchers who evaluate with research interests, to consulting firms, think tanks, agencies, ombudsmen, and specially appointed commissions or evaluation agencies; it is also common that executive agencies conduct self-evaluations. The same intervention or political effort can be evaluated multiple times by different actors. For example, the crash of a Dutch military cargo plane at Eindhoven Airport in 1996, and the subsequent crisis management, led to no less than 15 different evaluations from different actors (Goodin et al., 2008). While this event was extreme, it highlights the importance of evaluation in describing reality and providing recommendations for improvements of regulations, processes, and procedures, and also whether those regulations, processes, and procedures are effective in attaining the envisioned goals. The Eindhoven incident also highlights that different evaluators may reach different conclusions, a topic hitherto rarely attended to in the innovation literature. As our study will show, one of the aforementioned actors, the consultants, might be of special interest for those studying innovation policies.

During recent years, there has been a general trend in public administration toward an increased use of consultants. Although in many aspects it has been beneficial and efficient, the trend is also connected to several drawbacks. Scholars have pointed to reduced competence within public agencies, a confusion of responsibility between those contracted for a job and those ordering it, and a shift in values within the public sector: consultants bringing what can be referred to as instrumental rationality, a constant demand for efficiency, and evidence-based practices at the expense of normative judgments within the public sphere (van den Berg et al., 2019; Ylönen & Kuusela, 2019).

The field of evaluation is no exception to this trend. Although developing at different speeds in different countries, large organizations such as the AEA (American Evaluation Association) and the EES (European Evaluation Society) signify almost industry-sized evaluation markets connected to American and E.U. political reforms.

3   Empirical Setting: Innovation Policy in Sweden

In Sweden as in many other Western countries, evaluations are conducted throughout the entire public sector. Innovation policy presents no exception. Here, this policy area is amply funded as state grants only (not counting E.U., regional, and local investments) amount to more than €1 billion annually (Karlson et al., 2019). In

the United States, these figures have been gauged to be above $13 billion (Hunt & Kiefer, 2017).

As stated, evaluation can be conducted in several different ways, none of which is by default superior. However, once one has decided upon an evaluating policy and what to actually evaluate within each specific intervention, certain methods may be more preferred. Our initial premonition, supported, for instance, by an audit made by The Swedish National Audit Office (Swedish NAO) (2020), was that evaluative practices and judgment calls varied somewhat between evaluative actors. Here, the Audit Office states that *"there are considerable weaknesses in the effect evaluations of industrial policy that have been carried out by government agencies: only 2 out of 37 studied evaluations fulfill all three elementary criteria set up by the NAO regarding credible evaluations"* (2020, p. 4).

Apart from the report from the Swedish NAO, other studies provide initial concern. A few rather thorough research reports based on counter-factual methods contradict the otherwise quite favorable picture of the output of policies within the field and point to a lack of effects on firm turnover, number of employees, profits, or productivity (Daunfeldt et al., 2016; Gustavsson Tingvall & Deiaco, 2015).

4 Results

Innovation policy in Sweden is mainly organized through a few big, self-governing, state agencies, as is typical for Swedish public administration. Agencies such as Vinnova (the Swedish Innovation Agency), *Tillväxtverket* (the Swedish Agency for Economic and Regional Growth), and *Energimyndigheten* (the Swedish Energy Agency) are in charge of the lion's share of allocated resources.

Evaluations are also conducted by two independent agencies: *Tillväxtanalys* (The Swedish Agency for Growth Policy Analysis) and the previously mentioned Swedish NAO. Evaluation is also performed by researchers and by consultants, hired to evaluate specific tasks.

The empirical approach of the study involved reading and coding a total of 110 policy interventions from 2005 to 2019 with regard to the judgment calls made in the evaluations, the evaluative actor, the evaluative methods, the type of data used in the evaluation, as well as a few control variables. The results are presented below.

The study shows that evaluations of Swedish growth and innovation policies largely consist of positive reviews. Among the 110 evaluations examined, there are 67 positive, 37 neutral, and 6 negative evaluations. Figure 1 shows the frequencies of different results in the studied evaluations.

The low share of evaluations containing negative policy evaluations in Fig. 1 is noteworthy. One possible explanation based on these results is that Swedish growth and innovation policy overall shows quite remarkably effective and efficient results—rightfully resulting in positive evaluations. An alternative explanation would be that some actors embellish their evaluations and write evaluations that
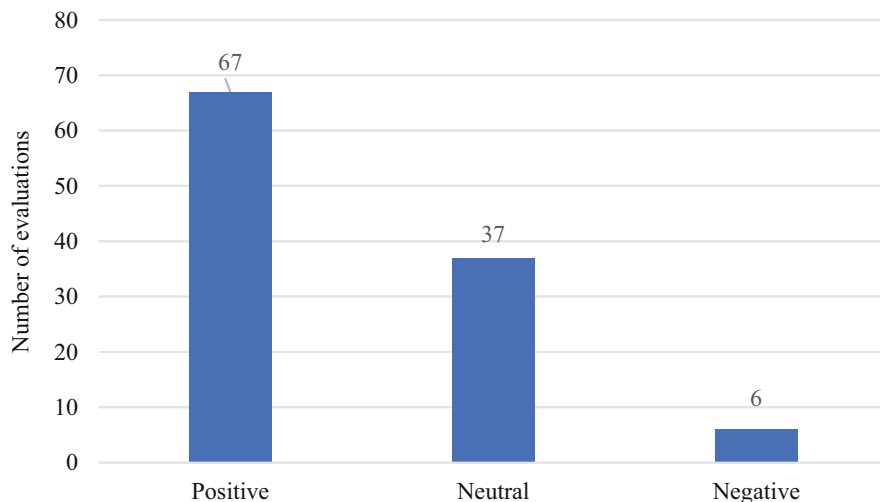
Fig. 1 Reviews given by evaluations of Swedish growth and innovation policies

give the impression that the policy seems to function better than it does. In the next section, we intend to probe various reasons behind the positive evaluations by analyzing the different actors responsible for the evaluations, as well as methods used in their evaluations.

4.1 Evaluators of Innovation Policy

Moving on to the different types of actors responsible for the evaluations of growth and innovation policies in Sweden, we see that most evaluations are carried out by consultants, either by self-employed consultants, larger firms, or several firms in constellation. Overall, slightly more than half (56 of 110) of all evaluations in our dataset are made by consultants. The second most frequent actor is evaluative agencies (31 out of 110 evaluations), followed by researchers or research groups (15 out of 110 evaluations). Public agencies evaluate themselves in the form of self-evaluations but such self-evaluations make up only 8 of the observed evaluations. In a few of the evaluations carried out by consultants and evaluative agencies, researchers have been invited to comment on the results, inform the evaluators about the evaluated field, or to carry out quantitative evaluations. In these cases, however, the researchers are not regarded as the evaluating actor because they only contribute to a small part of the work. Figure 2 shows the frequencies of different actors among the studied evaluations.

The fact that so many evaluations are carried out by consultants aligns with the general public administration trend pointing to a large and increasing use of consultants in public administration (van den Berg et al., 2019; Ylönen & Kuusela, 2019).
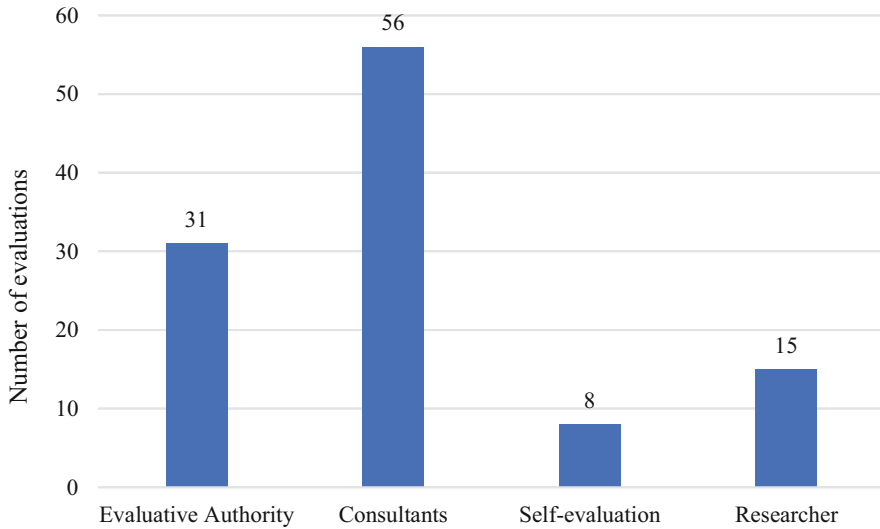
Fig. 2 Evaluative actors conducting evaluation of Swedish growth and innovation policies

4.2 Evaluation Methods and Data Sources

Regarding the methods used by evaluators, a few initial notes should be made: Evaluative practice could of course be studied and classified in different ways. One might, for instance, distinguish between methods focusing solely on goal accomplishment or on goal accomplishment as well as potential side effects. One could also focus on opinions from users or consumers of a certain policy, from the professionals implementing it or a larger society somehow affected by the policy (Vedung, 2009). Yet another way would be to evaluate the efficiency or effectiveness of the policy—focusing on the means spent to achieve a certain result (Vedung, 2009). Within each of these evaluative methods, more distinctions could of course be made.

In the current study, we have coded the methods as either *quantitative descriptive methods*, *qualitative methods*, *quantitative counterfactual (or experimental) methods*, or a *mix* of either the first two or all three of the methods.[2]

The study results show that qualitative methods are used to the greatest extent among the evaluations studied—qualitative methods occur in 61 of the cases. The second most common is that of mixed methods 1 (quantitative descriptive and qualitative methods), which occurs in 31 of the cases.

---

[2]Hence, the five categories of methods in the evaluations are: (1) *Quantitative descriptive methods*; (2) *Qualitative methods*; (3) *Quantitative contrafactual (or experimental) methods*; (4) *A mix of 1 and 2*; and (5) *A mix of 1, 2, and 3*.
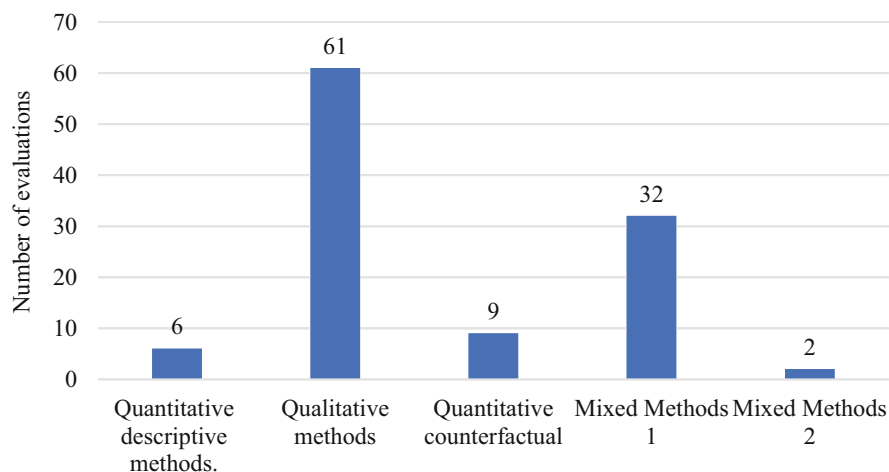
Fig. 3 Evaluative methods used in evaluations of Swedish growth and innovation policies

The quantitative counterfactual method was used in 9 of the cases and the qualitative descriptive method was used in 6. In 3 of the cases, mixed methods 2 (quantitative descriptive, qualitative, and quantitative counterfactual methods) were utilized. Figure 3 shows the frequencies of each method in the studied evaluations.

The fact that several of the evaluations utilize qualitative methods is an interesting observation. Several of the evaluations examined are not the type of goal and result evaluation usually associated with quantitative methods and the typical evaluation practice that characterizes New Public Management (NPM) (Hood, 1991). Rather, they are largely based on interpretation and understanding of user or stakeholder experiences. For example, in one of its reports, the public expert agency Growth Analysis examined how well state and regional business support responds to policy goals and the needs of entrepreneurs (Tynelius, 2016). This was done by comparing intentions and formulations in different documents with interview results and by interpreting and seeking an understanding of how entrepreneurs and prospective innovators perceive the support. Moreover, it should be mentioned that many of the evaluations studied are so-called mid-term evaluations, in which the evaluator examines whether established processes or application procedures match the goals of the policy. These mid-term evaluations are carried out when a project has begun or is half-finished and thus make it difficult to assess efficiency or effectiveness.

Finally, evaluators often base their reports on a mix of data sources. In our study, such data is defined as a combination of both objective data, defined as independent from the viewer and exemplified by, for instance, index data referring to company turnover, or gathered patents; and subjective data, like self-evaluations of people taking part in projects or other value statements from respondents. More than half of the evaluations studied, 67 of the 110, were based on mixed data. Twenty-three were based on subjective data (again, value statements from participants or beneficiaries)

Fig. 4 Data used in evaluations of Swedish growth and innovation policies
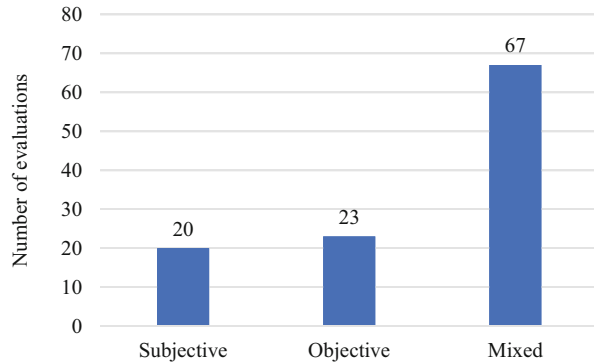
and 20 on objective data (index data). Figure 4 shows the frequencies of each data type used in the studied evaluations.

Public policy programs such as innovation policies are often quite complex in nature and studying different types of data to evaluate effects from such a policy hence seems a plausible approach. Apart from the variables presented above, two additional control variables were studied: *type of intervention* evaluated and *whether the evaluated program was ongoing or completed*. *Type of intervention* was coded in accordance with three possible types of interventions: *Financing intervention*, for example, grants or subsidies; *Rule changes*, such as permission to research new materials or regulatory relieves; and *Information efforts*, such as training in patent application or entrepreneurship. The evaluations examined concerned both completed and ongoing initiatives, which were coded by the dummy coding *ongoing* or *completed intervention*.

4.3 Evaluating Actors and Employed Methods

The next step in the analysis was to study the variation in evaluation judgments shown when divided based on the different types of actors. Among the 56 evaluations carried out by consultants, 45 (80.4%) were positive, the remainder neutral. For other types of actors, the distribution was much more even between the judgments distributed. Among other agencies, 11 (35.5%) evaluations were positive, 15 (48.4%) neutral, and 5 (16.1%) negative. Among researchers, 7 were positive, 7 neutral, and 1 negative; and among self-evaluations, there are 4 positive and 4 neutral evaluations. The results thus show that consultants provide considerably more positive evaluations than other actors. Figure 5 shows the frequencies of each judgment based on the actor conducting the evaluation.

To probe whether this correlation is statistically significant, Fischer's exact test was performed on the *actor* and *judgments* of evaluation variables (*p*-value: 0.001).

The dataset shows no major variation in the evaluations depending on which methods or data type they utilized, but great variation depending on the evaluating
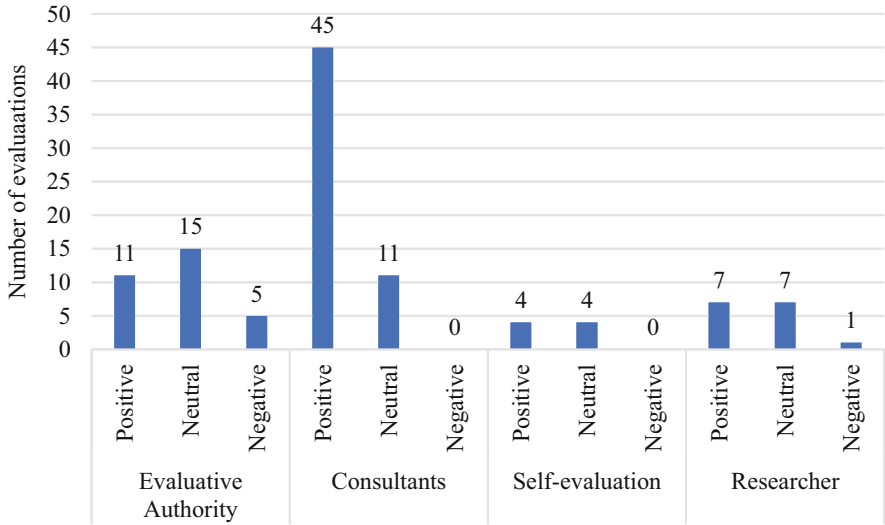
Fig. 5 Reviews of valuations of Swedish growth and innovation policies by evaluative actor
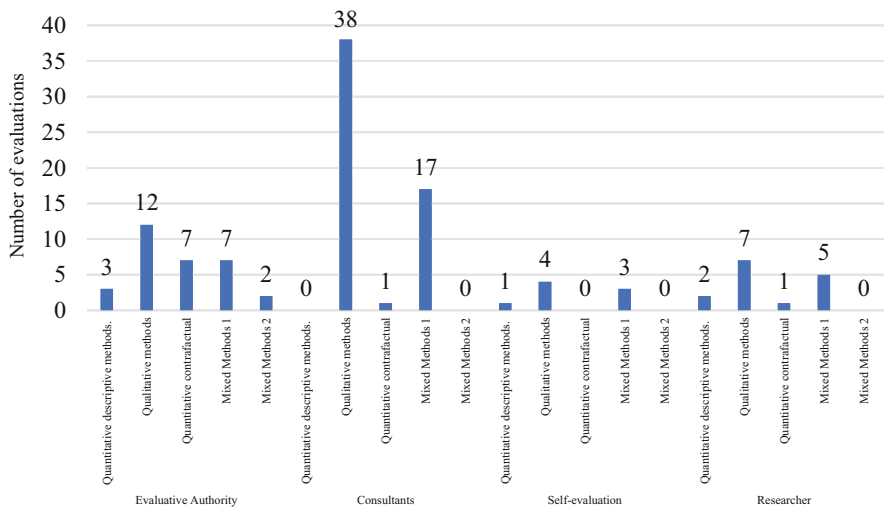

Fig. 6 Methods used in Swedish growth and innovation policy evaluation by evaluative actor

actor type. Figures 6 and 7 show frequencies of methods and data type based on the actor conducting the evaluation. The figures show a clear propensity among consultants to use qualitative and mixed methods while evaluative agencies have a slightly more even distribution between methods. The high number of qualitative methods could be attributed to the fact that a lot of the evaluations are conducted on ongoing projects, which makes quantitative approaches, often based on measuring effects
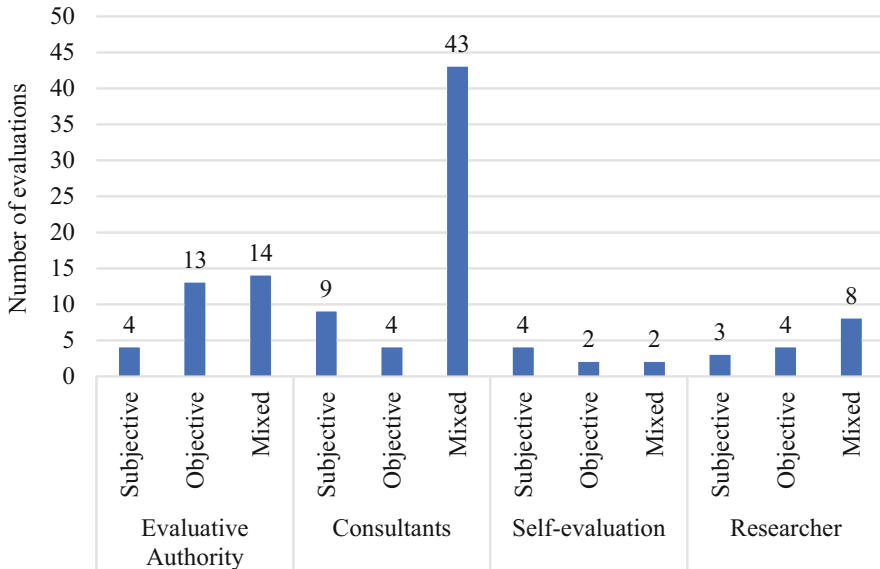
Fig. 7 Data used in Swedish growth and innovation policy evaluations by evaluative actor

through indicators such as employment, company turnover, or patents, somewhat difficult to perform.

To rule out other potential explanations and to map additional correlations, a logistical regression analysis was performed. The results, which are given in appendix 9.1, show high odds ratios and statistically significant p-values between a dummy for *positive judgment calls in the evaluations* and a dummy for the actor type *consultant*. We also observe a negative relationship between the method category qualitative methods and less positive evaluations, meaning that qualitative approaches are more like to result in positive evaluations. Notably, the "consultant effect" in terms of the strong correlation between the type of evaluating actor and their evaluations of policy remained statistically significant ($p = 0.02$), indicating that difference in, for instance, methods or data utilized in the evaluation, cannot explain the difference between different evaluators.

5 Discussion

Our results show that the vast majority of evaluations are positive, and few make use of quantitative evaluations in which real effects can actually be measured. Moreover, consultants are significantly more likely to conduct positive evaluations relative to the other evaluating actors. This does not seem to be due to the consultants using different methods, utilizing certain types of data material, or evaluating a certain type

of political intervention. What affects the results rather seems to be that it is specifically consultants that carry out the work.

The strong positive relationship between consultants as actors conducting evaluations and an evaluation being positive is the major finding of our study. Indeed, it is not a trivial finding. We have yet to confirm a causal relation between the two. In what follows, we, therefore, discuss possible explanations of the results and draw implications for future research.

5.1   What May Underlie Differences in Evaluations of Innovation Policy?

One plausible explanation could be found in public choice theory, according to which government agencies have an innate interest of looking out for themselves, partly through indicating positive results of their work (Niskanen Jr, 1994). Evaluated agencies can hence be expected to have strong incentives to choose evaluators they expect to give positive evaluations, since this gives them arguments for continued funding and support. Consulting companies are therefore likely, through competitive pressure, to become inclined to please their clients, which seems to mean that they come up with positive evaluations.

Vice versa, it can be argued that reviewing agencies such as the Swedish National Audit Office and Growth Analysis may have incentives to examine other agencies' efforts carefully and potentially more critically in order to identify problems and shortcomings and thereby justify their assignment as an examining agency.

Another explanation as to why consultants provide significantly more positive evaluations compared to other actors could be that they are hired to evaluate interventions that agencies already know have yielded positive results and therefore are considered easier to evaluate. Interventions that are more difficult to evaluate, and therefore often detect neutral or negative results, would, according to the same logic, be entrusted to evaluation agencies whose opinions should thus differ according to what we have observed. What speaks against such an interpretation would be that the positive evaluations studied often use methods that do not make it possible to draw conclusions on a scientific basis.

A potentially more reasonable interpretation of our results would be that evaluators are aware that the result affects the possibilities of obtaining further assignments from the agency in question. When a number of private, profit-maximizing companies compete with each other in a procurement procedure, significant sums are at stake. The winner of the procurement can hire additional staff at the next stage and charge by the hour in a way that benefits both superiors and shareholders. It would be strange if such an arrangement did not affect how evaluations are formulated, not least because this is a repetitive game in which the results from one evaluation can be expected to influence the outcome of the next procurement. The companies that carry out these evaluations are placed in an incentive structure in which it becomes very difficult to frame the results negatively.

Conversely, an evaluated agency is also in a challenging situation. With demands to be evaluated continuously and to report results to responsible politicians, the need for positive evaluations is apparent. As pointed out by the Swedish National Audit Office (2020), the government has, from time to time, presented results from evaluations to the Swedish parliament in more positive terms than are proportional to the results and methods of the evaluation. Such an observation also suggests that there is a demand for positive evaluations among responsible politicians. Thus, a pressure might exist on government agencies to generate positive results, as these are demanded by decision-making politicians.

Since the present study does not look at whether policies in the field actually work, it is not possible to determine exactly how these explanations should be judged. Further research in the field is hence important.

Every year, large sums are spent on innovation policies. Strictly speaking, results from the evaluation of any single policy can only be generalized to the specific policy intervention, and possibly similar ones carried out in the near future. Yet, there are well-articulated and important reasons for policy development and policy evaluation to "accumulate knowledge" and learn (Mazzucato, 2012). Hence, evaluating practices and quality remain central to any type of innovation policy that seeks to direct or enhance the sum, quality, or type of innovations in society.

The special nature of innovation policies, with limited funding in the form of often time-limited financial efforts, makes the results more difficult to directly apply to other policy areas. However, one area that is similar to innovation policy in this respect is foreign aid policy.

5.2 Future Research

Our novel approach to study a larger quantity of evaluations simultaneously has proven useful in exposing systematic differences between evaluators and could hence be beneficially utilized for similar future tasks. Future research could investigate incentives motivating evaluators, their relationships to the evaluated agencies, and their general evaluative competences.

One important thing to point out is that not every evaluation aims to measure both the *effectiveness* and *efficiency* of any single policy studied. Yet, a conceptual confusion exists between these concepts in the evaluations scrutinized. There are of course valid reasons to evaluate both of these concepts, by on the one hand evaluating the degree to which a policy is effective (i.e., whether it succeeds in meeting its goals), and on the other hand, evaluating its degree of efficiency (e.g., cost-effective vs expensive, simple vs cumbersome). Focusing on, for example, experiences of beneficiaries or the viewpoints of bureaucrats executing the policies provides valuable information that can help improve policy efficiency. However, and importantly: The latter type of more process-oriented evaluative methods should not, as is often done, be used to indicate *whether or not a policy is truly effective,* i.e., *accomplishing its designated results.*

From this perspective, it can be concluded that assessments of *effectiveness* seem to be almost completely absent among the studied evaluations, yet, many of the evaluations still contain phrases that can be interpreted as gauging policy effectiveness (i.e., goal accomplishment), even if that is not the explicit intention of the evaluation in question or if the methods employed do not enable assessment of policy effectiveness. It is one thing to evaluate whether or not one has achieved the expected goals but another to investigate whether one could have achieved the same with fewer resources, or achieved better effects with the same resources. If evaluations are to work as a safeguard of a society's common resources, such a perspective is truly warranted.

Moving from what can be said based on the study presented, an additional feature of the evaluative system operating close to the innovative policy field deserves to be mentioned: There are not that many agencies, firms, and people working with innovation policy (and likely other policy fields) and the evaluation of such policies in any smaller country. It is not uncommon that people start their career within an executive agency and then move to work for the evaluative branch of the complex, maintaining relationships with previous coworkers and the agency in question. In our pre-study, we came across examples of consultants winning procurements partly through such relationships or inside knowledge. While such relationships can of course provide good insight into how to evaluate, in quite a critical and efficient way they also demonstrate risks of—possibly unintentional—corruption. Studying the networks of people designing and executing policies, and those that evaluate the same policies, is therefore a pertinent area of study.

5.3   Policy Recommendations

Despite *evaluation* being a scientific field and higher education curriculum subject in economics, public policy, political sciences, psychology, and the educational sciences, there is no specific education or public certification for *evaluators* of public policy programs. Yet, evaluations of public policies proliferate, and today represent a large industry within and across countries. We have yet to discover if any specific common practice or ethos is present among evaluators but currently, few such indications have been found. An increased focus upon creating such an education or ensuring a common framework of evaluative practice could be an important step toward ensuring different types of evaluations are used and interpreted according to their separate purposes.

Another policy change to enhance evaluative practice could be to limit the type of evaluation allowed to be conducted (and financed) by executive agencies. It is important that such agencies are allowed to learn from and improve their implementation processes but to also assign the agencies the responsibility to evaluate their own policy efficiency or effectiveness is to create a system with distorted incentives. To solve this dilemma, such evaluations should be tasked to independent agencies and, in the case that private consultants are to be procured, such procurement should

involve criteria of both appropriate methods and independent practices. Such independence could potentially also be improved through some type of single-blinded system in which the agencies evaluated are unaware of who evaluated their policies.

6   Conclusion

In this chapter, we have explored evaluations of innovation policy. We add an important piece to the puzzle of innovation policy by studying a large sample of evaluations and looking for patterns across the data. Our results show that the overwhelming majority of evaluations are positive or neutral and that very few evaluations are negative. While this is the case across all categories of evaluators, we note that consulting firms stand out as particularly inclined to provide positive evaluations. The absence of negative or critical reports can be related to the fact that most of the studies do not rely upon methods that make it possible to discuss effects.

This discrepancy between so many positive evaluations on the one hand and comparatively weak evaluation methods, on the other hand, leads us to suspect that evaluators are not sufficiently independent. Consultants and scholars that are funded by a government agency in order to evaluate the agency's policies and programs are put in a position where it is difficult to maintain objectivity.

Our results indicate that further studies of how innovation policies are evaluated would be of interest, especially with regard to potential conflicts of interest.

Appendices

Fischer's Exact Test, Evaluating Actor and Evaluations of Public Innovation Policies

	Evaluations of Public Innovation Policies			
Actor Type	**Negative**	**Neutral**	**Positive**	**Total**
Evaluative agency	5	15	11	31
	(1.7)	(10.4)	(18.9)	(31.0)
Consultants	0	11	45	56
	(3.1)	(18.8)	(34.1)	(56.0)
Self-evaluation	0	4	4	8
	(0.4)	(2.7)	(4.9)	(8.0)
Researcher	1	7	7	15
	(0.8)	(5.0)	(9.1)	(15.0)
Total	6	37	67	110
	(6.0)	(37.0)	(67.0)	(110.0)
	p = 0.000			

Logistic Regression, Outcome Variable: Positive Policy Judgment

	Odds ratio	P-value
Evaluative agency	0.67	0.67
Consultant	9.00	0.02
Researcher	1.18	0.87
Quantitative descriptive methods	1.11	0.93
Qualitative descriptive methods	0.29	0.04
Quantitative Contrafactual methods	0.34	0.34
Subjective data	1.28	0.78
Mixed data	0.82	0.81
Financing policy[a]	5.91	0.00
Completed policy[b]	0.47	0.13
Constant	0.49	0.54
	n = 110	
	Pseudo R^2 = 0. 2485	

[a] Indicating a policy substantially consisting of grants or funding of a specific project. Compared to policies consisting of information, such as educative efforts or efforts to create networks or relationships between key innovative actors

[b] Indicating policies which, by the time of evaluation, were finished compared to ongoing policies. A majority of the studied evaluations were conducted on such, ongoing, policies

References

Bovens, M., Hart, P., & Kuipers, S. (2008). The politics of policy evaluation. In R. E. Goodin, M. Moran, & M. Rein (Eds.), *The Oxford handbook of public policy*. Oxford University Press.

Dahler-Larsen, P. (2007). Evaluation and public management. In E. Ferlie, L. E. Lynn Jr., & C. Pollitt (Eds.), *The Oxford handbook of public management*. Oxford University Press.

Daunfeldt, S.-O., Tingvall, P. G., & Halvarsson, D. (2016). Statliga innovationsstöd till små och medelstora företag–har de någon effekt? *Ekonomisk Debatt, 44*(1), 6–19.

Furubo, J.-E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation (Vol. 2002)*. Transaction Publishers, cop.

Goodin, R. E., Moran, M., & Rein, M. (Eds.). (2008). *The Oxford handbook of public policy*. Oxford University Press.

Gustavsson Tingvall, P., & Deiaco, E. (2015). *Do selective industrial policies cause growth?* Tillväxtanalys.

Hood, C. (1991). A public Management for all Seasons? *Public Administration, 69*(1), 3–19.

Hunt, R. A., & Kiefer, K. (2017). The entrepreneurship industry: Influences of the goods and services marketed to entrepreneurs. *Journal of Small Business Management, 55*, 231–255.

Karlson, N., Sandström, C., & Wennberg, K. (2019). *Markets or bureaucrats in innovation policy*. Ratio.

Knill, C., & Tosun, J. (2012). *Public policy a new introduction*. Palgrave Macmillan.

Mazzucato, M. (2012). *Mission economy: A moonshot guide to changing capitalism*. Penguin.

Mazzucato, M. (2018). Mission-oriented innovation policies: Challenges and opportunities. *Industrial and Corporate Change, 27*(5), 803–815.

Niskanen, W. A., Jr. (1994). *Bureaucracy and public economics (2:Nd ed., Vol. 1994)*. Edward Elgar.

Pollitt, C. (2003). *The essential public manager*. McGraw-Hill Education.

Power, M. (1997). *The audit society–rituals of verification*. Oxford University Press.

The Swedish National Audit Office. (2020). *En granskningsrapport från riksrevisionen* (s. 59). Riksrevisionen.

Scriven, M. (1991). *Evaluation Thesaurus* (Fjärde Upplagan, Vol. 1991). SAGE.

Tynelius, U. (2016). *De regionala företagsstöden (2016:01; s. 106) [Regional firm support systems]*. Tillväxtanalys.

van den Berg, C., Howlett, M., Migone, A., Howard, M., Pemer, F., & Gunter, H. M. (2019). *Policy consultancy in comparative perspective: Patterns, nuances and implications of the contractor state*. Cambridge University Press. https://doi.org/10.1017/9781108634724

Vedung, E. (2009). *Utvärdering i politik och förvaltning [Evaluation in politics and administration]*. Studentlitteratur.

Ylönen, M., & Kuusela, H. (2019). Consultocracy and its discontents: A critical typology and a call for a research agenda. *Governance, 32*(2), 241–258. https://doi.org/10.1111/gove.12369