JPM
JOURNAL OF POSITIVE MANAGEMENT
Vol. 10, No. 2, 2019, pp. 42–55

ISSN 2083-103X

http://dx.doi.org/
10.12775/JPM.2019.003

EVALUATION OF PUBLIC INTERVENTIONS IN A COMPLEX ENVIRONMENT: DEVELOPING GENERALIZABLE KNOWLEDGE FROM CASE STUDIES

Paulina Kubera

Poznan University of Technology, Poznan, Poland

e-mail: Paulina.Kubera@put.poznan.pl

Abstract

Purpose: In the debate on how to increase the effectiveness of public policy instruments, learning- oriented evaluation attracts considerable attention. The focus of interest has been shifted from 'what' questions to 'why' and 'how' questions (i.e. instead of asking what works/does not work we want to know why a particular public intervention works/ does not work). However, implementing public interventions in a complex environment which is characterised by feedback loops, adaptation by both- those delivering and those receiving the intervention does not allow to establish universal truths that apply anywhere, anytime. On the contrary, context matters and human agency cannot be taken for granted. Thus, we are more specific in our inquiry asking 'what works for whom in what circumstances'(the stance of the realistic evaluation approach). Case studies which have the explanatory power, do not necessary have to serve for one-off, discrete evaluation. The aim of this article is to address the dilemma of developing generalizable knowledge from case study research and on the basis of the extant evaluation literature, suggest approaches to enhance its external validity to enable the middle-ranged theories formulation, i.e. 'law-like'- regularities delimited in time and space, which can be used for learning beyond a particular case.

Methodology: The article has been written following a careful review of leading literature in the subject as well as a review of evaluation reports from the Science and Innovation Policy Evaluations Repository (the SIPER database) to provide insights into evaluation practice.

Findings: A case study approach is well recognised in evaluation practice in the field of research, development and innovation, however its full potential has not been exploited in terms of drawing lessons for future public interventions.

Originality/value: Given the complexity surrounding numerous public interventions the article suggests a wider utilisation of case study approach in evaluation along with the techniques to enhance the generalisability of knowledge gained from case study research.

Keywords: governance, theory-oriented evaluation, case study, casual mechanisms, realist synthesis

Paper type: Theoretical paper

1. Introduction

There is a growing need expressed to view public interventions [1] not only through their effects but also to consider how those effects are produced. This implies collecting better information on the conditions that influence an intervention's success or failure, delving into the inner (hidden) workings of a public intervention. The problem is that we live in a complex world and the actions taken mean intervening in a system which consists of many components that interact with each other. Distinct properties arise from those relationships [2]. As Byrne (2013) describes it aptly: '…in complex systems the cause will seldom be the intervention – something done to the system – taken alone. What matters is how the intervention works in relation to all existing components of the system and to other systems and their sub-systems that intersect with the system of interest' (p. 219). This undermines the possibility to transfer the knowledge gained in one setting elsewhere and at other times. However, does it mean that we should abandon efforts to build the broader evidence base for policy making as they are doomed to fail? On the contrary. Though, it requires taking a different approach- mechanism-based theorizing, what involves more than just establishing correlation among variables. These mechanisms should be distinguished from programme activities as they are cognitive, affective, social responses to an intervention which lead to desired outcomes (Weiss, 1997). They exhibit some regularities which can provide useful insights while developing new policy instruments in relevant settings. At the same time, however, we have to come to terms with the fact that the mechanism-based analysis will not equip us with universal laws, but rather law-like regularities, delimited in time and space. They are referred to, after Robert Merton, as 'middle-range theories', and are placed between universal social laws and mere description.

The aim of the article is to suggest how to build an evidence –based for policy making using a case-study approach, which seems to be an appropriate choice given the complexity surrounding numerous public interventions. This, in turn, implies tackling the perennial issue of developing generalizable knowledge from case study research.

The paper proceeds as follows. As a case study design is particularly relevant for those evaluations which are oriented on learning, (and not entirely on accountability), in the first section of the paper two main approaches to evaluation are discussed. The traditional one, which is frequently referred to as the 'black box' evaluation, with a focus on the input and output side of an intervention and the 'white box' evaluation, where all theory- oriented evaluations belong. This is the latter one which aims to unpack the black box to inspect the inner part of an intervention, the logic of a program. These two take different approaches, utilize different methods and ask different questions concerning the usability of the evaluation findings.

The second part of the article highlights the concept of casual mechanisms for explaining how and why public interventions work. It has its values and limitations, but importantly - the mechanism-based approach enables to work out the problem how multiple case studies can be aggregated or synthesized. In the third part several suggestions are put forward to enhance case study external validity in the context of evaluation practice in the field of science and innovation policy.

2. Unpacking 'black boxes' and theory- oriented evaluation

Evaluation traditionally has been organsied around the question whether a given public intervention (programme) works. However, recently more attention is paid to the variation in program effects and mechanisms through which these effects occur. Instead of asking 'what works' the question is 'what works for whom in what conditions'. Addressing these kinds of questions means moving beyond measuring average impacts of a public intervention and trying to understand what is inside the 'black box' of a policy instrument (Solmeyer and Constance, 2015; Granger and Maynard, 2015). Hence, these two types of evaluation are often referred to as 'black box' evaluations and 'white box' evaluations (Astbury and Leeuw, 2010), or method-led evaluations versus theory-led evaluations.

The 'black box' evaluation focuses on the input and output side of a given intervention. It deals with quantitative questions such as: how much has changed something? What was the direction of the change? How much was invested in a given program (input)? How much was achieved owing to its implementation (output)? It means a comparison of the situation before and after a given intervention and calculating its average effect. If entities who received a treatment (the programme beneficiaries) are, on average, better off than those who did not receive a treatment, the programme works. Quantitative questions imply quantitative research methods. Although, such evaluations shed some light on the efficiency and effectiveness of a given program they say little what should be done when average program effects are negative or not significant. Moreover, these overall outcomes hide differential outcomes for different sub-groups of the programme participants. Westhorp (2013), for instance, gives an account of the evaluation findings on The Early Head Start (EHS), an American program aimed at improving a range of outcomes for children over the life course by reducing risk factors and fostering protective factors, and while the evaluation found positive outcomes overall, it turned out that for more disadvantaged families the outcomes were less positive or on occasions even negative. Thirdly, if policy-makers are interested in gaining knowledge about the transferability of a programme into a different context or scale it up, they need to unpack the 'black box' of a given intervention and investigate the mechanisms through which the effects are produced.

To fill the above mentioned deficits the 'white box' evaluation, or in other words - the theory-oriented evaluation [3] has emerged, which rests on the idea of using a programme's theory as an 'explanatory account' of how the programme works (Schmitt and Beach, 2015). It can be defined as the analysis and valuation of the contribution made by an intervention to solve the identified social problems. The starting point in theory-oriented evaluations is provided by the objectives and assumptions on which a given intervention is based (Van der Knapp, 2004). Hence, it can be argued that public policy programmes are embodiments of theories in two ways. First, as they incorporate the expectation that a given intervention will lead to the desired outcomes (the alleviation of social problems). Secondly, as they rest on a set of assumptions about how and why programme activities and resources will produce the change. What is imperative is to establish a chain of evidence, make these programme assumptions explicit and test them empirically in a robust way (Astbury and Leeuw, 2010), ruling out the rival explanations. A programme theory can be constructed using a variety of methods such as: observation of the program in action, interviews with programme implementers, programme participants, programme document analysis, concept mapping exercises, investigation of research on similar initiatives or social science theory.

One of the forms of theory-oriented evaluation is realistic (realist) evaluation developed by Pawson and Tilley (1997). What sets this evaluation approach apart from other theory-oriented approaches is that a realist evaluator has 'a more explicit intent in uncovering programme theory. Such theory, rather than being about the nuts and bolts of programmes and their possible linkages, is more concerned with psychological and motivational responses leading to behaviour change' Thus, a realist evaluator tries to 'see what initiative fires in people's minds. This is what a realist means by mechanisms' (Blamey and Mackenzie, 2007, p. 446). Pawson and Tilley (1997) stress: 'we cannot simply treat programs as things, we have to follow them through into the choices made by recipients'(p.188). Qualitative methods and case study design [4] are commonly used by realist evaluators (Maxwell, 2004; Riege, 2003). They are also committed to constantly refine learning – they draw on other studies when formulating their theories and end with more refined propositions. That is why this approach is worth special attention from the point of view of the topic of this article. Moreover, many authors point at the suitability of the realist evaluation approach to investigating complexity (e.g. Astbury 2013; Marchal et al., 2012; Woolcock 2014).

3. The concept of mechanism: delving into the inner (hidden) workings of a public intervention. Realist synthesis

As a starting point, it has to be noted that the focus of evaluation endeavour does not need to be a public action per se (a programme), but 'interesting, puzzling,

socially significant regularities' (Pawson and Tilley, 1997, p.71), that might be relevant for broader application, and are referred to as ' mechanisms.' Public interventions usually carry not one, but more implicit mechanisms of action. Therefore, the success of a particular intervention depends on the cumulative success of the sequence of these mechanisms as the programme unfolds (Pawson et al., 2004). By relying on the accumulated evidence of the mechanisms at work, instead of an intervention as such, policy makers are in the position to notice that many seemingly 'novel' interventions share in fact common underlying mechanisms of change.

The concept of 'mechanism' has been introduced to evaluation research thanks to Chen and Rossi (1989), who also stressed the importance of theory-oriented evaluation in better understanding the casual linkages between the treatment and effects. The suitability of the concept for bringing greater explanatory power to evaluation has been also recognised by Weiss (1997) or Donaldson (2007). However, these are Pawson and Tilley who made it imperative to identify the trio of explanatory components known in the terminology of realistic evaluation as the 'context-mechanism-outcome configurations' (CMOs). Thus, the explanatory theory sought by realistic evaluators is a 'generalizable mechanism that explain why an individual or group of individuals (within a particular context) respond in a particular and relatively predictable way to an intervention (or aspects of an intervention)' (Blamey and Mackenzie, 2007, p. 446). The stress is on the characteristics of the programme recipients (their individual background) and the context as they both influence the outcome. As Pawson et al. (2004) point out: "The hard slog of realist synthesis is about building up a picture of how various combinations of such contexts and circumstance can amplify or mute the fidelity of the intervention theory" (p. iii).

This is in contrast with the 'successionist' model of causal explanation or variance theory that fit in the positivist/empiricist position. Successionists examine the associations between variables, which are considered as 'the vital causal agents'. In the first step, a variable that capture 'the ouput/ outcome' is identified (the dependent variable 'Y') and then other explanatory variables (the independent variable(s) 'X') which are considered to be responsible for the variation of 'Y', or to put it simply, influence the outcome of an intervention. It is based on an analysis of the contribution of differences in values of particular variables to differences in other variables. A fundamental critique of the successionist approach lies in neglecting the contextual features. Barnes, Matka and Sullivan (2003) raise an argument against interpreting the context as a purely external factor, as context is, on one hand, shaped by the actors, and on the other hand, constraints their actions. Moreover, the successionist approach is unable to explain the causal connection, i.e. the process leading from cause to effect, what happens in-between cause and effect, therefore typically involves a 'black box

approach' to the problem of causality. Generative causation, which underpins the realist perspective, by contrast, attaches great importance to this transformation, as it tries to provide fine-grained explanation of the behaviour of specific actors (thinking, decision-making, action) in a given context with specific resources, opportunities and constraints (financial benefits, social rewards, institutional structures, anything that constitutes an incentive or an obstacle for a specific behaviour or decision). (Befani, 2016). As Pawson (2003) notes: 'Intervention work when the resources on offer (material, cognitive, social, and emotional) strike a chord with programme subjects'. However, 'programme resources resonate much more for certain subjects in certain contexts' (p.473–474). While other accounts, based on variables and attributes, taken in different settings result in a change of all the coefficients and configurations, the great advantage of the generative account lies in the fact that it describes the processes that are generic, strengthen our understanding of how and why public interventions work, with whom, and under what circumstances and can be used for a family of related interventions (Pawson, 2006).

Ylikoski (2018) makes a useful distinction between causal scenarios and mechanism schemes. While the former denotes a representation of a particular process responsible for some concrete event or phenomenon, the latter one is more of an abstract nature. The belief-formation mechanisms such as self-fulfilling prophecy is a good example of a mechanism scheme that finds its application in many various settings. It occurs when 'an initially false belief of a situation evokes behaviour that eventually makes the false conception come true' (Hedström and Swedberg, 1998, p.18). This mechanism scheme can be used to explain various phenomena such as placebo, the Hawthorne effect, or how teachers interact with disadvantage students (i.e. if a teacher expects disadvantage students to underperform, they will probably underperform) (see: Astbury and Leeuw, 2010). Concrete application of the this social mechanism provides also Ylikoski (2018) who reconstructs Espeland and Sauder's (2016) case study of the effects of rankings on US legal education. The author argues that case studies can contribute to the 'theoretical toolbox' in several ways. First, as they provide evidence about a new mechanism or a combination of already known mechanisms. Second, they can be instrumental to enhance our understanding of a particular mechanism, e.g. about the necessary background conditions for a mechanism to operate. Third, they can help to learn about other effects of the mechanism that proved to be substantial diagnostic evidence about its operation. Ultimately, they can indicate avenues for future theoretical development 'by bringing to the fore puzzles that show the limits of the current theoretical ideas' (Ylikoski, 2018, p. 4).

4. External validity of case study research

Yin (2018) suggests that one should favour choosing case study research, compared with the others, when: (1) the main research questions are 'how' and 'why' questions, (2) the researcher has little or no control over behavioural events, and, (3) the focus of study is a contemporary (as opposed to entirely historical) phenomenon. At the same time Yin (2018) admits that: 'Doing case study research remains one of the most challenging of all social science endeavours'(p. 3), one should acknowledge its strengths and limitations. Evaluation which is focused on learning attempts to develop knowledge that is in some way generalizable. However, despite the advantages of the case study method (as it allows for an extensive 'in-depth' description of the phenomenon under study, grasped in its totality), its external validity is doubtful, to say the least. In crude terms, external validity refers to the validity of applying the findings of the study outside the context of that study, e.g. across other locations, actors or times.

For the ensuing discussion, it is essential to clarify the type of generalisation expected form case study-based evaluation. While quantitative research aims at statistical generalisation as a form of attaining external validity, case study research is grounded in analytical generalisation, which is distinct form statistical generalisation in that it does not draw inferences from data to population, i.e. the findings from a sample are not claimed to apply to its universe. In analytical generalisation particular findings are generalised to some broader theory. According to Yin (2018) analytical generalisation is 'the logic whereby case study findings can apply to situations beyond the original case study, based on the relevance of similar theoretical concepts or principles' (p. 286). Thus, in the field of evaluation this type of generalisation involves making projections about the application of findings form one study, based on a theoretical analysis of the factors producing outcomes and the effect of context. Realistic evaluation with its context-mechanism-outcomes configurations are especially predestined for this purpose.

In the subsequent part three techniques will be suggested to enhance external validity of case study research and be confronted with evaluation practice in the field of science and innovation policy. These are: comparing evidence with extant literature, defining scope and boundaries of reasonable analytical generalisation for the research and use of replication logic in multiple-case studies. The SIPER database (The Science and Innovation Policy Evaluations Repository) maintained by the RISIS Horizon 2020 project will serve as a reference base to shed some light on evaluation practice in the subject matter. The database contains evaluations in the field of science and innovation policy conducted after the year of 2000 with the focus on the OECD countries.

Taking into account the data analysis method, 229 out of 539 evaluations (42,5%) gathered in the database so far [5] use case study. This makes case study

the third most used method in evaluations of public interventions supporting science and innovation, following: descriptive statistics (493) and qualitative/ quantitative text analysis (343). However, it is almost never used as a single research method [6], but as one of the several methods in mixed method design. The rationale of using case study is usually: (1) to obtain more in-depth knowledge on problems which cannot be captured by other methods (such as econometric analysis, descriptive statistics etc.), (2) to illustrate certain topics within an evaluation, or (3) to present good practices in the field.

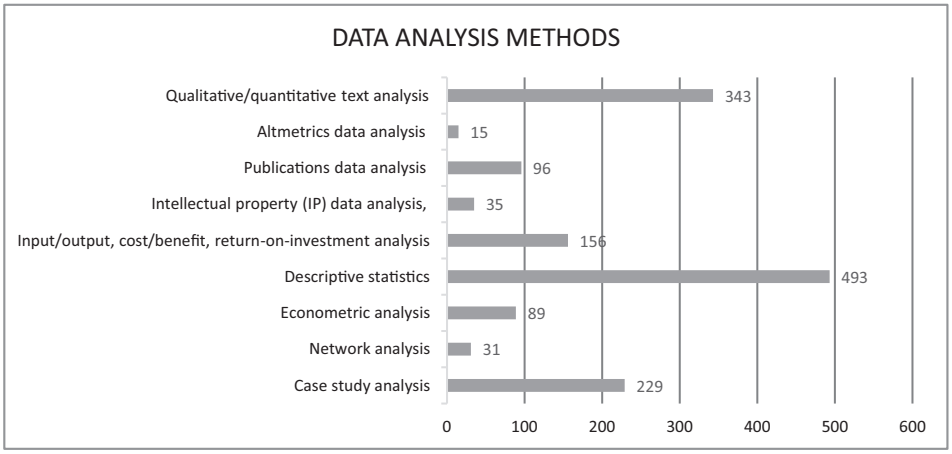

Figure 1. Data analysis method used in evaluations of science and innovation policy

Source: own elaboration based on the SIPER.

In order to enhance external validity of case study research as a starting point, it is advisable to link analytic generalisation to the related research literature by means of investigating overlaps as well as gaps (Yin, 2013). The majority of the evaluations gathered in the SIPER has literature review as one of the activities provided for in the research design or while formulating inferences makes at least explicit references to the extant literature or other studies in the field, making the final conclusions more credible. 'A formative evaluation of Collaboration for Leadership in Applied Health Research and Care (CLAHRC): institutional entrepreneurship for service innovation'(2014) can serve as an example. In order to better understand the context of an initiative, the evaluation team have conducted a literature review relating to knowledge translation (i.e. the exchange and utilisation of knowledge in practice) in health care and institutional entrepreneurship, on the basis of which they developed five schematic archetypes of knowledge translation as a means of framing their research. As they admitted: 'institutional theory provides helpful analytical concepts with which to understand the disciplinary knowledge silos and contrast ways of organising for knowledge production and its application' (p. 6). Accordingly, evidence gathered in

conducted case studies were linked to the extant theory, and the authors discussed the convergence and research gaps.

Secondly, what is imperative for reasonable analytical generalisations is defining scope and boundaries while designing case study research, as it sets empirical and theoretical limits on the extent to which an inference can be generalised. They are reflected in the research questions, stated propositions and the 'case(s)' identified for being studied. In reference to the explanatory case study, Aus (2005) writes about the 'scope conditions', that is 'the circumstances or a set of institutional and political conditions under which a causal mechanism or set theoretic relationship between causal mechanisms empirically holds' (p.4). One of the methods in multi-case design which has the potential for specifying the scope conditions of theoretically competing causal mechanisms and to study the relationship between context and outcomes in projects (see e.g. Verweij and Gerrits, 2012) is the qualitative comparative analysis (QCA) [7]. However, none of the evaluations in the SIPER Repository applies this approach and generally the evaluations under investigation are not specific about scope and boundaries of reasonable analytical generalisation. Apparently, the reason for it lies in the purposes for which the evaluators resort to case studies. They are descriptive and exploratory case studies rather than explanatory. In this regard I see great potential to enhance use of case study in evaluation research. In the next step – replication logic in multiple case study will be proposed, however, the argument is, that a proper within-case analysis is a prerequisite for any ensuing analytical generalisation. It is worth quoting here one of the most discussed study in comparative social sciences aptly conveying the idea- Theda Skocpol's States and Social Revolutions (1979). It is argued that what makes her study convincing is not her cross-case comparison but more her analysis of revolutionary processes. Hence, 'the fact is that many comparative case studies drew their strength less from the way they compare cases but more by their within-case analysis' (Blatter and Haverland, 2012, p. 3)

The third way to enhance external validity of case study is replication logic in multiple-case studies. Replication can be literal or theoretical (Yin, 2018). Literal replication refers to the situation when cases are selected for the study to predict similar results, that is, to corroborate each other. A previously developed theory serves as a template with which we compare the empirical results of the case study. Replication is claimed when two or more cases support the same theory. Theoretical replication, in turn, means a situation where cases are selected to predict contrasting results, however, for anticipated reasons. In both types of replication a theoretical framework/theory have to be developed that states the conditions under which a particular phenomenon is likely to be found (a literal replication) and the conditions when it is not likely to be found (a theoretical replication). For example, in "Evaluation of the investment

readiness demonstration projects and fit4finance' (2004) the Mason & Harrison model to enhance 'investment readiness' amongst SME population has been utilised to assess the effectiveness of public support in six cases-the projects, in 'Evaluation of the Networks of Centres of Excellence Program (2007) the appropriateness of a network structure model is tested in each of the eight case studies – the NCR networks. Although muti-case design, i.e. the situation when the same study contains more than a single case, is a norm in evaluation practice under investigation, nevertheless, a replication logic is frequently missing as there is no theory developed against which cases are examined and which can be used to generalise to new cases. At best a common themes are identified in cases under investigation which led to the conclusions further reinforced by the findings from other sources of data. For example, in ' Impact assessment of the SME-specific measures of the Fifth and Sixth Framework Programmes for Research on their SME target groups outsourcing research (2010), based on the case studies, eight factors have been identified that contribute positively to the success of a project (p. 11–12). In a similar vein, in ' Implementation Evaluation of the Comprehensive Rural Development Programme' (2013), eighteen case studies have been conducted around four core themes which allowed for the comparisons of the findings in each particular case. The next example is: 'Managing Innovation Prizes in Government' (2011), which draws on case studies widely recognised as successful technology programmes to develop a list of key factors and recommendations to increase the impact of prize programmes that are articulated in general terms so as to be applicable to a broader range of types of prizes and technologies (p. 28–29). However inspirational they might be, the weakness of such approach is, that the knowledge gained is rather fragmented and do not form a mechanism (e.g. the context-outcome-mechanism configuration). The point is that these are the combinations of aspects of cases (factors) that produce an outcome, or in other words – the effect of a an aspect of a case is contingent upon the other conditions. It must be acknowledge that there might be not a single success recipe and different conditions (aspects of a case) can produce the same outcome as well as the same condition (aspect of a case) can produce different outcomes depending on the other coexisted conditions. Understanding the links between them are instrumental for explanatory case study and hence for making causal interferences that are of the paramount importance in evaluation of public interventions.

5. Conclusions

Case study as an evaluation tool provides the following advantages. First, it enables to capture the complexity of an evaluand (a subject of an evaluation, e.g a programme), including relevant changes over time. Second, it takes fully into

account the contextual conditions, also those which potentially interact with the case. And although other evaluation methods are in the position to assess the outcomes of an intervention, case study offers the opportunity to examine the relevant processes (Yin, 2018). The typical research questions formulated in the case study research are 'how' and 'why' questions (e.g. how a given intervention will lead to a specific change). Thus, it can be safely assumed that case study approach is in the position to play an instrumental role in evaluation oriented on learning. Its appropriateness for evaluation of public interventions in a complex environment, such as those implemented within science and innovation policy, is recognised in evaluation literature (e.g. Verweij and Gerrits, 2012; Befani, 2013).

In fact, the examination of methods used in evaluations included in The Science and Innovation Policy Evaluations Repository leads to the conclusion that case study is a recognised method (42% of all of the evaluations in the repository use case study approach and it is the third most used method in evaluations under investigation). Nonetheless, as it has been demonstrated in the article, the full potential of case study has not been exploited. It is almost never applied as a single research method, but to complement other methods used. This is apparently for the fear that case study is by its nature low in external validity. However, while findings about a particular case may not be generalisable the underlying principles (mechanisms) often are. Simons (2015) calls this something of a paradox saying: 'the 'real' strength of case study lies in the insights we gain from in-depth study of the particular. If we study the singular case in sufficient depth, and are able to capture its essence – what makes it unique – in all its particularity, (…) we will also discover something of universal significance'. Hence, 'the more you capture the particulars of one person, context, programme, policy, its context and circumstance, the more likely you are to discover something universal' (p. 181), or, to put it more modestly – something which can be applicable beyond the particular case.

The central assertion of the article is that context and human agency matter and they are difficult to capture by other approaches such as (quasi-) experimental designs. Although the problem as such is not new and is addressed by many scholars still it seems justified to restate the value of using case study- based evaluations in a political climate that privileges inferences from large sample studies and experimental (quasi-experimental) designs. What is imperative though is to resort to measures which will enhance external validity of case study research. The article suggests three such measures: comparing evidence with extant literature, defining scope and boundaries of reasonable analytical generalisation for the research and use of replication logic in multiple-case studies. While the first idea, can be concluded, is realised in evaluation practice the other two are almost not existent in the evaluation practice. The problem of generalisation, i.e. the relevance of the conclusions reached for other cases, not

under the investigation, is hardly addressed at all. Recommendations are, as a rule, formulated for the here and now, case studies are conducted for descriptive and exploratory purposes. However, it is not surprising taking into account the current institutional arrangements in which professional evaluators operate that can be characterised by market demands and contractual obligations, evaluation agendas driven by sponsors who want specific answers to their specific questions (Astbury, 2013). Presenting successful cases are more welcome than delving into intricate workings of public interventions leading not always to the desired outcomes. Such situation is not conducive for building a broader evidence base about how generic types of intervention function. However, with the emphasis shifted towards evaluation oriented on learning there is a great potential of case study approach to be utilised in evaluation research, where the problem of generalisation of findings are addressed explicitly. A promising avenues for future research include further integrating qualitative methods with formal and statistical methods (e.g. fussy set qualitative comparative analysis), or those attempts which aim to introduce more rigour in the selection process of case studies. In order to ensure cumulation of knowledge it is imperative to select cases for investigation self-consciously and with a view to maximising inferential leverage.

Notes

[1] 'Public interventions' is a broad term denoting a wide repertoire of public actions ranging from the coercive measures (such as requirements, prohibitions), through the catalytic instruments – those that establish external catalysts to induce the desired behaviour (e.g. financial incentives) to hortatory instruments, which rely more on the use of symbols and values to motivate to the desired behaviour (e.g. labelling). A unit of analysis of an evaluation is typically a project, programme or a whole policy. However, as argued in the article it should not be necessarily the case.

[2] They can be characterised, among others, by non-linearity, emergence, dynamics, adaptation, uncertainty and co-evolution (Patton 2011).

[3] The term 'theory-oriented' evaluation is used to avoid confusion and denote ,inter alia, 'theory-driven' evaluation by Chen and Rossi, 'theory-based evaluation by Weiss or realistic evaluation by Pawson and Tilley.

[4] While case study research is frequently associated with qualitative research methods, a case study researcher in his or her quest to understand a given phenomenon in-depth uses whatever data is available, either qualitative or quantitative.

[5] Access: 12/09/192019.

[6] Rare example are 'The National Institute for Health Research at 10 years. An impact synthesis. 100 Impact Case studies' (2016) RAND Europe and the Policy Institute at King's, „The nature, scale and beneficiaries of research impact. An Initial analysis of Research Excellence Framework (REF) 2014 impact case studies" (2014) King's College London and Digital Science, Ex-post Impact Assessment FP6 sub-priority "Global Change and Ecosystems" (2008), „Evaluation of the investment readiness demonstration projects and fit4finance (2004) SQW Limited.

[7] Qualitative Comparative Analysis (QCA), introduced by C.Ragin (1987) is a hybrid alternative that integrates the generic patterns of variable-oriented approach with the idiosyncratic events of

case-based studies. It overcomes the trade-off between a fine-grained understanding of complex causal relations and the ability to generalise findings from small, medium and large number of cases. It handles asymmetric and multiple-conjunctural causality additionally to counterfactual reasoning. This allows necessity and sufficiency to be analysed separately, recognising the relevance of causal packages and multiple causal paths leading to the same outcome (see: Befani, 2013).

References

Astbury, B. (2013), "Some reflections on Pawson's Science of Evaluation: A Realist Manifesto", *Evaluation*, Vol. 19 No. 4, pp. 383–401.

Astbury, B., Leeuw, F. (2010), "Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation", *American Journal of Evaluation*, Vol. 31 No. 3, pp. 363–381.

Aus, J. (2005), "Conjunctural Causation in Comparative Case-Oriented Research. Exploring the Scope Conditions of Rationalist and Institutionalist Causal Mechanism", ARENA Working Paper No. 28, November 2005.

Barnes, M., Matka, E., Sullivan, H. (2003), "Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems", *Evaluation*, Vol. 9 No. 3, pp. 265–284.

Befani, B. (2013), "Between complexity and generalization: Addressing evaluation challenges with QCA", *Evaluation*, Vol. 19 No. 3, pp. 269–283.

Befani, B. (2016), "Causal frameworks for assessing the impact of development programmes", UEA seminar, 9 March 2016.

Blamey, A., Mackenzie, M. (2007), "Theories of Change and Realistic Evaluation. Peas in a Pod or Apples and Orangaes", *Evaluation*, Vol. 13 No. 4, pp. 439–455.

Blatter, J., Haverland, M. (2012), "Two or three approaches to explanatory case study research? Paper prepared for the presentation at the Annual Meeting of the American Political Science Association", New Orleans, August 30-September 2, 2012.

Byrne, D. (2013), "Evaluating complex social interventions in a complex world", *Evaluation*, Vol. 19 No. 3, pp. 217–228.

Granger, R., Maynard, R. (2015), "Unlocking the Potential of the "What Works" Approach to Policymaking and Practice: Improving Impact Evaluations", *American Journal of Evaluation*, Vol. 36 No. 4, pp. 558–569.

Hedström, P., Swedberg, R. (1998), *Social mechanisms: An analytical approach to social theory*, Cambridge University Press, Cambridge.

Marchal, B., van Belle, S., van Olmen, J., Hoerée T., Kegels G. (2012), "Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research", *Evaluation*, Vol. 18 No. 2, pp. 192–212.

Maxwell, J. (2004), "Using Qualitative Methods for Causal Explanation", *Field Methods*, Vol. 16 No. 3, pp. 243–264.

Patton, M. (2011), *Developmental evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, SAGE, Thausand Oaks.

Pawson, R. (2003), "Nothing as Practical as Good Theory", *Evaluation*, Vol. 9 No. 4, pp. 471–490.

Pawson, R. (2006), *Evidence-based policy: A realist perspective*, SAGE, London.

Pawson, R., Greenhalgh, Y., Harvey, G., Walshe, K. (2004), "Realist synthesis: an

Introduction", ESCR Research Methods programme, University of Manchester RMP methods Paper No. 2.

Pawson, R., Tilley, N. (2007), *Realistic Evaluation*, SAGE, London.

Ragin, C. (1987), *The Comparative Mthod: Moving Beyond Qualitative nad Quantitative Strategies*, University of California Press, Berkley.

Riege, A. (2003), "Validity and reliability tests in case study research: a literature review with 'hands-on' applications for each research phase", *Qualitative Market Research: An International Journal*, Vol. 6 No. 2, pp.75–86.

Schmitt, J., Beach, D. (2015), "The contribution of process tracing to theory-based evaluations of complex aid insruments", *Evaluation*, Vol. 21 No. 4, pp.429–447.

Simons, H. (2015), "Interpret in context: Generalizing from the single case in evaluation", *Evaluation*, Vol. 21 No. 2, pp. 173–188.

Skocpol, T. (1979), *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*, Cambridge University Press, Cambridge.

Solmeyer, A., Costance, N. (2015), "Unpackingthe 'Black Box' of Social Programs and Policies: Introduction", *American Journal of Evaluation*, Vol. 36 No. 4, pp. 470–474.

Van der Knapp, P. (2004), "Theory-based Evaluation and Learning: Possibilities and Challenges, *Evaluation*, Vol. 10 No. 1, pp. 16–34.

Verweij, S., Gerrits, L. (2012), "Understanding and researching complexity with Qualitative Comparative Analysis: Evaluationg transportation infrastructure projects", *Evaluation*, Vol. 19 No. 1, pp. 40–55.

Weiss, C. (1997), "Theory-based evaluation: Past, present and future", *New directions for Evaluation*, No. 76, Jossey-Bass, San Fransisco, pp. 41–55.

Westhorp, G. (2013), "Developing complexity-consistent theory in a realist investigation", *Evaluation*, Vol. 19 No. 4, pp. 364–382.

Woolcock, M. (2013), "Using case studies to explore the external validity of 'complex' development interventions", *Evaluation*, Vol. 19 No. 3, pp. 229–248.

Yin, R. (2013), "Validity and generalisation in future case study evaluations", *Evaluation*, Vol. 19 No. 3, pp. 321–332.

Yin, R. (2018), *Case study research and applications: design and methods*, SAGE.

Ylikoski, P. (2019), "Mechanism-based theorizing and generalization form case studies", *Studies in History and Philosophy of Science*, No. 78, pp. 14–22.