

Policy Effectiveness through Configurational and Mechanistic Lenses: Lessons for Concept Development

Valérie Pattyn , Priscilla Álamos-Concha , Bart Cambré , Benoît Rihoux & Benjamin Schalembier

To cite this article: Valérie Pattyn , Priscilla Álamos-Concha , Bart Cambré , Benoît Rihoux & Benjamin Schalembier (2020): Policy Effectiveness through Configurational and Mechanistic Lenses: Lessons for Concept Development, Journal of Comparative Policy Analysis: Research and Practice

To link to this article: <https://doi.org/10.1080/13876988.2020.1773263>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 26 Jun 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Policy Effectiveness through Configurational and Mechanistic Lenses: Lessons for Concept Development

VALÉRIE PATTYN ^{*}, PRISCILLA ÁLAMOS-CONCHA^{**},
BART CAMBRÉ^{**}, BENOÎT RIHOUX [†], & BENJAMIN SCHALEMBIER[‡]

^{*}Institute of Public Administration, Leiden University, Leiden, The Netherlands, ^{**}Antwerp Management School, Antwerp, Belgium, [†]University of Louvain (UCLouvain), Department of Political and Social Sciences, Université Catholique De Louvain, Louvain-la-Neuve, Belgium, [‡]Department Work and Social Economy, Flemish Government, Brussels, Belgium

ABSTRACT *The aim of this article is to build up a concept-informed research design to answer “why and how” a policy can make a difference. It demonstrates the potential and challenges of an innovative multimethod approach, which combines a configurational and mechanistic view to policy effectiveness. The article hereto draws on experiences in applying Qualitative Comparative Analysis and Process Tracing in one single evaluation. The study calls for a rigorous treatment of concepts, especially to avoid the risk of mechanistic heterogeneity. It unpacks important lessons in concept formation and operationalization, so as to ensure concept validity and to make strong causal inferences.*

Keywords: Qualitative Comparative Analysis; Process Tracing; policy evaluation; causality; policy effectiveness

1. Introduction

Shedding light on the “whys” and “hows” behind policy success or failure is quintessential for policy learning. At present, though, evaluation research that addresses such questions is not yet very common. Accordingly, there are not many effectiveness studies applying a configurational approach to causality, such as Qualitative Comparative

Valérie Pattyn is assistant professor at the Institute of Public Administration of Leiden University and partially affiliated with KU Leuven Public Governance Institute.

Priscilla Álamos-Concha is senior researcher at Antwerp Management School, and affiliated as scientific collaborator with the Center for political science and comparative politics (CESPOL) of UCLouvain.

Bart Cambré is Vice Dean of Antwerp Management School and professor of Business Research Methods at Antwerp Management School, the University of Antwerp and MCI (Austria).

Benoît Rihoux is full professor in political science at the University of Louvain (UCLouvain, Belgium), Chair of the Centre for Political Science and Comparative Politics (CESPOL) and coordinator of the global COMPASS network ([HYPERLINK “http://www.compass.org”](http://www.compass.org) www.compass.org) around QCA and related methods.

Benjamin Schalembier is a researcher at the Department of Work and Social Economy at the Flemish Government.

Correspondence Address: Valérie Pattyn, Faculty Governance and Global Affairs, Institute of Public Administration The Hague 2501 EE, The Netherlands. Email: v.e.pattyn@fgga.leidenuniv.nl

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Analysis (QCA), or that have a mechanistic understanding of causality, such as Process Tracing (PT). And to the best of our knowledge, studies in which the two questions are combined in one single evaluation (“why *and* how”) are almost non-existent. This is unfortunate, given the promises of such a multimethod design. A study relying on QCA and PT has the potential to combine the strengths of cross-case causal inference of multiple policy interventions (equifinality) and within-case causal inference through within-case analysis of individual interventions (causal mechanisms) (Schneider and Rohlfing 2013, 2013; Goertz 2017).

Applying a multimethod design is not free from challenges, however. This particularly applies to the treatment of concepts (Goertz 2005; Collier and Gerring 2009), and the risk of mechanistic heterogeneity in a QCA–PT design. Whereas existing literature on multimethod designs has dealt with concept formation (Beach and Pedersen 2016, 2019; Goertz 2017), there is only little empirical guidance available when it comes to ensuring strong causal inferences in such a design, and avoiding flawed generalizations.

Our contribution addresses this issue and is unique in this respect. We do so by drawing on our experiences in combining QCA and PT during an evaluation commissioned by the Flemish authorities in which we analyse the effectiveness of in-house training programmes (funded by the European Social Fund – ESF) in Flanders-based firms. Rather than discussing the actual findings of the study, we unpack some important lessons one should consider when developing the research design of such a multimethod study.

Our aim is twofold: first and foremost, we explain how to ensure concept validity when engaging in research that combines QCA and PT. Indeed, even if the two methods are consecutively applied (sequential design), it is imperative to anticipate the use of both methods at the early stage of developing the research design. In particular, we illustrate concept formation (conceptualization and operationalization) of conditions, contexts and outcome, so as to achieve better causal inferences when tackling “why and how” questions. Secondly, our study can be read as a call to expand the toolbox for comparative policy analysis and highlights the potential of alternative approaches to policy effectiveness research, other than the more mainstream experimental designs. While drawing on an evaluation example, the study can be inspirational beyond evaluation research.

The contribution is structured as follows: first, we present different conceptualizations of policy effectiveness, and explain what it entails to apply a configurational or mechanistic approach to policy effectiveness, and how this differs from experimental approaches. This sets the stage for the introduction of our empirical study, which serves as a basis for illustrating the merits of an evaluation in which QCA and PT are combined. A subsequent section explains how we dealt with the challenging issue of concept formation, and gives illustrations of how to avoid mechanistic heterogeneity in practice. We conclude with some implications for future evaluations.

2. Different Conceptualizations of Policy Effectiveness

What constitutes effective policy has been a key issue in the policy sciences for decades, and has received reinvigorated attention, with a wave of recent studies revolving around the “new design” orientation (see Howlett and Mukherjee 2018; Peters et al. 2018; Bali

et al. 2019). From a policy evaluation lens, policy effectiveness is mainly conceived as the study of the effectiveness of particular policy instruments or policy measures (Mukherjee and Singh Bali 2019) in bringing about changes in the actions of policy targets. When considering guidelines for policy evaluations, governments tend to focus on “whether the policy works” (Stern et al. 2012), or related notions. Strictly speaking, demonstrating that policies “work” requires testing the “attribution” of a government intervention to a particular effect. Experimental designs such as Randomized Controlled Trials (RCTs), which rely on counterfactual comparisons between situations with (“policy on”) and without the intervention (“policy off”), are particularly suited for this objective. While experimental evidence can be very insightful for policy makers, particularly for accountability purposes (Pattyn 2019), not all policy settings lend themselves to the application of RCTs. Importantly, the “policy works” claim relies on the assumption that the intervention is the primary cause of the effect of interest (Stern et al. 2012, p. 38). Experimental approaches serve to maximize confidence that the observed effect is indeed attributable to the treatment. In many complex policy settings, however, an intervention will be but a “contributory” cause of a particular effect, while the impact of the intervention will also often depend on the context in which it is embedded. The attribution–contribution distinction is therefore of major importance for the study of policy effectiveness. From such a contributory lens, one can assume that an intervention may be a necessary part of a causal package of factors that together may be sufficient to produce the intended effect (Stern et al. 2012). If indeed a policy intervention turns out to be a vital part of a causal package that is sufficient in triggering particular effects (a so-called “INUS” condition – that is, an Insufficient but Necessary part of a configuration that is in itself Unnecessary but Sufficient; Schneider and Wagemann 2012, p. 79), it can be said that the policy “makes a difference” in bringing about change in the actions of policy targets (Mukherjee and Singh Bali 2019), which is more than a mere rephrasing of the notion that the “policy works”.

Policy evaluators having the ambition to identify in more depth “why the policy makes a difference” can rely on different methods, of which QCA is very appropriate. In this approach, addressing the “why” question should be conceived as unravelling the (combinations) of conditions under which a policy makes a difference. QCA is particularly built on the idea of complex causality, and on the assumption that social phenomena have multiple and conjunctural causes (Berg-Schlosser et al. 2009; Rihoux and Ragin 2009; Fischer and Maggetti 2017; Ragin 1987). Evaluations applying QCA therefore do not focus on the average difference an intervention makes, in contrast to those that implement experimental designs. Instead, they are oriented to identify the diverse performance of the intervention in different settings (Befani 2016, p. 17), no matter whether the setting is an outlier or not. Taking context variables into account is hence inherent to QCA, which makes it especially suited for evaluations for learning purposes (Pattyn et al. 2019). That said, while an evaluation study resorting to QCA has the potential to address the conditions under which policy effectiveness was realized, it will not provide an answer to the question of “how the policy made a difference”, i.e. in terms of (causal) mechanisms. The same can be said about experimental approaches. Pawson and Tilley (1997) attributed this to the “successionist theories of causality” of experimental designs. Because of their setup, we can only know that something has changed from status “a”

(without stimulus) to status “b” (with a stimulus), but it is difficult to attribute change to a stimulus, or a policy intervention.

To answer the “how” question, there is no other way but to open the causality black box. This brings us to a third type of effectiveness designs, such as theory-based approaches to evaluations, that particularly have this ambition. A plethora of theory-based approaches exists (Stame 2004), but they all require identification of the assumptions on which a programme is based (Birckmayer and Weiss 2008, p. 408). Assumptions concern the logic model connecting programme activities to intended or observed outcomes (Pawson and Tilley 1997). In many instances, no well-developed theory exists as to how a particular policy intervention works. In these circumstances, realistic approaches to theory-based evaluation may be relevant for building or testing theory (Stern et al. 2012). Specific about realistic approaches to theory-based evaluation is the central role of generative mechanisms in understanding policy effectiveness, and how these operate in context. While in counterfactual and configurational approaches the explanation is located in respectively variables or in (combinations of) conditions, realistic evaluation focuses on what brings about the relationship between policy intervention and effects (Pawson 2008). At the basis is the idea that causal relationships only occur when triggered by a generative mechanism.

From such a viewpoint, the mere connection between the mechanism and the occurrence of the outcome is not fixed, but will be contingent on context. Generative causal explanation therefore requires the identification of the so-called CMO configuration, i.e. contexts (C) and mechanisms (M) that account for outcomes (O). A policy intervention, from this perspective, can be considered as an opportunity that actors can choose to seize, but the outcomes will depend on how the mechanisms work in context (Stame 2004; Stern et al. 2012). As White and Phillips (2012, p. 43) put it: “The context signifies the precise circumstances into which a particular intervention is introduced, and the mechanism is the precise way in which this measure works within a given context to produce a particular observable outcome. The CMO configurations for a given intervention bring together the different programme contexts with the multiple potential mechanisms which together might produce various outcomes”. Identifying a generative mechanism operating within the causal process can help to achieve some middle range theorizing about the effectiveness of interventions (Pawson and Tilley 1997, p. 124; Stame 2004).

To reconstruct the causal process between the causes and effects, policy evaluators can resort to PT as a within-case method (e.g. Schmitt and Beach 2015; Befani and Stedman-Bryce 2017). To be sure, different versions of process tracing exist. In this article, we rely on Beach and Pedersen’s (2016, 2019) approach to PT, while acknowledging the merits of other variants (Rohlfing and Schneider 2018) that apply a counterfactual understanding of causality at the level of causal mechanisms. The latter draws attention to the differences that possible variations in action linkages make for the outcome across cases (Beach and Pedersen 2019, p. 32). The approach that we have adopted in this article is more in line with the above-mentioned realistic approach. Unlike the counterfactual approach to PT, we focus on the treatment of causal mechanisms as middle-range theories which apply to a population of cases sharing the presence of particular contextual conditions. The idea is to understand and learn how the actual process works in a particular setting, with the aim to generalize to similar cases.

Table 1. A comparison of three approaches to effectiveness research

Topic	Experiments	Method	
		QCA	PT
Evaluation question	Did the policy work?	Why/under which (combinations of) conditions did the policy (not) make a difference?	How did the policy make a difference?
Policy-oriented ambition	Accountability	Policy learning: Explaining	Policy learning: Understanding
Focus	(Average) causal effects	Causal complexity	Causal processes
Mode of causal explanation	Counterfactual comparison (succession logic)	Cross-case comparison of combinations of conditions and outcomes	Mechanistic within-case inference

Source: own elaboration.

Table 1 provides a concise overview of the main attributes of the three effectiveness questions discussed, that can help evaluation stakeholders to decide which evaluation question(s) are of most interest in a particular setting. Our overview is not exhaustive, either in terms of all possible approaches to causality or in listing all methods compatible with a particular type of question (for a more extensive overview, see Stern et al. 2012). It is nonetheless indicative for the main dimensions of effectiveness research that are commonly at stake, and of the methods that one can resort to.

If technically and politically appropriate, and provided that resources are available, evaluators ideally address multiple questions. As such, one can achieve a multifaceted outlook on policy effectiveness. Combining several type of effectiveness questions, however, also implies combining different approaches to causality, to which one should pay attention to already at the stage of establishing an evaluation design. In the remainder of the article, we unpack what it entails to combine a configurational and a mechanistic approach in one single effectiveness study.

3. The Evaluation of Training Transfer Effectiveness

To illustrate the establishment of a design that combines a configurational and mechanistic understanding, we rely on our experience in an evaluation study, which focuses on the effectiveness of soft skills training (such as leadership skills and stress management) in Flemish (Belgian) firms. The evaluation of the training was commissioned by the Flemish ESF Agency, which also subsidized them.

Training programmes can work towards a multitude of effects or outcomes. In consultation with the commissioner of the evaluation, it primarily focuses on effects at individual worker's level, rather than at firm level. Supported by educational theories (Kirkpatrick 1994; Holton 1996), we examined "training transfer effectiveness", with which we refer to effects at the behavioural level. Later in the article, we detail how we conceptualized this outcome. Although previous counterfactual research demonstrated

the positive impact of training subsidies in Flanders (Baert et al. 2014), it also revealed that there is not always transfer of what is learned (in the training programme) to the working environment (Botke et al. 2018). This observation constituted the main rationale for our evaluation and triggered the commissioner to switch focus from “whether the subsidized training works” to the conditions under which training programmes make a difference in the working environment, and to the mechanisms that can make us understand how successful training works.

The evaluation was launched in 2017 and reached its final stages at the time of writing. The study entails an analysis of 10 soft skill training programmes followed by 203 participants. Data collection, for the QCA part of the study, consisted of a survey that we sent to all training respondents before and after the training. We could thus track differences in employees’ skills before and after having attended the training. For the PT part of the study, we rely mainly on interviews among relevant stakeholders engaged in the training programme, i.e. attendees themselves, but also their colleagues, responsible managers, and employees in charge of developing the in-firm training philosophy. For the QCA part, the employees that attended one of the training courses constitute the cases of our evaluation; these cases (individuals) are nested in firms that receive ESF training subsidies. In PT, when performing within-case analysis, the cases are instances of a causal process, linking causes with the outcome (Beach and Pedersen 2016, p. 5), and are the units in which a given causal relationship plays out, from the cause to the theorized outcome (Beach and Pedersen 2016, p. 5). For the actual analysis, we proceeded only with those employees who participated in different data collection rounds and for which we had no missing data. Indeed a QCA analysis cannot be conducted if information on particular conditions (or the outcome) is missing (Rihoux and De Meur 2009). This rigorous restriction inevitably resulted in the loss of some cases along the way. Eventually, 51 cases were kept in the analysis.

We opted for a sequential multimethod design in which we first applied QCA, and then PT. The surveyed literature points to a broad range of different explanations of why training is successful (or unsuccessful) in a specific situation. Unlike PT, QCA proves especially useful, when faced with such a list of potential determinants, to identify what are necessary and sufficient (combinations of) conditions (Beach 2018, p. 13) for training transfer effectiveness. QCA also helps to produce a rigorous mapping of the population of potential cases where a given causal mechanism might operate, and hints at potentially relevant contextual conditions that facilitate such functioning. This sets the basis for a further within-case analysis in a well-justified selection of successful cases (see below) to address the “how” question. Note that this is not the only available strategy, since one can also choose positive cases that are non-uniquely covered by the same combination of conditions (Beach and Pedersen 2018). Given the sensitivity of mechanisms to context, in a realistic understanding such strategy could be relevant to unravel whether the presence of other conditions matters to “activate” the mechanism linking X and Y.

A further interest of the “QCA-first sequence” is that it enables us, assuming there are no omitted causal/contextual conditions (Beach and Pedersen 2016), to cautiously generalize our findings from the ensuing PT to other, similar, cases. True, PT could also precede QCA, where insights gained during the PT case could be used as input for the QCA, either by identifying other causes or contexts that might not have been addressed in available literature, or by conceptualizing concepts in an alternative way (Beach and

Pedersen 2016, 2019). However, our position is that neither sequence is superior to the other. Which method is implemented first will to a large extent depend on available empirical evidence and theory in terms of what explains the effectiveness of particular policy interventions.

4. Investigating Policy Effectiveness through a Configurational and Mechanistic Approach

4.1 Approach to Causality

Prior to discussing how we dealt with concepts in our study, we draw attention to the causal underpinnings of such a design (Beach and Pedersen 2019, p. 1), and lay down the three main assumptions that we have taken on board in our evaluation.

First, as we hinted at above, QCA and PT share an asymmetric approach¹ where the causal role of a condition should always refer to “only one of the two qualitative states – presence or absence – in which a given condition can potentially be found” (Schneider and Wagemann 2012, p. 78) and “any solution term always refers to only one of the two qualitative states – presence or absence – in which an outcome can be found” (Schneider and Wagemann 2012, p. 78).

Second, causation is conceived in a deterministic fashion – although not by default in QCA, because it “allows for deviations from perfect sufficiency and perfect necessity” when applying more sophisticated procedures such as parameters of fit and also when working with fuzzy values of sets (Schneider and Wagemann 2012, pp. 316–317). Considering this conception of causation, when one claims that a particular condition is necessary for policy effectiveness, one is also claiming that this outcome would not occur if this given necessary condition is absent. Thus, the ontological determinism here is “claiming that a condition is necessary” for a given outcome to occur (Beach and Pedersen 2019, p. 18) and hence “that things do not happen by chance, although our empirical knowledge of why something happened will always be imperfect” (Beach and Pedersen 2019, p. 18).

Third, we combine cross-case analysis and within-case analysis (Goertz 2017). With PT, one can only make inferences that the studied causal mechanism is present or not in the case(s) of interest, but it does not indicate whether the cause(s) was (were) necessary or sufficient (Beach 2018, p. 3). To identify such causation at the cross-case level, we engage in a QCA-first study. The combination of both methods enables us to address “why” and “how” questions via certain methodological alignments. These alignments are related to the way in which we conceptualize our concepts. The main aim is to avoid the risk of causal heterogeneity at the level of mechanisms (i.e. “mechanistic heterogeneity”; Beach and Pedersen 2016, 2019). Mechanistic heterogeneity is at stake when the same outcome would be produced by different conditions, or when the same conditions yield different outcomes. The latter would imply that different causal mechanisms exist, linking conditions and outcomes. From a causal inference point of view, this is most problematic. Mechanistic heterogeneity jeopardizes making generalizations about the presence of a mechanism in a particular positive case to other positive cases in the population.

4.2 *The Treatment of Concepts*

From the above, it should be clear that the combination of QCA and PT is anchored in the notion that a deterministic ontology is compatible with the study of causal mechanisms (Mahoney 2008; Russo and Williamson 2011; Goertz and Mahoney 2012; Illari and Russo 2014) in real-world single cases with PT. This overarching assumption is key in how we established our research design, and especially how we approached the concepts in our design. Here, below, we systematically compare our approach to a situation in which we would have engaged in a QCA-only or PT-only study (i.e. a classic single-method study).

4.2.1 Distinguishing between Contexts and Causal Conditions. In QCA-only designs, conditions can in principle be “contexts” or “causes”, unless one is explicitly applying a research design aimed at understanding remote and proximate conditions (Schneider 2019). In PT-only designs, however, it is essential to make a clear-cut distinction between contexts and conditions (Beach and Pedersen 2016, 2019). Conditions only qualify as “causal” if they are active and capable of producing a certain phenomenon. Contexts, on the other hand, are conceived and conceptualized as scope conditions: they are passive in terms of productivity, but they need to be present for the correct functioning of a causal mechanism. A causal condition, in turn, is defined as an “activator” (Capano et al. 2019, p. 5), as something that triggers a mechanism, activating the causal forces in a productive relationship (Beach and Pedersen 2016) “through which the behaviour of individuals, groups and subsystems is altered to achieve a specific outcome” (Capano et al. 2019, p. 5). In other words, a causal condition “does something”, whereas the contextual condition is the enabler, a factor that determines whether a causal relationship functions as theorized (Beach and Pedersen 2016).

Therefore, combining QCA and PT involves the application of some methodological alignment where one method is to be given priority over the other. In concrete terms, given the specific understandings of contexts and conditions in PT, one also needs to apply such a clear-cut distinction in the QCA part of the analysis, in order not to jeopardize the implementation of the PT part of the study. Likewise, conditions in a multimethod design need to be formulated in a causally relevant manner in which they are distinguished from contexts.

In our evaluation, for instance, we treated the “supervisor support” condition as a causal condition of training transfer effectiveness, consistent with insights from the education research literature. Commonly, supervisor support is understood in terms of “sources of encouragement, assistance, reinforcement, opportunities and guidance (feedback) for employees on their use of new knowledge at the workplace” (Lancaster et al. 2013). In the same vein, and consistent with the Beach and Pedersen guidelines (2016), we defined the concept in a causally active way: supervisor support is the superior’s commitment to facilitate the retention and motivate the use of the acquired content in a training to the job by employees, during and after a training program takes place (Lancaster et al., 2013, Nijman et al., 2006).

By contrast, we treat the “training designed as active learning instructional method” condition more as a context factor, that would need to be present to make the given causal mechanism work correctly (i.e. an “enabler”, as discussed above). We hereto rely again on the training transfer literature, where “training as active learning” is merely seen

as a building block or stimulus for high-impact learning (Dochy and Segers 2018). The idea resonates with Dewey's (1997) "learning by doing" approach, who argued that the mere passive consumption of knowledge, skills and attitudes is an ineffective learning method. From such a perspective, instructional learning methods are seen as facilitators for training transfer to take place, which will make the causal mechanism function, so that it leads to training transfer effectiveness.

4.2.2 Selecting and Conceptualizing the Outcome and Conditions. Following multimethod research good practice, it is a prerequisite that conditions and the outcome are formulated in a way that makes them applicable to all the methods utilized. For our design specifically, it is hence important to ensure that the conclusions drawn from the QCA are informative for the subsequent PT analysis. As mentioned, both in QCA-only and PT-only designs, causal claims are asymmetric claims, in the sense that they only concern the causes of the outcome, and not what causes the absence of the outcome. That being said, there are major differences in how concepts are defined in these respective methods. In QCA-only designs, concepts can be of all kinds, i.e. there are no restrictions about how we may define them. Scholars may nonetheless have particular preferences, reflecting different schools of conceptualization and measurement, such as philosophy, economics, and latent variable (cf. Goertz *forthcoming*). In PT, the way in which concepts are formulated is much more restrictive: the prominent view of concepts is based on ontological attributes, i.e. following an essentialist position about the constituent parts of a given phenomenon, "where the goal of our definitions is to capture the essence of what the concept means as a cause or outcome (Sartori 1984) instead of thinking of conceptualization in terms of choosing indicators of latent variables" (Beach and Pedersen 2019, pp. 56–57).

Given these distinctions, so that QCA and PT can operate within a multimethod design, concepts need to be defined in a set-theoretic way (Beach and Rohlfing 2016; Rohlfing and Schneider 2018). Set theory as used in social science methodology defines causes and outcomes in terms of the attributes that determine whether a given case is a member of the set of the concept, and conceives theoretical relationships between causes and outcomes as subset relationships (e.g. a necessary condition is one in which the cases that are members of the outcome are a subset of cases that are members of the necessary condition) (Ragin, 1987; Ragin 2000; Schneider and Wagemann 2012). Such an approach has deep implications for both the selection and the conceptualization of the outcome and of the conditions.

4.2.2.1 Selecting and Conceptualizing the Outcome. In QCA-only studies, an outcome represents a given phenomenon of interest. QCA requires an outcome to be clearly defined, circumscribed and conceptualized in a set-theoretic manner, in its positive and negative poles (Beach and Pedersen 2019). When the outcome has been treated as variable (for instance, when using Likert scales), one can systematically compare differences in outcome values, enabling one to identify "the impact that differences in values of independent variables have across the full range of values of the dependent variable across a set of cases" (Beach and Pedersen 2019, p. 61).

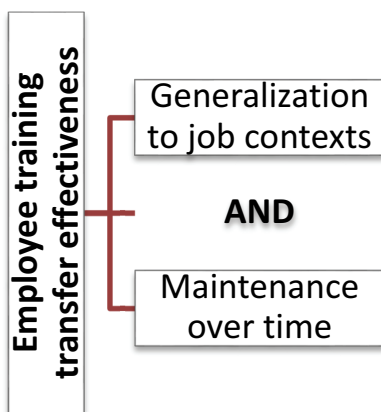
By contrast, as mentioned above, in PT-only designs one may define the outcome in terms of attributes that have the "capacity" to receive some impact from the respective causes and causal mechanisms (Goertz 2017; Capano et al. 2019). When implementing

our QCA–PT multimethod design, we thus have to prioritize the PT ontological view on concepts; this is the only strategy that enables us to feed the insights of the QCA analysis into the PT analysis. Outcome concepts are therefore to be conceived as attributes that are essential to, or a constituent part of, a phenomenon. To know which attributes to consider, it is wise to think along the lines of questions such as “what does the phenomenon mean?” – e.g. what is policy effectiveness? What is it not? And how can we recognize the existence of policy effectiveness in a given case?” In other words, the outcome is to be unpacked in terms of attributes that constitute the outcome itself. Outcomes need to be defined as something that can be affected by certain causal attributes.

By way of illustration, in our evaluation our outcome is *employee training transfer effectiveness at the workplace*. Following an attribute-based understanding, we defined it as “the application of the learned knowledge (content, skills, attitudes) acquired in a training programme to the job by trainees AND maintained over a period of time” (see Figure 1).

For the presence of training transfer effectiveness, i.e. the positive pole of the concept, we assumed that the two attributes of the concept need to be present together to make it exist (positive outcome). We hereto relied on definitions mapped from the “training transfer” literature which most commonly conceives it as “the application of what is learned from the training to the workplace” (Dochy and Segers 2018, p. 163). This use of new knowledge into the job is also referred to in the literature as “generalization”, meaning that the trainees are able to “activate the resources” acquired in one context (e.g. training) in another context (e.g. the job) (Kirwan and Birchall 2006), or as “productive use of acquired knowledge and skill” (Gegenfurtner 2011, p. 154). Similarly, some scholars introduce the notion of “maintenance of the learned material over a period of time on-the-job” (Kirwan and Birchall 2006, p. 5) when they refer to an

Figure 1. Conceptualization as attributes of the “Employee training transfer effectiveness” outcome



Source: Own elaboration.

effective transfer, arguing that the continued application can lead to a certain standard over time (Broad and Newstrom 1992, p. 6).

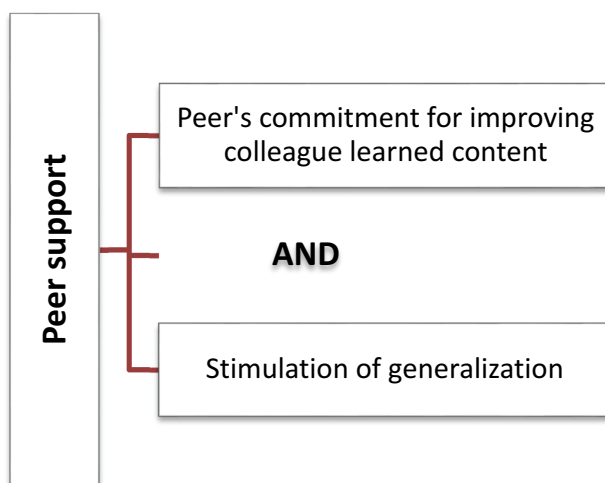
We further assumed that the absence of one of these attributes implies the absence of the concept, and hence a failure of “effective employee training transfer” (negative outcome). In principle, in QCA it is important to consider both the positive and the negative pole of concepts (Goertz 2005; Yamasaki and Rihoux 2009; Coppedge 2012). In PT, by contrast, the study of causal mechanisms is primarily geared towards the “positive pole”, in line with the asymmetric nature of causal claims about mechanisms. Causal powers only concern the positive pole of a concept (Beach and Pedersen 2016, 2019). To do justice to both approaches’ distinguishing traits, we conceived a negative pole for our concepts. Concepts qualify as “absent” if at least one of the attributes is missing.

Further, in the training transfer literature, concepts tend to be understood in a conjunctural way (see below). Having this negative pole defined is therefore essential for the QCA part of the study. Yet in the search for relevant conditions, and with PT only oriented towards the positive pole, we primarily put efforts into identifying conditions for an effective employee training transfer (the positive outcome), and in identifying the causal mechanisms linking the conditions and the positive pole of the outcome, within certain necessary contexts. Note that, in a QCA application, it is also possible to lay more emphasis on the positive outcome, even if it is still a good practice to perform QCA analyses (minimizations) both for the positive and for the negative outcomes – thus: the QCA and PT rationales are not fully opposed in this respect.

4.2.2.2 Selecting and Conceptualizing (Causal) Conditions. In QCA-only designs, the selection of conditions is geared towards explaining the outcome of interest. At the same time, the conceptual approach to conditions is much more flexible in QCA than in PT. In the latter, conditions must be understood as “activators” (Capano et al. 2019) that have causal powers to trigger processes that produce the outcome. If the selected condition cannot trigger a process, it is probably a contextual condition that could play a relevant role in the well-functioning of the mechanism, but not as a causal condition.

Similar to the conceptualization for outcomes, we thus approach the conceptualization of conditions in a process-oriented manner. Our primary focus is placed on ontological attributes that have causal powers to trigger causal mechanisms, which can in turn produce the outcome of interest. Besides this, we treat concepts “thickly”, so as to create a causally homogeneous population which is defined by contextually specific concepts and which is compatible with mechanistic claims. Again, for the study of causal processes, we solely focus on the positive pole of conditions, so as to enable causal inferences via the PT analysis.

In our evaluation, for instance, “peer support” is one of the key causal conditions, which we define as “colleagues’ commitment for employees to improve the trainee’s learned content and stimulate the trainee’s use of learned material to the job”. Based on a literature review, we identified the following main attributes of peer support: (1) peer’s commitment for employees to improve the trainee’s learned content in the job, (2) stimulation of the generalization of learned knowledge by trainees at the workplace (Nijman et al. 2006). For the negative pole of the concept, we posited that peer support does not exist if any one of the necessary attributes is absent (see Figure 2).

Figure 2. Conceptualization as attributes of the “Peer support” condition

Source: Own elaboration.

Both attributes are connected via the “logical AND”, to avoid sources of mechanistic heterogeneity that could otherwise lead to potentially flawed generalizations. As explained, mechanistic heterogeneity means that a given (combination of) condition(s) is linked to the same outcome in different cases through different mechanisms (cf. Schneider and Rohlfing 2016, p. 555; Beach and Pedersen 2019). Thus, at the level of concept formation (either at the abstract level or at the level of technical macro-concepts with QCA) we avoided multi-attribute concepts. The latter are usually formed with logical OR (i.e. either substitutable or family resemblances of concepts – see Goertz 2005), but this entails the risk that attributes that are equivalent from a cause–effects perspective at the cross-case level trigger different mechanisms at within-case level (Beach and Pedersen 2019, p. 120).

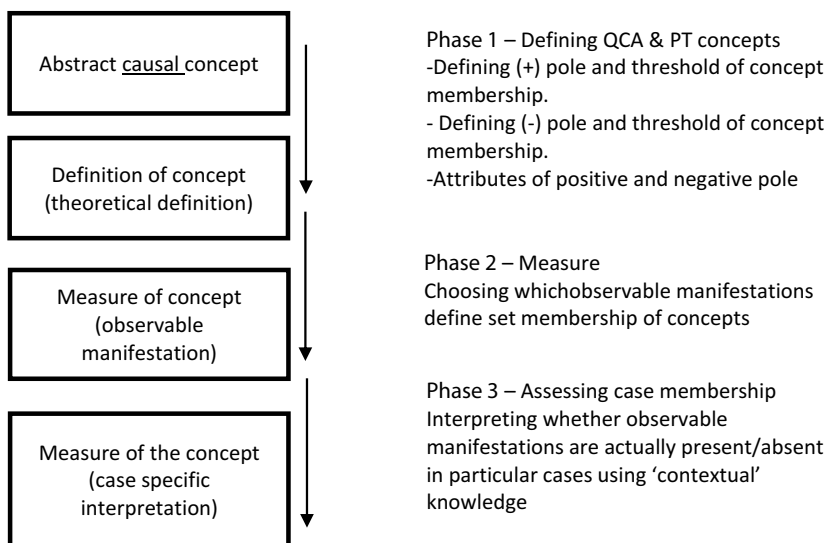
To explain further the importance of mechanistic homogeneity, let us consider the process that links peer support with the training transfer effectiveness. The process is conceived as consisting of 14 parts (see Online Appendix 1) structured around six blocks: (1) following the training; (2) building up common understanding; (3) intervention; (4) adaptability and application; (5) intervention after adaptability (feedback loop) and (6) new work thinking. Imagine that we would have proceeded with a conceptualization of peer support in which attributes are connected with a logical OR, in which “peer’s commitment for improving colleague learned content” (AT1) and “stimulation of generalization” (AT2) are conceived as functional equivalents. Cases could then have a positive score on the overall concept of peer support, without necessarily sharing the same attributes. Importantly, this would have entailed the risk that for cases scoring positive solely on AT1, the process leading to training transfer looks different from those cases with membership in AT2 solely. For example, while the blocks “following the training” and “sharing common understanding” (parts 1–4 in the Online Appendix) can

be found relevant for AT1, AT2 could occur through the other blocks (parts 5–14). Thus, positive cases that are members of the whole condition but with different membership in the respective attributes should not be treated as equivalent with regard to the processes within which they are engaged, precisely because elements of multi-attribute concepts can have different effects at the level of mechanisms.

4.2.3 Operationalization. In QCA-only designs, one can operationalize concepts based on validated scales from previous theoretical and empirical studies. However, in PT-only studies, concepts are built up from ontological attributes, which implies that the operationalization should follow the same analytical purpose. In line with our ontological understanding of concepts, we operationalize them in a common PT manner. We hereto rely on the heuristics of so-called “observable manifestations” as coined by Beach and Pedersen (2016, 2019) for operationalizing and measuring causal concepts and outcomes. Observable manifestations can be understood as empirical fingerprints left by the attributes in real-world cases, and they focus on what a given attribute looks like in real-world cases. When a concept consists of multiple attributes, multiple such observable manifestations need to be envisaged and measured. If one attribute corresponds to multiple observations, one needs to spell out the relationship between these observations, which can be conjunctions, substitutes or display so-called family resemblance.

Figure 3 details how to proceed in a stepwise manner from the definition of an abstract causal concept to its concrete tangible measurement. In the first phase, we must define the positive and negative pole, because we are engaged in a comparative QCA-first study. The way in which we unpack concepts here is based on ontological attributes (Goertz 2005; Beach and Pedersen 2016, 2019). Following this step, we may operationalize our

Figure 3. Conceptualizing causal concepts for the combination of QCA and PT



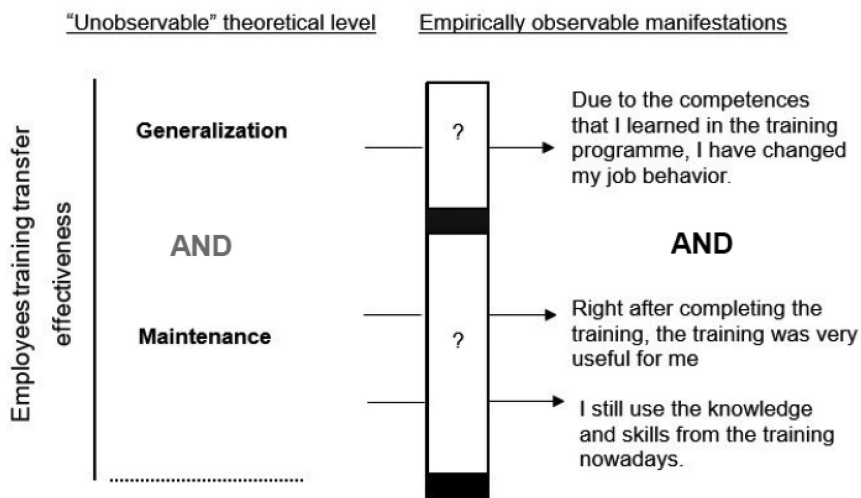
Source: Based on Beach and Pedersen 2016, p. 96 (inspired by Adcock and Collier 2001, p. 531).

concepts by considering the observable manifestations (i.e. fingerprints) and the relationship between such fingerprints (substitutable, conjuncture, family resemblance; see further Goertz 2005). We may define the set membership of each attribute in relation to the whole concept (Do all attributes need to be present to form the concept under study?). Finally, using empirical data, we may evaluate if the respective fingerprint is present or not in a given case, by considering the importance of the context, for its interpretation.

By way of illustration, in Figure 4, we detail how we operationalized “training transfer effectiveness”. As can be derived from this figure, for training transfer effectiveness, each attribute of the outcome needs to be present (logical “AND”). This also applies to the level of the individual attributes, where the two observable manifestations of “maintenance over time” are also in a conjunctural relationship. On top of this, we assumed that all three observable manifestations of training transfer effectiveness as a whole are “necessary” for that effectiveness. In other words, an employee for which just one of the attributes could not be observed is considered as a case in which training transfer effectiveness is absent.

A combination of QCA and PT thus comes with a very stringent view on case membership in conditions and in the outcome, especially if one wants to avoid sources of mechanistic heterogeneity. In spite of these challenges, this multimethod design offers major potential. To illustrate, the QCA analysis (see Online Appendix 4 for the conservative solution) revealed for instance that two cases (coded “Dec_coa_ld” and “Mat_sm”) share the same solution term in the same contextual circumstances (see Online Appendix 5 for potential cases with PT). Provided that one collects enough evidence on the presence of the workings of each part of the process (Beach and Pedersen 2016, 2019), tracing the causal mechanisms in these two cases enables us to

Figure 4. Observable manifestation of policy effectiveness (training transfer effectiveness)



Source: Own elaboration (figure style based on Beach and Pedersen 2016).

make strong causal inferences. Besides, the design enables us to gain in external validity, due to the contextual sensitivity of mechanism-based explanations (for more information see Online Appendices 2 and 3).

5. Conclusion

Searching for effective solutions to public policy problems is at the core of policy design (Peters et al. 2018). From a policy evaluation angle, three types of ex post effectiveness questions are usually distinguished, each highlighting a different dimension of the functioning of policy interventions: “did the policy work”, “why”, and “how”? When taking policy effectiveness seriously, public authorities ideally engage in multimethod research, in which at least two questions are addressed, thereby providing a multifaceted outlook on the impact of public policy.

It is however not self-evident to combine several questions in one single study, in spite of the strong potential of such a strategy. In this article, we unpacked what it entails to combine a “why” and a “how” outlook on policy effectiveness, in such a way as to avoid the risk of mechanistic heterogeneity. Our study especially calls for a rigorous treatment of concepts, and for bringing this into the focus of any multimethod design. After all, conceptual misalignment between two different approaches on causality may jeopardize concept validity, and as such also hinder the external validity of the evaluation findings. This especially applies to a study in which QCA and PT are combined.

Taken altogether, it can be said that QCA has a more relaxed view on causality and concepts than PT. Nonetheless, to build and implement a robust multimethod design sequencing QCA and PT, one should adopt PT’s processual and ontological understanding of causes, also in the QCA stage of the research. This first requires making a clear-cut distinction between context and causal conditions in a way that is PT compatible. The same applies to the need to define and unpack concepts in terms of ontological attributes, at least when it comes to the calibration of conditions and outcomes in QCA. Third, the relationship between ontological attributes is to be defined in set-theoretic terms, where multi-attribute concepts are ideally avoided, so as to prevent mechanistic heterogeneity. Fourth, the operationalization of concepts should be conducted along the same ontological lines: consistent with PT, concepts should be operationalized in terms of “empirical fingerprints” left by the attributes in real-life cases. Engaging in multimethod research with QCA and PT is hence not a walk in the park. Yet the potential to combine the strengths of cross-case causal inference of multiple policy interventions and within-case causal inference through within-case analysis of individual interventions strongly makes up for this. Further, applying such design does full justice to the contributory nature of policy interventions to societal effects, which cannot be fully accounted for with mainstream “what works” designs, and the corresponding experimental methods. For policy evaluations to take the important step towards combining how and why questions into one framework, the proposed design, however challenging, can become one of the most promising designs for future research.

Notes

1. This is currently subject to debate, as some argue that QCA rather relies on a causal notion of difference-making, i.e. following a regularity approach (Baumgartner and Falk 2019), versus the asymmetric approach based on counterfactual thinking. It seems to us that both views can be convincingly upheld, but we have opted for what constitutes up to now the most widely accepted view, following Rohlfing and Schneider's (2018) argument.

Supplemental Data

Supplemental data for this article can be accessed [here](#).

References

- Adcock, R. and Collier, D., 2001, Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* **95** (3), pp. 529–546. doi:10.1017/S0003055401003100.
- Baert, L., Decramer, S., and Reynaerts, J., 2014, *KMO Portefeuille - Pijler Opleiding. Een Evaluatie Van De Opleidingsubsidies in Vlaanderen. Beleidsrapport STORE-B-14-003* (Leuven: Steunpunt Ondernemen & Regionale Economie).
- Bali, A. S., Capano, G., and Ramesh, M., 2019, Anticipating and designing for policy effectiveness. *Policy and Society*, **38** (1), pp. 1–13. doi:10.1080/14494035.2019.1579502.
- Baumgartner, M. and Falk, C., 2019, *Boolean Difference-Making: A Modern Regularity Theory of Causation*. The British Journal for the Philosophy of Science. doi: 10.1093/bjps/axz047.
- Beach, D., 2018, Achieving methodological alignment when combining QCA and process tracing in practice. *Sociological Methods & Research*, **47**(1), pp. 64–99. doi:10.1177/0049124117701475.
- Beach, D. and Pedersen, R., 2018, Selecting appropriate cases when tracing causal mechanisms. *Sociological Methods and Research*, **47**(4). 837–871 doi:10.1177/0049124115622510.
- Beach, D. and Pedersen, R. B. (Eds.), 2016, *Causal Case Study Methods. Foundations and Guidelines for Comparing, Matching, and Tracing*(Ann Harbour: University of Michigan Press).
- Beach, D. and Pedersen, R. B. (Eds.), 2019, *Process Tracing Methods: Foundations and Guidelines* (University of Michigan Press).
- Beach, D. and Rohlfing, I., 2016, Integrating cross-case analyses and process tracing in set theoretic research: strategies and parameters of debate. *Sociological Methods and Research*, **47**(1), pp. 3–36. doi:10.1177/0049124115613780.
- Befani, B., 2016, Pathways to change: Evaluating development interventions with QCA, Rapport till Expertgruppen för biståndsanalys [Report for the Expert Group for Aid Studies-EBA], Expert Group for Aid Studies. Report 05/16. Stockholm, Sweden: EBA. Retrieved from <http://eba.se/en/pathways-to-change-evaluating-development-interventions-with-qualitative-comparative-analysis-qca/#sthash.nyGxIej9.dpbs>.
- Befani, B. and Stedman-Bryce, G. 2017, Process tracing and Bayesian updating for impact evaluation. *Evaluation*, **23**(1), pp. 42–60. doi:10.1177/1356389016654584.
- Berg-Schlosser, D., De Meur, R. B., and Ragin, C. C., 2009, Qualitative Comparative Analysis (QCA) as an approach, in: B. Rihoux and C. Ragin (Eds) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks and London: Sage), pp. 1–18.
- Birkmayer, J. D. and Weiss, C. H., 2008, Theory-based evaluation in practice: What do we learn? *Evaluation Review*, **24**(4), pp. 407–431. doi:10.1177/0193841X0002400404.
- Botke, J. A., Jansen, P. G., Khapova, S. N., and Tims, M., 2018, Work factors influencing the transfer stages of soft skills training: A literature review. *Educational Research Review*, **24**, pp. 130–147. doi:10.1016/j.edurev.2018.04.001.
- Broad, M. L. N. and Newstrom, J., 1992, *Transfer of Training: Action-Packed Strategies to Ensure High Payoff from Training Investments* (Reading (Mass.): Addison-Wesley).
- Capano, G., Howlett, M., and Ramesh, M., 2019, Disentangling the mechanistic chain for better policy design, in: G. Capano, et al. (Ed.) *Making Policies Work: First- and Second-Order Mechanisms in Policy Design* (Cheltenham: Edward Elgar Publishing), pp. 2–13.

- Collier, D. and Gerring, J. (Eds), 2009, *Concepts and Method in Social Science: The Tradition of Giovanni Sartori* (London: Routledge).
- Coppedge, M., 2012, *Democratization and Research Methods* (Cambridge: Cambridge University Press).
- Dewey, J., 1997, *Experience and Education* (New York: Touchstone).
- Dochy, F. and Segers, M., 2018, *Creating Impact through Future Learning the High Impact Learning that Lasts (HILL) Model*, 1st ed. (Abingdon, Oxon /New York: Routledge/Taylor & Francis Group).
- Fischer, M. and Maggetti, M., 2017, Qualitative comparative analysis and the study of policy processes. *Journal of Comparative Policy Analysis: Research and Practice*, **19**(4), pp.345–361. doi:10.1080/13876988.2016.1149281.
- Gegenfurtner, A., 2011, Motivation and transfer in professional training: A meta-analysis of the moderating effects of knowledge type, instruction, and assessment conditions. *Educational Research Review*, **6**(3), pp. 153–168. doi:10.1016/j.edurev.2011.04.001.
- Goertz, G., 2005, *Social Science Concepts: A User's Guide* (Princeton: Princeton University Press).
- Goertz, G., 2017, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach* (Princeton: Princeton University Press).
- Goertz, G. (Ed.), forthcoming, *Social Science Concepts: A User's Guide*, 2nd ed. (Princeton: Princeton University Press).
- Goertz, G. and Mahoney, J., 2012, *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences* (Princeton: Princeton University Press).
- Holton, E. F., 1996, The flawed four-level evaluation model. *Human Resource Development Quarterly*, **7**(1), pp. 5–21. doi:10.1002/hrdq.3920070103.
- Howlett, M. and Mukherjee, I., 2018, *Routledge Handbook of Policy Design* (New York: Routledge).
- Illari, P. and Russo, F., 2014, *Causality: Philosophical Theory Meets Scientific Practice* (Oxford: Oxford University Press).
- Kirkpatrick, D., 1994, *Evaluating Training Programmes. The Four levels* (San Francisco: Berrett-Koehler Publishers).
- Kirwan, C. and Birchall, D., 2006, Transfer of learning from management development programmes: Testing the Holton model. *International Journal of Training and Development*, **10**(4), pp. 252–268 doi:10.1111/j.1468-2419.2006.00259.x.
- Lancaster, S., Di Milia, L., and Cameron, R., 2013, Supervisor behaviours that facilitate training transfer. *Journal of Workplace Learning*, **25**(1), pp. 6–22. doi:10.1108/13665621311288458.
- Mahoney, J., 2008, Toward a Unified Theory of Causality. *Comparative Political Studies*, **41** (4–5), pp. 412–436. doi:10.1177/0010414007313115.
- Mukherjee, I. and Singh Bali, A., 2019, Policy effectiveness and capacity: Two sides of the design coin. *Policy Design and Practice*, **2**(2), ppp. 103–114. doi:10.1080/25741292.2019.1632616.
- Nijman, D. J. J. M., Nijhof, W. J., Wognum, A. A. M., and Veldkamp, B. P., 2006, Exploring differential effects of supervisor support on transfer of training. *Journal of European Industrial Training*, **30**(7), pp. 529–549 doi:10.1108/03090590610704394.
- Pattyn, V., 2019, Towards appropriate impact evaluation methods. *The European Journal of Development Research*, **31**(2), pp.174–179. doi:10.1057/s41287-019-00202-w.
- Pattyn, V., Molenveld, A., and Befani, B., 2019, Qualitative comparative analysis as an evaluation tool: lessons from an application in development cooperation. *American Journal of Evaluation*, **40**(1), pp.55–74. doi:10.1177/1098214017710502.
- Pawson, R. 2008. Causality for beginners. ESRC/NCRM research methods festival. Available at <http://eprints.ncrm.ac.uk/245/> (accessed 11 March 2020).
- Pawson, R. and Tilley, N., 1997, *Realistic Evaluation* (London: Sage).
- Peters, B. G., Capano, G., Howlett, M., Mukherjee, I., Chou, M.-H., and Ravinet, P., 2018, *Designing for Policy Effectiveness: Defining and Understanding a Concept* (Cambridge: Cambridge University Press).
- Ragin, C. C., 1987, *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies* (Berkeley/Los Angeles/London: University of California Press).
- Ragin, C. C., 2000, *Fuzzy-Set Social science* (Chicago: Chicago University Press).
- Rihoux, B. and De Meur, G., 2009, Crisp-set qualitative comparative analysis (csQCA), in: B. Rihoux and C. Ragin (Eds) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks and London: Sage).
- Rihoux, B. and Ragin, C. C. (eds.), 2009, *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks and London: Sage).

- Rohlfing, I. and Schneider, C. Q., 2018. A unifying framework for causal analysis in set-theoretic multimethod research. *Sociological Methods and Research*, **47** (1). 37–63 doi:10.1177/0049124115626170.
- Russo, F. and Williamson, J., 2011, Generic versus single-case causality: the case of autopsy. *European Journal of the Philosophy of Science*, **1** (1), pp. 47–69. doi:10.1007/s13194-010-0012-4.
- Sartori, G., 1984, Guidelines for concept analysis, in: G. Sartori (Ed.) *Social Science Concepts: A Systematic Analysis* (Beverly Hills, CA: Sage), pp. 15–85.
- Schmitt, J. and Beach, D., 2015, The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation*, **21**(4), pp. 429–447. doi:10.1177/1356389015607739.
- Schneider, C. Q. and Rohlfing, I., 2013, Combining QCA and process tracing in set-theoretical multi-method research. *Sociological Methods and Research*, **42**(4), pp. 559–597. doi:10.1177/0049124113481341.
- Schneider, C. Q. and Rohlfing, I., 2016, Case studies nested in fuzzy-set QCA on sufficiency: formalizing case selection and causal inference. *Sociological Methods and Research*, **45**(3), pp. 526–568. doi:10.1177/0049124114532446.
- Schneider, C. Q. and Wagemann, C., 2012, *Set-Theoretic Methods for the Social Sciences A Guide to Qualitative Comparative Analysis* (Cambridge: Cambridge University Press).
- Schneider, C.Q., 2019, Two-step QCA revisited: the necessity of context conditions. *Qual Quant*, 53, pp.1109–1126. doi:10.1007/s11135-018-0805-7. 3 doi:10.1007/s11135-018-0805-7
- Stame, N., 2004, Theory-based evaluation and types of complexity. *Evaluation*, **10**(1), pp. 58–76. doi:10.1177/1356389004043135.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and Befani, B., 2012, *Broadening the Range of Designs and Methods for Impact Evaluations* (London: DFID). Department for International Development (DFID) Working Paper 38.
- White, H. and Phillips, D., 2012, *Addressing Attribution of Cause and Effect in Small N Impact Evaluations: Towards an Integrated Framework* (New Delhi: International Initiative for Impact Evaluation).
- Yamasaki, S. and Rihoux, B., 2009, A commented review of applications, in: B. Rihoux and C. Ragin (Eds) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks and London: Sage), pp. 123–146.