doi: 10.15678/ZP.2021.58.4.02

Zarządzanie Publiczne
Public Governance
No. 4(58)/2021
ISSN 1898-3529

Paulina Kubera

Mechanism-Based Explanations in the Impact Evaluation of Public Interventions: Digging Deeper into Why and How Programmes Worked or Failed to Work

Abstract

Objectives: The aim of the article is to consider the evaluation of public interventions through the prism of evidence-based policy (EBP) as well as, more specifically, its potential to address the problem of how to produce in the process of impact evaluation usable knowledge that can help improve policymaking and policy implementation which can be accumulated over time, where evaluations will not be single endeavours and one-off studies, but will contribute to the growing body of knowledge.

Research Design & Methods: The article provides a critical overview of the research literature on evaluation approaches and methods as tools for gathering and apprising evidence relevant for policymaking and policy implementation.

Findings: Building upon the identified limitations of the traditional input/output approach to impact evaluation of public interventions, alternative approaches to evaluation are considered that make use of a theory that properly explicates the causal mechanisms linking programme activities with programme outcomes, confronting the mechanisms with empirical observations. As a strategy for synthesising the gained knowledge, the realist synthesis is considered as being more appropriate for reviewing research on complex social interventions (rather than traditional meta-analysis).

Implications / Recommendations: The article demonstrates how theory-based evaluation with mechanistic explanation and realist synthesis can contribute to growing evidence for policy needs, identifying also their limitations and practical problems related to their implementation.

Contribution / Value Added: The article contributes to the existing pool of knowledge by providing important insights into how to use mechanism-based explanations in impact evaluation to make stronger causal claims and enhance policy-learning.

Keywords: evidence-based policy, evaluation, theory of change, causal mechanisms, realist synthesis

Article classification: conceptual article

JEL classification: H11, H43, H83, L38

Paulina Kubera – Faculty of Engineering Management, Poznań University of Technology; ul. J. Rychlewskiego 2, 60-965 Poznań; e-mail: Paulina.Kubera@put.poznan.pl; ORCID: 0000-0002-6246-6952.

Introduction

The primary question of impact evaluation of public interventions – regardless of whether these are projects, programmes or policies – is 'what works'. However, finding 'what works and what does not work' rarely does the trick and is not as simple as identifying best practices and disseminating them widely. In many instances, interventions that turned out to be successful in one setting did not bring the expected results when implemented elsewhere. The principal reason for it 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. Therefore, there is a strong need to better recognise the contextual factors that influence the intervention's success or failure. 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 begs the fundamental question: how to build the evidence base for policymaking in complex environments?

The overarching aim of the article is to consider evaluation of public interventions through the prism of evidence-based policy (EBP) and, more specifically, its potential to address the problem of how to produce in the process of impact evaluation usable knowledge that can help improve policymaking and policy implementation which can be accumulated over time, where evaluations will not be single endeavours and one-off studies, but will contribute to the growing body of knowledge. To this end, the following lines of inquiry are pursued that are discussed in the subsequent parts of the paper. First, in order to take a broader perspective, the question is raised why it is so hard, in general, to implement evidence-based practice in public policy and management, and what are the limitations of the traditional input/output approach to impact evaluation that has little to offer to inform policymaking about the likely transferability of findings from one evaluation research to other settings. Secondly, based on a critical overview of the research literature on evaluation approaches and methods as tools for gathering and apprising evidence relevant for policymaking and policy implementation, theoretical inquiry is made into the potential contribution of the theory-oriented evaluation approach that uses mechanistic explanation to answer the question why and how the intervention worked or failed to work. To this end, various definitions of the term 'mechanism' are discussed in order to clarify the conceptual confusion that surrounds the term. As they are grounded in different types of causation and analysed using different research methods, they provide different answers and serve different policy needs. Thirdly, unpacking causal mechanisms, exploring what works and what does not work in a certain evaluation endeavour is one thing. Another one is to add to a growing body of knowledge. Therefore, a realist synthesis is discussed that is more suitable for reviewing research evidence on complex social interventions than traditional meta-analysis. Finally, as suggested solutions are hardly ever perfect in reality, the article ends with indicating the drawbacks and practical problems related to the application of the mechanism-based approaches and realist synthesis that should be accounted for by policymakers, as well as promising directions for future research are indicated.

Opening the 'black box' of public interventions

The movement of evidence-based practice, i.e. the idea that professional practice should be informed by scientific inquiry, has been first institutionalised in medicine, and with success (Rousseau, 2006). However, its realisation differs in areas such as public policy and management, for a couple of reasons. First, this is due to the +distinctive features of social science that form evidence base for policymaking and managerial

decisions. Contrary to natural sciences that deal with the physical world, social sciences deal with interactions and human behaviours, which makes it hard to copy ready-to-use solutions. In medicine, which is a combination of natural science and human science (Dreyfus, 2011), one can expect with a reasonable degree of certainty that a particular medicine or treatment will work against a particular disease no matter where the medicine is administered or where treatment is given. In the case of public interventions, the same programme can produce different results when implemented elsewhere. It suffices that the programme is presented to the beneficiaries in a different way, the cultural context is different, the programme is implemented by a less experienced administration, etc. (Józefowski, 2012). Nevertheless, it should be added that also in evidence-based medicine, it is acknowledged that using scientific evidence while making clinical decisions should not be automatic nor unreflective. It should be combined with clinical experience that doctors had acquired during years of clinical practice as well as patient values and preferences. Following these indications, Rousseau (2006) defines evidence-based management as "a paradigm for making decisions that integrate the best available evidence with decision makers expertise and client/customer preferences to guide practice toward more desirable results" (p. 258). (With regard to evidence-based policy, one could say 'beneficiaries' or 'recipients' of public policies' instead of 'clients' or 'customers', although the term 'recipients' has been criticised for having passive connotations; see European Commission, 2017).

Another point to note about why it is harder to make evidence-based practice a reality in the public policy and management is the way in which decisions are made. In medicine, this is principally a doctor that individually makes decisions and prescribes the treatment to a particular patient. In turn, political decision-making and managerial decision-making in the public sector as well as in the private sector can be characterised by bargaining, entrenched commitments, and the interplay of diverse stakeholder values and interests (Head, 2010).

All of those factors that influence policymaking should be taken into account in the realisation of the evidence-based policy postulates. However, the main problem around which this article is structured is of methodological nature, namely – how to produce in the process of impact evaluation usable knowledge that can help improve policymaking and policy implementation which can be accumulated over time?

Looking at the issue of the transferability of findings from one evaluation study to other settings, it can be argued that the main problem lies in the fact that public interventions are primarily viewed in terms of their effects, with not enough attention paid to how those effects are brought about. The traditional approach to impact evaluation is focused on the input and output side of an intervention, i.e. how much was invested in a given programme and how much was achieved owing to its implementation. It entails a comparison of the situation before and after the intervention, and calculating its average effect. Certainly, making such cause-and-effect connections is at the heart of impact evaluation, as well as evidence-based practice. Nevertheless, causation is not the same as explanation, and such an approach has little to offer to inform policymaking about whether the programme can be successfully scaled up, implemented elsewhere or towards other entities, or what should be done when a programme does not yield expected effects. Frequently, diverse and contradictory findings regarding the impact of public interventions (in particular in reference to the effectiveness of international aid programmes) were not taking account of diversity of social and institutional contexts, and failed to distinguish between aid channels, instruments and modalities, as well as neglected the technological aspect and capacity-building benefits of aid (Picciotto, 2012). For that reason, randomised control trials (RCTs) – which are so successfully applied in medicine and once labelled as a golden standard also in relation

to social policy programmes[1] for eliminating selection bias[2] and addressing the attribution question in scientific and rigours fashion – had to give finally its way to the mixed methods doctrine and methodological triangulation that put quantitative and qualitative approaches on the same footing (Picciotto, 2012; Saunders, 2011). Ultimately, producing scientific knowledge does not entail merely the verification of whether one event follows another, but, rather, the explanation of the relationship between events by use of a theory (Chaney, 2016). Hence, the same expectation is placed on evaluation as a device to build evidence base for policymaking.

The limitations of the traditional approach to impact evaluation can be illustrated on the example of aid for research, development, and innovation (RDI). The input/output approach rests on three assumptions that do not necessarily hold true for innovation (Bach & Matt, 2005). The first one concerns a linear model of innovation, i.e. that there is a clear link between input and output of innovation activities. It views innovation as an orderly and one-way process, starting with the discovery of new knowledge, moving through various development stages, and emerging in a final viable form. In turn, the innovation process tends to involve continuous feedback loops between the different stages, the interplay between supply sources of science, and the demand forces of the market. The systemic approach to innovation implies that, to understand the specific challenges and opportunities with respect to innovation, it is critical to examine the way in which the various actors, institutions, and structures interact and, thereby, influence driving forces and capabilities for innovation (Andersson et al., 2004). Secondly, that returns to the scale of the innovation activities are constant and divisible; one can frequently reap benefits from investments in innovation only after reaching a certain threshold. Partial knowledge or a part of a new technology is useless. Thirdly, there is no difference in the nature of the output generated by public and private funding. In turn, the role of the public sector in enhancing innovative activity is not about encouraging innovation as much as about directing the efforts to diverse and important social challenges and sustainable solutions. Thus, to arrive to a more meaningful conclusions concerning the effectiveness of public interventions and its applicability to other settings, there is a strong need to understand how and why public intervention unfolds the way it does, how inputs are transformed to outputs, how undertaken activities and involved resources lead to the observed effects. In other words, one needs to open the 'black box' of public interventions, regardless of whether one will fully inspect the contents of the box or just peek inside the box, depending on the adopted definition of a mechanism (mechanisms as systems versus the minimalist understanding of mechanisms discussed later in the article).

[1] Esther Dufo, a co-founder of the MIT Poverty Action Lab and the 2019 recipient of the Nobel Prize in Economics, is known for saying during the World Bank Conference held in 2003: "Just as randomized evaluations revolutionized medicine in the 20th century, they have the potential to revolutionize social policy during the 21st century" (Picciotto, 2012, p. 214).

[2] Selection bias arises when participants of a specific programme (intervention) are systematically different than non-participants even before they receive public support. There are two main types of selection bias: self-selection and committee-selection (agency selection). The former one occurs when companies voluntarily decide whether to apply for aid or not. In the case of aid schemes for innovative projects, companies that are more growth-oriented are more likely to become beneficiaries of such programmes and, owing to their motivation and dynamism, they are also more likely to perform better even without public assistance compared to those who lack ambition and creativity. The committee selection, in turn, occurs in the case of aid schemes where only a fraction of potential beneficiaries are awarded public support. It is argued that public authorities may follow a 'picking the winner' strategy – for instance, companies which are more engaged in RDI activity in the first place are also more likely to receive more aid for RDI. In short, firm-specific characteristics can influence in a systematic way the probability of receiving public support and heterogeneity in the treatment effect.

Linking programme activities with the programme outcomes through causal mechanisms in the programme theory

The paper is based on the premise that mechanism-based explanation can play an instrumental role in enhancing evidence-based policy for its capacity to answer questions as to why and how programmes worked or failed to work. A widely held but unsubstantiated belief is that causal inferences are made on the basis of, and protected by, sound research technique, and that the good design and analysis are the foundation stones of the ability to make causal statements (Pawson, 2008). Yet, there is a difference between establishing whether or not two variables or events are causally related to one another, and developing good explanations. Establishing causal relationships is necessary, but rarely sufficient for giving an acceptable explanation (Hedström & Wennberg, 2017).

To fill the above-mentioned deficit, theory-oriented evaluation approaches have emerged[3]. What they have in common is the idea that public intervention is a particular kind of a theory of change. However, it is far more than a simple logic model or a result chain that is so prevalent in evaluation practice, as it identifies causal assumptions as to why and under what conditions each of the link in the impact pathway is expected to work (Mayne, 2015)[4]. Thus, a robust theory of change describes not only a sequence of steps in getting from activities to impact (how does it happen), but also the contextual factors (internal and external) that should be taken into account to make change work. For example, the reason why many publicly-funded venture capital programmes in the UK underperformed against initial expectations was not addressing properly the aversion of SMEs to equity finance as a form of investment (BIS, 2011). Collaboration between companies and universities will not boost innovation if absorption capacity – i.e. an organisation's ability to identify, assimilate, transform, and use external knowledge – is not high enough (Biedenbach et al., 2018).

(Causal) mechanism is an important but insufficiently recognised component of the theory of change. The term 'mechanism' denotes a set of parts that work together to perform a particular function. These can be various cogs and wheels which are organised in such a way as to transfer motion from one part to another. Important is not only the structure but also the dynamics – how the movement is transferred from one part to another. In a similar vein, behavioural mechanisms explain how programme activities and spent resources ("appropriate ideas and opportunities") affect participants' responses – their attitude, knowledge, and behaviour in a particular context. They are situated in the interaction between programme's activities and the programme's participants. Therefore, mechanisms are usually non-observable and are highly context-specific (Lemire et al., 2020). They should be distinguished from programme's activities, as they are cognitive, affective, social responses to an intervention, leading to desired outcomes (Weiss, 1997). They explain behaviour of specific actors (thinking, decision-making, action) in a given context with specific resources, opportunities, and constraints. Providing a training course is not a mechanism; it is, for example, how the training spark an 'eye-opener' for some participants, as they recognise the relevance and value of the contents of the training to their day-to-day work (Punton et al., 2016). Intervention works when the involved resources and undertaken activities 'strike a chord' with programme's participants, and public policy-makers should acknowledge that the provided resources and activities resonate much more for

[3] Three main theory-oriented approaches can be distinguished: theory-driven evaluation by Chen and Rossi (1989), theory-based evaluation by Weiss (1997), and realistic (realist) evaluation by Pawson and Tilley (1997). For a more detailed comparison, see: Stame, 2004.

[4] Similar distinction between logic models and theories of change can be found in Patton (2008), who argues: "Specifying the causal mechanisms transforms a logic model into a theory of change" (p. 336).

certain people in certain contexts than others (Pawson, 2003). These regularities can provide useful insights while developing new policy instruments in relevant settings.

Behavioural mechanisms in the meaning described above are typically analysed in case-based research and are grounded in the generative type of causation. Generative causation differs significantly from other approaches to causation. Firstly, it does not rely merely on associations between one single cause and effect to explain the observed effects of a programme, as is the case with regularity and counterfactual approach, and takes account for interaction of causes. Social phenomena, though, are usually the effect of the combination of factors that jointly produce a change. This is something different than a claim that the effect results from many (independent) variables. Secondly, it is not limited to identifying a 'package of causes', as is the case with configurational approach, but, instead, describes the whole causal process taking place between cause and effect. Using an analogy to a recipe – generative causation offers not only a list of ingredients that are needed to make a certain meal (what combination of conditions produce the desired effect), but also the instructions on what to do with them, i.e. how to dose them, in what order, etc. (Befani, 2012).

However, with the increased interest in causal mechanisms, the way they are defined and analysed in evaluation literature and practice has become more diverse. Most of the scholars follow the above-mentioned realist accounts of causation in their conceptualisation of mechanisms (Astbury & Leeuw, 2010). Still, some take the minimalist view of a mechanism. From the minimalist perspective, mechanisms are often described as a form of intervening factors between a cause and its outcome. Such a minimalist conceptualisation of mechanisms can be present in variance-based research, where mechanisms are defined as mediator variables which are positioned on the path between an intervention and the outcomes (mediation analysis – see, e.g., Bruder et al., 2020), but also in case-based research (congruence method – see, e.g., Beach & Pedersen, 2019). In this light, Beach and Pedersen (2016) divide mechanisms into two groups: mechanisms as systems and the minimalist understanding of mechanisms. The former one enables the full inspection of the content of the 'black box' of a public intervention, while the latter one merely peeks inside, as the causal process between cause and outcome is not unpacked in detail.[5]

Realist synthesis – a way of accumulating knowledge on public interventions

Gaining insights into underlying causal mechanisms, exploring what works and what does not in a certain evaluation endeavour is one thing. Another one is to avoid the perpetual, regressive habit of 'starting from scratch', as each evaluation should respond to and develop from what is already known (Pawson, 2013). It needs to add to a growing body of knowledge. In other words, there is a need for some kind of synthesis. Evidence reviews are an important element of evidence-based policy, i.e. building on evidence which already exists. This, however, is a daunting task given the specificity of public interventions and the complexity of social problems they are aimed to deal with. One of the viable approaches is realist synthesis (also known as a realist review), which has been developed in an attempt to overcome some of the limitations of the traditional meta-analysis.

As both of the strategies aim to systematically pull together findings of a particular research problem, it can be argued that the former one is theory-driven, while the latter one is method-driven. In meta-analysis, the quality and trustworthiness of primary evidence is assessed in terms of a methodological hierarchy, in which priority is

[5] Schmitt (2020) indicates two axes around which causal mechanisms can be classified. These are: the level of analysis (behavioural mechanisms and process mechanisms that describe cause-and-effect relationships across multiple steps of the theory of change) and the methodological approach (case-based and variance-based).

given to experimental and quasi-experimental designs as the most rigorous and robust research method of determining whether a cause-and-effect relationship exists between an intervention and an outcome. Qualitative and case study designs are ranked lower and often only primary research results are used in meta-analysis. Moreover, an attempt is made to identify research results that can be generalised across contexts (Jagosh, 2019). For example, Storey (2000) introduced the 'Six steps to heaven' framework, which has been incorporated into the OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes (2008), in order to categorise them according to the sophistication in the assessment procedure. The assumption behind the Storey's Framework is that the higher position a method occupies, the more robust and convincing the results are. Realist synthesis, in turn, rejects methodological hierarchies and operates on the basis of a variety of data sources, including grey literature such as action research, administrative records, legislative analysis, conceptual critique, or personal testimony. This, however, does not imply that research quality is irrelevant, but, rather, that decisions about quality require complex contextualised judgements instead of following one pre-established checklist. Moreover, the unit of analysis in not a study, but a programme theory. Therefore, for example, only one element of a primary study can be under investigation by a realist reviewer to test a very specific hypothesis about the link between a context, a mechanism, and an outcome (Pawson et al., 2004), e.g. how to reach a target group or what are the trusted sources for those to whom an intervention will be addressed. Finally, the overarching aim of the realist synthesis is not seeking generalisable lessons or universal truths, but, rather, to find out under what conditions the intervention is more or less likely to work, or, to put it differently, what should be considered and what can be done to enhance the chances that the given intervention will succeed.

It is argued that realist synthesis is more explanatory than judgemental in nature as it seeks to find out how various combinations of context can amplify or mute the fidelity of the programme theory. This issue is well-explained by Pawson (2006) in his article provocatively entitled "Digging for Nuggets: How 'Bad' Research Can Yield 'Good' Evidence":

> "…research synthesis follows a disciplined, formalized, transparent and highly routinized sequence of steps in order that its findings can be considered trustworthy – before being launched on the policy community. The most characteristic aspect of that schedule is the appraise-then-analyse sequence. The research quality of the primary studies is checked out and only those deemed to be of high standard may enter the analysis, the remainder being discarded. This paper rejects this logic, arguing that the 'study' is not the appropriate unit of analysis for quality

Figure 1. A comparison between realist synthesis and meta-analysis

	REALIST SYNTHESIS	META-ANALYSIS
focus	theory-driven	method-driven
unit of analysis	a programme theory	a single study (research)
data sources	various data sources, including grey literature; sympathetic to the usage of a multi-method, multi-disciplinary evidence-based designs	results of primary studies hierarchy of research designs, priority given to experimental and quasi-experimental designs
sequence of steps	analysis precedes appraisal	appraisal precedes analysis
guiding question	what works for whom in what circumstances – searching for contextual success factors	what works and what does not – searching for results that can be generalised across contexts

Source: Own elaboration based on: Pawson et al., 2004; Pawson, 2006; Jagosh, 2019.

> appraisal in research synthesis. There are often nuggets of wisdom in methodologically weak studies and systematic review disregards them at its peril." (p. 127)

Hence, the key point is that there is much to be learned about a given intervention when exploring the specifics of each study. While in a traditional meta-analysis appraisal goes first and then is followed by analysis, in realist synthesis it is other way round – analysis precedes appraisal (for an example of how a realist synthesis is carried out, see: Rycroft-Malone et al., 2012).

Concluding remarks

The paper demonstrates how the feedback from evaluation can be used to enhance policymaking and policy implementation. Building upon the identified limitations of the traditional input/output approach to the impact evaluation of public interventions, theory-oriented evaluations that focus on mechanistic evidence coupled with the realist synthesis are suggested to support usable knowledge creation and its accumulation for public policy needs. To this end, the role of theory in evaluation is discussed, providing clarification of the term 'mechanism', as different conceptualisations of the term can be found in the literature and evaluation practice. To complement the posed approach – the idea of a realist synthesis for reviewing research on complex social interventions is presented and contrasted with the traditional meta-analysis.

The following practical implications and recommendations for practice can be formulated to answer the questions about why and how programmes worked or failed to work. First, one needs an evaluation approach that makes use of a theory that properly explicates the causal mechanisms linking programme activities with programme outcomes, confronting the mechanisms with empirical observations. From the point of view of evidence-based practice, one needs knowledge which not only describes what happened, but which also explains the events that took place. This allows predictions about the future, thereby making it possible to affect the future (i.e. improve future public interventions of the same kind or improve the ones in operation). Second, as different approaches to causation capture different aspects of causal relationships, distinctive features of mechanism-based causation must be recognised against other types of causation (regularity, counterfactual, configurational). None of them should be considered superior, as they answer different questions and serve different policy needs. The strength of mechanisms grounded in generative causation lies in capturing the complexity of social phenomena. They provide a fine-grained description of causal mechanisms at work, strengthen our understanding of how and why public interventions work, with whom, and under what circumstances, thereby allowing the inferences about the effectiveness of the intervention in other settings (opening the 'black box' of public interventions). However, there is nothing that would stop one from combining different approaches and methods to identify causal relationships. Quite contrary – it is even recommended. Cook (2000), for example, writes about "false choice" between theory-based evaluation and experimentation. Peck (2020) asserts that "operating experimental evaluations with a theory-based framework is ideal as both kinds of evaluations are made stronger by being in partnership" (p. 146). Frequently, the counterfactual approach is only the first step in the evaluation process, as in the second step, in-depth qualitative research is conducted to explain the observed relationship and the mechanisms that govern it.

Nevertheless, the suggested approaches are not free from certain practical problems and limitations. Providing mechanism-based explanations is very demanding in terms of time and data requirements. An in-depth understanding of micro-mechanisms at work enables gaining knowledge of more general significance, but requires taking into account also mechanisms working at the higher levels, i.e. meso and macro, emerging from a high number of micro-mechanisms being activated at the level of agents, which makes the causal

inference about public intervention a challenging task. Given the fact that the distance between the macro level and the micro level might be too big to explain certain phenomena properly (e.g. individual-level entrepreneurial outcomes through the prism of country characteristics and its institutional quality; see: Kim et al., 2016), exploring meso-level mechanisms is a promising directions for future research. As regards realist synthesis, it does not follow a standardised procedure, it is inherently pluralist and flexible, and, therefore, it is not easy to reproduce. Thus, it is not for novices. It requires substantial knowledge and expertise to identify and apprise the quality of evidence appropriately. In order to preserve the accuracy and reliability of the made judgements, one needs what Pawson (2013) calls "organised scepticism", which means that any scientific claim must be exposed to criticism. For that reason, it is imperative that any judgmental, discretionary decisions must be transparent enough to be openly challenged and scrutinised.

References

Andersson, T., Appelquist, J., & Serger, S. (2004). "Public Research and Innovation Policy for the good of society: How to assess the way forward?" Background Paper at joint IKED/INNOVA seminar. November 8, 2004. Berns, Salonger, Stockholm, pp. 17–34. Available at: http://www.iked.org/pdf/Evalproceedings05.06.03.pdf (accessed: 09.10.2021).

Astbury, B., & Leeuw, F. (2010). Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation*, *31*(3), 363–381.

Bach, L., & Matt, M. (2005). From economic foundations to S&T policy tools: A comparative analysis of the dominant paradigms. In P. Llerena & M. Matt (Eds.), *Innovation Policy in a Knowledge-Based Economy* (pp. 17–46). Springer.

Beach, D., & Pedersen, R. (2016). *Causal Case Study Methods. Foundations and Guidelines for Comparing, Matching, and Tracing.* University of Michigan Press.

Befani, B. (2012). Development Models of Causality and Causal Inference. Appendix pp. 1–24, In E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, & B. Befani (Eds). *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38. Department for International Development.

Biedenbach, T., Marell, A., & Vanyushyn, V. (2018). Industry-university collaboration and absorptive capacity: An empirical study in a Swedish context. *International Journal of Technology Management*, *76*(1/2), 81–103.

BIS (2011). *Equity finance programmes qualitative reviews of: a) ukhtf and b) the bridges fund*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32220/11-1009-bis-equity-finance-qualitative-reviews-ukhtf-bridges.pdf (accessed: 07.05.2022).

Bruder, M., Polak, J., Guffler, K., & Scheinert, L. (2020). Using mediation analysis to uncover psychological mechanisms of attitude change in a development volunteer pro-gram. *New Directions for Evaluation*, *167*, 131–143.

Byrne, D. (2013), Evaluating complex social interventions in a complex world. *Evaluation, 19*(3), 217–228. https://doi.org/10.1177/1356389013495617.

Chaney, B. (2015). Reconsidering Findings of "No Effects" in Randomized Control Trials: Modeling Differences in Treatment Impacts. *American Journal of Evaluation*, *37*(1), 45–62. https://doi.org/10.1177/1098214015573788.

Chen, H., & Rossi, P. (1989). Issues in the Theory-driven Perspective. *Evaluation and Program Planning*, *12*(4), 299–306.

Cook, T. (2000). The false choice between theory-based evaluation and experimentation. *New Directions for Evaluation*, *87*, 27–34.

Dreyfus, H. (2011). Medicine as combining natural and human science. *The Journal of Medicine and Philosophy*, *36*(4), 335–341. https://doi.org/10.1093/jmp/jhr027.

European Commission (2017). *Quality of Public Administration. A Toolbox for Practitioners*. Publications Office of the European Union.

Head, B. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society*, *29*(2), 77–94. https://doi.org/10.1016/j.polsoc.2010.03.001.

Hedström, P., & Wennberg, K. (2017). Causal mechanisms in organization and innovation studies. *Innovation: Organization & Management*, *19*(1), 91–102. https://doi.org/10.1080/14479338.2016.1256779.

Jagosh, J. (2019). Realist Synthesis for Public Health: Building an Ontologically Deep Understanding of How Programs Work, For Whom, and In Which Contexts. *Annual Review of Public Health*, *40*, 361–372. https://doi.org/10.1146/annurev-publhealth-031816-044451.

Józefowski, B. (2012). Evidence-based policy – nowa koncepcja dotykająca starych problemów. *Zarządzanie Publiczne*, *1*(19), 147–152.

Kim, P. H., Wennberg, K., & Croidieu, G. (2016). Untapped Riches of Meso-Level Applications in Multilevel Entrepreneurship Mechanisms. *The Academy of Management Perspectives*, *30*(3), 273–291.

Lemire, S., Kwako, A., Nielsen, S. B., Christie, C. A., Donaldson, S. I., & Leeuw, F. L. (2020).What is this thing called a mechanism? Findings from a review of realist evaluations. *New Directions for Evaluation*, *167*, 73–86.

Mayne, J. (2015). Useful Theory of Change Models. *Canadian Journal of Program Evaluation*, *30*(2), 119–142.

OECD (2008). *Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes*. DOI: 10.1787/9789264040090.

Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). Sage.

Pawson, R. (2003). Nothing as Practical as Good Theory. *Evaluation*, *9*(4), 471–490. https://doi.org/10.1177/135638900300900407.

Pawson, R. (2006). Digging for Nuggets: How 'Bad' Research Can Yield 'Good' Evidence. *International Journal of Social Research Methodology*, *9*(2), 361–372. https://doi.org/10.1080/13645570600595314.

Pawson, R. (2008). *Causality for beginners*. In: NCRM Research Methods Festival 2008. (Unpublished) Available at: https://eprints.ncrm.ac.uk/id/eprint/245/ (accessed: 14.10.2021).

Pawson, R. (2013). *The Science of Evaluation. A Realist Manifesto*. Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: An introduction*. ESRC Research Methods Programme, University of Manchester.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Sage.

Peck, L. (2020). Leveraging experimental evaluations for understanding causal mechanisms. *New Directions for Evaluation*, *167*, 145–160.

Picciotto, R. (2012). Experimentalism and development evaluation: Will the bubble burst? *Evaluation*, *18*(2), 213–229. https://doi.org/10.1177/1356389012440915.

Punton, M., Vogel, I., & Llyod, R. (2016). *Reflections from a realist evaluation in progress: Scaling ladders and stitching theory* (CDI Practice Paper). Institute of Development Studies. Available at: https://archive.ids.ac.uk/cdi/publications/reflections-realist-evaluation-progress-scaling-ladders-and-stitching-theory-0.html (accessed: 30.04.2020).

Rousseau, D. (2006). Is There Such a Thing as "Evidence-Based Management"? *Academy of Management Review*, *31*(2), 256–269.

Rycroft-Malone, J., McCormack, B., Hutchinson, A., DeCorby, K., Bucknall, T., Kent, B., Schultz, A., Snelgrove-Clarke, E., Stetler, C., Titler, M., Wallin L., & Wilson V. (2012). Realist synthesis: Illustrating the method for implementation research. *Implementation Science*, *7*(33), 1–10.

Saunders, M. (2011). Capturing effects of interventions, policies and programmes in the European context: A social practice perspective. *Evaluation*, *17*(1), 89–102. https://doi.org/10.1177/1356389010394480.

Schmitt, J. (2020). The causal mechanism claim in evaluation: Does the prophecy fulfil? Causal Mechanisms in Program Evaluation. *New Directions for Evaluation*, *167*, 11–26.

Stame, N. (2004). Theory-based Evaluation and Types of Complexity. *Evaluation*, 10(1), 58–76. https://doi.org/10.1177/1356389004043135.

Storey, D. J. (2000). Six Steps to Heaven: Evaluating the Impact of Public Policies to Support Small Business in Developed Economies. In D. L. Sexton & H. Landstrom (Eds.), *Handbook of Entrepreneurship* (pp. 176–193). Blackwell.

Weiss, C. (1997). Theory-based Evaluation: Past, Present and Future. *New Directions for Evaluation*, *76*, 41–55. https://doi.org/10.1002/ev.1086.