Evident and Effective?
The Challenges, Potentials and Limitations of Evaluation Research on Preventing Violent Extremism

Inga Nehlsen[i]
Inga.Nehlsen@bmi.bund.de

Janusz Biene[ii]
biene@legato-hamburg.de

Marc Coester[iii]
Marc.Coester@hwr-berlin.de

Frank Greuel[iv]
greuel@dji.de

Björn Milbradt[iv]
milbradt@dji.de

Andreas Armborst[i]
Andreas.Armborst@bmi.bund.de

[i] National Center for Crime Prevention, Bonn
[ii] Legato, Osnabrück
[iii] Berlin School of Economics and Law, Berlin
[iv] German Youth Institute, Munich

Vol. 14(2)/2020

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

1

Evident and Effective?
The Challenges, Potentials and Limitations of Evaluation Research on Preventing Violent Extremism

Inga Nehlsen[i]
Janusz Biene[ii]
Marc Coester[iii]
Frank Greuel[iv]
Björn Milbradt[iv]
Andreas Armborst[i]

[i] National Center for Crime Prevention, Bonn
[ii] Legato, Osnabrück
[iii] Berlin School of Economics and Law, Berlin
[iv] German Youth Institute, Munich

Radicalization and its prevention have increasingly become the subject of public debate in academia and in the political arena. The impact of prevention efforts is subject to public scrutiny – not only because these efforts, if successful, contribute to the common good, but also because many countries have increased public spending on prevention based on this justification. Evaluations can analyze impact and effectiveness and thereby advance knowledge about how prevention operates within different social contexts and what kind of outcomes it produces. However, there are significant challenges to developing a robust basis of evidence in the field of Preventing Violent Extremism (PVE). On the one hand, practitioners and scholars debate the notion of "evidence-based approaches", e.g. with regard to methods of data collection or appropriate evaluation designs, while some fundamentally question the concept itself. On the other hand, expectations regarding the capability and feasibility of outcome evaluations are often inflated and incompatible with real-world conditions in PVE practice. This article addresses some of the challenges that researchers face when conducting outcome evaluations in the field of PVE and it suggests pragmatic solutions. It sheds light on the state of evaluation in PVE, focusing on the German context, and gives recommendations pertaining to the commissioning, planning, implementation and utilization of (outcome) evaluations.

Keywords: radicalization, extremism, prevention, evaluation, evidence-based

The topic of preventing political and religious extremism has been experiencing a boom in the last years. The UN Secretary-General presented his "Plan of Action to Prevent Violent Extremism" in early 2016 (United Nations 2015) and multiple UN entities have been working on this issue for a number of years.[1] A "High-Level Expert Group on Radicalisation" was set up by the EU Commission in 2017 and completed its work in mid-2018, recognizing "the valuable work already done, and the achievements of EU initiatives such as the Radicalisation Awareness Network". Of 28 EU member states (at the time of writing), 23 have

[1] See, e.g., https://en.unesco.org/preventing-violent-extremism; https://www.undp.org/content/undp/en/home/demo-cratic-governance-and-peacebuilding/preventing-violent-extremism.html; 28 February 2019.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

2

some sort of national prevention or counterterrorism strategy, and many states list more than one initiative.[2] In Germany, the so-called "National Prevention Program against Islamist Extremism" (NPP) as well as a federal program aimed at promoting democracy ("Demokratie leben!" – "Live Democracy!") both saw substantial increases in funding in the last two years (bpb 2018; BMFSFJ n.d.). With increased public spending, pressure has also been rising to demonstrate the effectiveness of such programs. Referring to the national plans of action for preventing violent extremism, the UN report states that "effective monitoring and evaluation mechanisms for these plans are essential to ensuring that policies are having the desired impact" (United Nations 2015, 13) – leading us directly into the main topic of this paper.

An evaluation, in principle, can target all possible aspects of a project, such as its prerequisites, underlying concepts, processes and outcomes. This paper focuses on evaluations that investigate the possible outcomes or effects of interventions. It focuses largely on the context of Germany, aiming to give researchers, practitioners and policy-makers from other countries unique insights into this national case. Our aim is to enable a comparative perspective that will enrich the international debate on challenges, potentials and limitations of evaluating the prevention of violent extremism.

As demonstrated by the plethora of actors and agencies involved and projects funded under this label, the "Prevention of Violent Extremism" (PVE)[3] encompasses a wide range of different approaches. The German PVE landscape, in particular, is broad in scope, ranging from socio-pedagogical and educational initiatives for promoting democracy to exit and deradicalization programs.[4] While this diversity is

seen as advantageous (Unger 2016; Nordbruch 2017), experts also deem it problematic given that the enormous variety of measures is based on a rather indistinct concept of prevention (Greuel 2018). This diversity also presents a challenge when evaluating PVE projects as evaluations must remain flexible in terms of their approaches and they must be adapted to their object of inquiry, i.e. the respective prevention activity. This point is reflected in the abundance of different approaches and methods applied in the context of evaluations.

Even when focusing only on outcome evaluations as one of the many types, a broad array of possible evaluation approaches and methods remains. Selecting an appropriate methodological approach allows social researchers to design evaluations that are flexible, tailored to the object of investigation and topic-specific. At the same time, however, this plurality of methods often gives rise to disputes among professionals regarding the suitability of certain research and evaluation methods. This particularly applies to the context of outcome evaluations and the related debate surrounding "evidence-based" approaches. Even given identical research objectives, different positions can lead to greatly differing decisions as to which methods and quality standards are to be applied and, ultimately, also to divergent results.

The following sections critically address the concept of evidence-based approaches and shed light on the state of evaluation of PVE in Germany. Section one begins by defining the central concepts related to this field: radicalization, prevention (of radicalization), "evidence-based" approaches and evaluation. Section two presents the debate surrounding evidence-based approaches that has influenced the PVE evaluation environment in Germany. Section three offers insights into the German evaluation landscape for the prevention of right-wing extremism as well as Islamist extremism. Section four then turns to specific approaches for evaluating effectiveness using qualitative and quantitative research methods. We also look into the logic model and the realist approach and highlight possibilities for combining qualitative with quantitative data that can be used to identify potential impact mechanisms. Section five addresses the question of determining uniform quality standards for evalua-

---

[2] https://ec.europa.eu/home-affairs/what-we-do/networks/ radicalisation_awareness_network/ran-and-member-states/ repository_en; 28 February 2019.

[3] There are a number of terms and labels used to describe different efforts directed against radicalization and extremism (see section 1) with varying connotations. Acknowledging this heterogeneity, this paper, for the sake of simplicity, uses the term "PVE" interchangeably with "prevention of radicalization" to refer to any such efforts comprehensively.

[4] For an English review of the German PVE infrastructure (for Islamist extremism), see Said and Fouad (2018). For an overview in German (for right-wing, left-wing and Islamist extremism), see Gruber and Lützinger (2017).

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

3

tions. In closing, section six offers practical recommendations that are meant to help "establish a supporting culture of evaluation that combines commitment, certain scientific standards, transparency and also a culture of accepting mistakes" (El-Mafaalani et al. 2016, 27).[5] One way of promoting this would be to more clearly articulate evaluation mandates and to realistically examine whether the conditions for an outcome evaluation are actually met: evaluations that are not tailored to their object of investigation may go on to do more harm than good for the involved actors.

1 Defining the Core Concepts

As the central concepts used in the present article are already addressed in detail in this volume by Abay Gaspar et al. (2020) and Baaken et al. (2020), we will only discuss them briefly here to illustrate their interdependencies.

1.1 Radicalization

The widespread use of the term radicalization within public discourse might suggest that we are dealing with a functional concept with clear boundaries and no need for further clarification. However, this is far from being the case. Radicalization is a fundamentally disputed concept marked by a lack of consensus regarding several of its components. This includes the question of what should be categorized as "radical", how to define the beginning and end of radicalization processes, and to what extent a willingness to use violence or its actual use are an inevitable consequence of radicalization (Abay Gaspar et al. 2020; Neumann 2017b, 43; Schmid 2013, 5). This is relevant for PVE projects and the corresponding evaluations as the concept's indeterminacy has to be considered within both project design and evaluation research. A definition of the object of prevention that is plausible, appropriate and precise with regard to the project in question must be ensured. Only once this conceptual groundwork has been established will we be able to clarify what should be prevented or reduced. Based on this premise, evaluators can determine if and how preventive measures (might) bring about positive impacts (Kiefer 2017; Kober 2017).

Following the argumentation of Pisoiu et al. (2020) and Baaken et al. (2020), radicalization, as addressed in this paper, is understood as a social process through which an individual or a group of individuals adopt extremist views.[6] With regard to the areas of right-wing and religious extremism,[7] extremist views may include positions that oppose the principles of an open and pluralist society and/or those that deny certain groups equality (due to their ethnic background or beliefs).[8] Although the legitimation of or willingness to employ violence may accompany radicalization, the actual application of violence is just one of its many manifestations (Pisoiu et al. 2020, 1; Frindte et al. 2016, 3).

When radicalization is understood as a process, it follows that there are various points of contact – at different levels and points in time – for preventing radicalization, removing individuals from a radicalized environment, or jointly working towards their deradicalization (Biene et al. 2015, 9). Which point of access one ultimately chooses depends on the respective case, context and the availability of resources. Ideally, this should already be reflected upon in the program development phase as well as in the implementation phase, drawing on the research expertise generated through program monitoring. In terms of evaluating PVE projects, this highlights the necessity to constantly reflect on the suitability of one's course of action in light of the specific situation and context.

1.2 Preventing Radicalization

CVE/PVE is an umbrella term for strategies and approaches that aim to prevent or mitigate radicalization and/or extremism. This terminology varies in the details. The Organization for Security and Cooperation in Europe speaks of "Preventing Terrorism and

---

[5] All quotes from German sources have been translated by the authors.

[6] For the purpose of this paper, we thus adhere to a narrower understanding of radicalization than that used by other authors such as Abay Gaspar et al. (2020).

[7] Due to the sparse amount of existing research and prevention projects on left-wing militancy, as well as a lack of corresponding evaluations, we have excluded this area from this paper.

[8] This understanding – according to which radicalization refers to an increase in the extremist views and potential actions outlined above, and not in radicality – is borrowed from Quent (2016) and Pfahl-Traughber (2015). For more on the questionable connection between radicality and radicalization, see Schmid (2013, 3).

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

4

Countering Violent Extremism and Radicalization that Lead to Terrorism" (OSCE 2014) while the American debate uses the term Countering Violent Extremism (CVE). In the context of the United Nations, the commonly used term is Preventing Violent Extremism (PVE) (see Frazer and Nünlist 2015; Neumann 2017a). While there are minor conceptual differences among the respective strategies described by these terms, at the operational level, prevention work targeting radicalization is well established among international actors.

Approaches to prevention differ in the point in time at which they are carried out and in their target population. In this context, the terminology of primary, secondary and tertiary prevention (Caplan 1964) as well as universal, selective and indicated prevention is commonly applied (Gordon 1983).[9] Moreover, all three types of prevention can target single persons individually or take a systemic approach addressing individuals and their social environment (Perrez 1998).

Primary or universal prevention targets a broader population and attempts to counteract undesired developments at an early stage, in a fundamental manner. Secondary or selective prevention addresses individuals or groups of individuals that already exhibit initial signs of problematic developments or are deemed to be "at risk". Thus, the aim here is to prevent the process of radicalization from evolving. This is supplemented by training courses for professionals potentially working with vulnerable individuals. Tertiary or indicated prevention addresses people among whom the targeted problem has already fully developed, aiming to prevent recurrences or relapses. Strictly speaking, tertiary or indicated prevention describes a reactive measure tackling problems that are already manifest. Despite this terminological vagueness, demobilization and deradicalization are gener-

ally allocated to the area of tertiary or indicated prevention.

Tertiary or indicated approaches to PVE, ideal-typically, can be subdivided into demobilization, disengagement and deradicalization. The aim of demobilization work is to get an individual involved in an extremist environment to refrain from committing (further) crimes, renounce violence and withdraw from that environment (Ceylan and Kiefer 2018, 73) – although the last point is not a necessary condition for disengagement or demobilization (Neumann 2013, 8; Köhler 2015). The key feature of these approaches is that they focus on the behavioral level.

Deradicalization work, on the other hand, addresses the level of attitudes. Its objective, depending on the approach, is not only to extract the individual from an extremist group or to get him or her to renounce violence but also for them to reject extremist attitudes (also see Baaken et al. 2020). This entails reflecting upon one's individual needs, conflictual family constellations, right-wing extremist or Jihadist ideologies, personal identities shaped by these, and the possibility of social (re)integration into non-extremist milieus. Along with educational and social-work interventions, such individual casework may also include psychological interventions (Sischka 2018). Demobilization work can be understood as one step on the path towards cognitive deradicalization (Baaken et al. 2020, 5).

1.3 Evaluation and Evidence-Based Prevention of Radicalization

The attribute "evidence-based" suggests that a (scientifically) verifiable interdependency should exist between an intervention and the desired prevention goal (e.g., avoiding an undesired situation or the reduction of risk factors).

Theory, research, evaluation and the practical implementation of the respective findings together pave the way towards evidence-based prevention of radicalization.[10] In this context, evidence-based means that measures aimed at preventing social problems are developed, implemented and evaluated on a scientific basis (Beelmann 2017, 14–15). Accordingly, along with theoretical and empirical basic research into the

---

[9] Even though the concepts are not identical, for the sake of simplicity, we use the terms universal/primary, selective/secondary and indicated/tertiary prevention interchangeably in this paper, falling into line with a wide-spread understanding of these concepts (also see Trautmann und Zick 2016, 7; Ceylan and Kiefer 2018). The difference between the two concepts lies in the fact that, with the designations universal/selective/indicated, Gordon refers to the target group while Caplan, using the designations primary/secondary/tertiary, distinguishes prevention measures based on the point in time at which they occur (also see Greuel 2018).

[10] For debates about the concept of evidence, see section 2 of this article.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

5

causes of social problems (in this case: radicalization), evaluations also constitute a fixed component of evidence-oriented prevention approaches.

The meaning of "evidence-based" beyond this definition and the question of which criteria must be met remains contested (see, e.g., Bullock and Tilley 2009; Brown et al. 2018; DeGEval 2018). Three broad positions can be identified within this controversial debate. The first only assigns the status of "evidence" to results derived from evaluation studies using randomized control groups (randomized controlled trial, RCT) and excludes all other methods for generating knowledge. Running contrary to this view, the second position fundamentally rejects control-group studies as a suitable way of generating evidence. Finally, situated in between these two poles, there are numerous nuanced positions that dismiss the dogmatic conception of "evidence-based" and instead look for pragmatic alternatives, e.g., by employing mixed-methods approaches and triangulating evidence from qualitative and quantitative sources (see e.g. Beelmann 2011; de la Chaux, Kober, and Moussa Nabo 2018; Pawson and Tilley 2004).

The present paper cannot provide a comprehensive overview of the various forms of evaluation and we instead focus on one of its variations: outcome evaluations, also called effectiveness evaluations or impact evaluations.[11] These are closely connected to the frequently articulated demand for an evidenced-based approach to prevention. By focusing on this topic, we do not, however, intend to obscure the fact that other forms of evaluation play an equally important role for the development and quality assurance of PVE approaches. This includes, e.g., ex-ante evaluations conducted prior to the implementation of an intervention for the sake of supporting program development, process evaluations conducted during implementation with a consulting function, and systematic reviews of various evaluation studies. These types of evaluations

can also be a meaningful addition to an outcome evaluation (e.g., Widmer, Beywl, and Fabian 2009).

2 The Debate Around Evidence-Based Approaches in Germany

The conversation in Germany surrounding the effects of approaches to PVE is embedded in two interconnected social trends. For the context of religious extremism, in particular, there has been a noticeable trend towards "securitizing" the debate in relation to prevention, demobilization and deradicalization.[12] This mirrors trends in the domestic PVE strategies of other countries as well as in international development work, in which the danger of securitizing PVE efforts has also been criticized (see e.g. Anderlini, Rosand, and Holmes 2019, 12; Mastroe 2016, 58; Attree n.d.). The relatively recent phenomenon of Jihadist suicide bombings generally invokes a greater sense of insecurity among broad segments of society than do right-wing terror attacks, thereby giving rise to a political discourse that increasingly makes PVE a matter of internal security. Consequently, PVE measures are increasingly being assessed based on national security standards, which, according to the critics, subjects them to demands that are neither realistic nor appropriate to the nature of the work (Schuhmacher 2018a). The second trend has been towards "evidence-based approaches" (cf. Baruch et al. 2018; Commission for Countering Extremism 2019: 8), complicating this matter further. Given the role extremism plays in public perceptions of security,[13] PVE efforts are granted utmost political priority. Consequently, the question of "what works" is also posed with more vehemence. As such, according to a narrow perspective marked by a desire for (internal) security, an effective prevention program is supposed to prevent terror attacks. From a professional perspective, it seems self-evident that the total number of attacks or the number of individuals listed as a threat to national security cannot serve as a realistic measure for judging the effectiveness of single prevention projects, many of which have educational aims. Regardless, large segments of the public

---

[11] For the purpose of this paper, we use these terms interchangeably, although they are sometimes used to refer to different types of results-oriented evaluations. "Outcome" seems to be the broadest term, while "impact" often refers to longer-term results of a program and "effectiveness" sometimes specifically addresses the *extent* to which effects have been achieved. However, the use of these terms is by no means consistent across evaluation research and practice (see, e.g., Chen 2015; Schalock 2001; Salabarría-Peña, Apt, and Walsh 2007).

[12] For an overview of more recent developments, see Biene et al. (2016).

[13] https://www.ruv.de/presse/aengste-der-deutschen/grafiken-die-aengste-der-deutschen; 6 June 2018.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

6

and even many policy-makers often do not accept the claim made by experts that there are no easy answers to the question of "what works?".

Against the backdrop of expectations for easy answers, many practitioners in Germany view undifferentiated demands for "evidence-basing" to be presumptuous and obtrusive. Instead, they call for confidence in the years of practical experience held in their area of expertise as well as in research findings on radicalization processes. They deem it to be inappropriate and problematic to demand (supposedly) explicit evidence of effectiveness by all means. In contrast, others find such an evidence-based approach to be a suitable means for the results-oriented development and dissemination of successful approaches to prevention and for allocating resources more efficiently. At the same time, (demands for) evaluations can serve as effective leverage for questioning or reinforcing the legitimacy of prevention programs vis-à-vis political actors. Here, the research community is responsible for providing convincing, differentiated answers to questions of effectiveness that can help steer this debate in a fair and objective manner.

3   The State of PVE Evaluation in Germany

Following these insights into the debate surrounding evidence that influences the PVE evaluation landscape in Germany and potentially elsewhere, we now turn to concrete examples from the field. While we focus on some of the developments and current activities in the German context, these, of course, have been and are influenced by developments at the international level.

Some of the decisive international milestones in evaluation research include the request by the United States Congress to produce a report about the effectiveness of crime prevention programs. The so-called Sherman Report (Sherman et al. 1998) also significantly influenced the prevention field in Germany and contributed to the establishment of the evidence-based paradigm.[14] Today, the main focus of this strategy is to develop quality standards for evaluations and program planning, meta analyses and reviews (e.g., the systematic reviews by the Campbell Collabo-

ration)[15] as well as identify best-practice models and disseminate them via the corresponding databases.[16] One of the pioneers of the evidence-based approach in Germany was the State Prevention Council of Lower Saxony (Landespräventionsrat Niedersachsen), which spread this concept among policy-makers and practitioners by issuing the Beccaria Standards[17] and the "Green List of Prevention" database (Grüne Liste Prävention).[18]

The evidenced-based approach is also increasingly being adopted within the field of PVE (Lindekilde 2012). The Radicalisation Awareness Network (RAN),[19] founded by the European Commission in 2012, seeks and develops best practices in this field across Europe. In 2015, Madriaza and Ponsot (2015) published one of the few reviews about international approaches to the prevention of radicalization processes, Christmann (2012) compared evaluated approaches in the context of Great Britain, and Litmanovitz et al. (2017) presented the first systematic review for the Campbell Collaboration about the risk factors of radicalization with inferences for prevention. Toolkits for conducting (outcome) evaluations in PVE are also becoming increasingly available – such as the RAND Corporation's Program Evaluation Toolkit for Countering Violent Extremism (Helmus et al. 2017) or the Online Evaluation Toolkit of the EU research consortium IMPACT Europe[20].

For Europe, in particular, the possibilities of comparing, transferring and determining the success of specific approaches between countries and cultures are of great interest, and evaluations serve as an important basis for this purpose. The following section assesses the various evaluation activities in Germany.

---

[14] For a more comprehensive overview of the development of evaluation research internationally and in Germany, see Coester (2018).

[15] https://www.campbellcollaboration.org; 2 May 2018.

[16] E.g., http://www.blueprintsprograms.com; 2 May 2018.

[17] https://www.beccaria.de/nano.cms/de/Beccaria_Standards1; 6 July 2018.

[18] http://www.gruene-liste-praevention.de/nano.cms/datenbank/information; 6 July 2018.

[19] https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network_en; 16 April 2018.

[20] http://impacteurope.eu; 7 April 2018.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

7

3.1 Evaluations in the Prevention of Right-Wing Extremism

Strategies for evaluation have been the subject of much debate in the PVE field in Germany in recent years (see Lynen von Berg and Roth 2003; Strobl, Lobermeier, and Heitmeyer 2012). The approaches underlying such evaluations have been just as diverse as the programs implemented at the national, state and regional levels. The present section focuses on some of the federal programs launched by the German Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (BMFSFJ) and the German Federal Ministry of the Interior, Building and Community (BMI).

The first large-scale federal program following German reunification – the "Action Program against Aggression and Violence" (Aktionsprogramm gegen Aggression und Gewalt, AgAG) – already included funding for monitoring and evaluation. The program promoted 124 projects in focus regions across eastern Germany between 1992 and 1996. It should be noted, however, that this was not purely a prevention program but one that also pursued infrastructure-promoting objectives in the states of former East Germany. Accordingly, as Möller (2003) has pointed out, the evaluation (Bohn and Münchmeier 1997) also focused on infrastructural effects rather than on the concrete program impacts related to reducing violence and aggression. To Möller, rising numbers in right-wing extremist violence and group membership at the time "clearly deny any resounding and sustainable effects of the AgAG on the reduction of extremism" (2003, 28).

Between 2001 and 2014, multiple large-scale federal programs were implemented that aimed at promoting democracy and tolerance and preventing right-wing extremism and xenophobia. They funded thousands of initiatives, projects and interventions and employed a range of different evaluative approaches and evaluators, mirroring the variety of prevention approaches, formats and subject areas represented.

One example is the evaluation of two of those federal programs[21] carried out by the German Youth Institute (DJI) (Bischoff et al. 2011). They used a mixture of methods ranging from document analyses to gathering monitoring data and surveying external experts (external perspective) as well as experts involved in the project (internal perspective). They also conducted an outcome evaluation based on a quasi-experimental longitudinal design including before-and-after surveys of the target group of a model project. Using the results of the program evaluation, the evaluators were able to provide input for adjustments during program implementation (process evaluation) as well as give recommendations for further developing both programs (Bischoff et al. 2011). The current programs by both ministries[22] also include a monitoring and evaluation component.

When looking at the federal programs in Germany, the sheer diversity and complexity of the field becomes evident. These national programs aimed at promoting democracy and preventing radicalization, violence and prejudice are designed, first of all, to establish the necessary structures as well as to develop and test educational approaches, points of access to the target groups and possibilities for cooperation. As such, the success of these programs as an object of evaluation is not primarily defined by, for instance, the broad reduction of prejudice (let alone terror attacks); instead, they are judged by their capacity to develop approaches for promoting democracy and preventing radicalization, which can be transferred into existing child and youth welfare structures, as well as corresponding structures at the federal, state and municipal levels. This entails questions as to the tasks and the function of evaluations: What, exactly, are they supposed to investigate? What criteria should be used to measure success or failure? Which causalities or cause-effect relationships are assumed? (How) can the results of educational work be observed in the biographies of children and adolescents?

---

[21] "Vielfalt Tut Gut. Jugend für Vielfalt, Toleranz und Demokratie" ("Diversity Is Good. Youth for Diversity, Tolerance and Democracy") and "kompetent. für Demokratie – Beratungsnetzwerke gegen Rechtsextremismus" ("competent. for Democracy – Counseling Networks Against Right-wing Ex-

tremism")

[22] For more information (in German) on the model projects under the BMFSFJ program "Demokratie leben!" addressing far-right extremism as well as religious extremism, see Birtsch et al. (2016).

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

8

3.2 Evaluation in the Prevention of Religious Extremism

The prevention of religious extremism as an occupational field in Germany emerged relatively recently. Consequently, few evaluation studies in this field exist to date. Without aiming to be exhaustive, this section introduces a few of the evaluations that are currently available.

In 2017, a report was released about the work of the "Advice Centre on Radicalisation" (Beratungsstelle Radikalisierung) under the German Office for Migration and Refugees (BAMF) and its four civil society counseling partners at the federal level.[23] This report is not an outcome evaluation but rather a process and network analysis as well as an evaluation of the logical framework structuring the work of the partner organizations (Uhlmann 2017, 11). One of the main findings is that, while the counseling field is becoming increasingly professionalized – e.g., by developing a system of standards within the context of this very evaluation – further systematization of "standards, indicators of deradicalization and metrics to measure success" (Uhlmann 2017, 52) is necessary. The report points out that only once these conditions are present will a comprehensive impact analysis of the counseling work be possible.

The evaluation of the Hessen counseling center of the "Violence Prevention Network (VPN)" was published at the beginning of 2018. The stated aim of this formative evaluation was to "expand knowledge about the target groups of individuals at risk of radicalization as well as already radicalized individuals" and identify "promising possibilities for triggering and supporting disengagement processes" (Möller and Neuscheler 2018, 4). Among other things, this evaluation also contributed to further professionalization: evaluators and those being evaluated worked together to create a system of objectives that facilitates the identification and verification of impacts (Möller and Neuscheler 2018, 12).

Focusing on conceptual and structural aspects, the evaluation of "Legato" – a counseling center in Hamburg with an explicitly systemic approach – was published in November 2018 (Schuhmacher 2018b). The evaluation decidedly excluded "the question of effects on the target group" due to a lack of resources and pre-conditions that would make it possible to examine such effects. However, it did identify specific outputs (e.g., which kinds of activities were developed) as well as impacts concerning the target groups that were reached or the quality of network structures.

One of the few evaluations that explicitly assess impact is that of the KOMPASS counseling center, also run by VPN (Schroer-Hippel 2018). Although it emphasizes that the impact assessment should be viewed with caution and that it is based on preliminary criteria, the evaluation takes a courageous and important step in developing such criteria in the first place. Combining quantitative and qualitative data, the evaluators also focus on key impact factors influencing the success of counseling work.

At the time of writing, further evaluations of counseling centers and other PVE programs are either being conducted or planned. Thus, over the next few years, we can expect to see a successive expansion of the knowledge base about the structural and process logics as well as the effectiveness of prevention, demobilization and deradicalization practice in the area of religious extremism in Germany.

3.3 Interim Conclusion

Despite these generally flourishing evaluation activities, the PVE field in Germany still suffers from a lack of knowledge-based strategies. This owes to the persistent scarcity of relevant evaluations, despite advances that have been made and the positive examples that exist (see sections 3.1. and 3.2.). Another issue is a lack of transparency, as evaluation results sometimes remain unpublished.[24]

The debates surrounding monitoring and evaluation of federal, state and regional projects in the area of

---

[23] These four Local Partners are the *Hayat* counseling center, located at ZDK Gesellschaft Demokratische Kultur gGmbH, the *Grenzgänger* counseling network located at IFAK e.V., the *kitab* counseling center located at Verein zur Förderung akzeptierender Jugendarbeit e.V. (Vaja) and the *Violence Prevention Network* e.V.

[24] It should be noted that full transparency can impede building a relationship of trust between evaluators and the evaluated. Ultimately, unlimited transparency may, in turn, lead to non-transparency if, for example, those being evaluated only report their own achievements out of a fear of their mistakes and difficulties being published.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

9

PVE and democracy promotion in Germany also reflect the predominant intra-academic paradigms and controversies. The protagonists of evidence-oriented research based on statistical, control-group studies, in particular, complain of the lack of relevant studies measured on the basis of the concomitant criteria. Frindte and Preiser come to the conclusion that numerous projects rest on a "hitherto unstable foundation. Their effectiveness has often neither been theoretically grounded nor empirically verified" (2007, 35).

The German Federal Criminal Police Office (BKA) conducted a survey of 721 PVE projects in the areas of right-wing extremism, left-wing extremism and Islamism implemented in 2014/2015. With regard to the question of evaluation, the authors conclude that "the existing information about evaluation activities must be described as extremely scanty in terms of both quantity and quality" (Lützinger and Gruber 2017, 20). This survey also provides important insights on how different actors conceive of evaluation. It becomes evident that, among practitioners, evaluation research is repeatedly met with partially justified skepticism as to the objectivity, utility, informative value and feasibility of evaluation studies.

Other studies echo these conclusions. Summarizing the prevention programs targeting Islamist radicalization that are offered and implemented in Germany, Trautmann, Kahl, and Zick (2017, 3) state that: "The positive effects are often difficult to measure and the evaluation approach is sometimes met with sensitivities on the part of the actors. In some areas, there is a lack of sufficient activities or the quality is questionable". In a review of evaluations in the context of preventing Islamist extremism, Kober (2017, 233) reaches the following conclusion: "It is therefore not possible to identify any studies or evaluation reports in Germany that allow statistically sound statements, using quantitative and/or qualitative methods, on the effectiveness of measures or projects aimed at preventing religious radicalization".

Despite this somewhat deficient state of research, one clear trend is a move towards more knowledge-based prevention practice. The more recent evaluations mentioned in section 3.2 bear witness to this development. The idea of advancing knowledge- and evidence-based approaches is also firmly established in the current "National Prevention Program against Islamist Extremism"[25] with the aim of "increasing the effectiveness of extremism prevention" (5). Germany is able to rely on effective research structures and thereby satisfies a core prerequisite for addressing social issues using knowledge- and evidence-based approaches and for conducting evaluations that are appropriate to their target object.

4 Approaches to Outcome Evaluations

Even for experts in the field, the spectrum of different evaluation approaches seems difficult to manage at times. The term evaluation tends to be used excessively (Döring and Bortz 2016, 978), without a clear idea as to who evaluates what or whom, which criteria are applied, with which research interest the evaluation is conducted, or which research methods should be used.

There are no blueprints for investigating the impacts of PVE approaches. The prototype for effectiveness research, i.e., an experiment with an experimental and a control group, is not always applicable to the investigation of social programs. As a result, experts from various disciplines are searching for alternative approaches to determine the effects generated by "complex intervention programs in prevention" (Robert Koch Institute 2012; see also Baruch et al. 2018; Gielen 2015). The problem in itself is not a new one. A study by the Bertelsmann Foundation published in 2004 titled "Evaluating Civic Education – Can Impact be Measured?" addresses the questions of "[...] whether one can capture, let alone 'measure', such complex processes of change" (Uhl, Ulrich, and Wenzel 2004, 9–10) and Pawson and Tilley developed their realist approach precisely in response to this problem (Hewitt, Sims, and Harris 2012, 253).

In order to conduct an outcome evaluation, one needs a collection of observable characteristics that the program aims to influence (so-called indicators). According to Köhler, "the possibility of comparing the set objectives with the program organization and the results that are ultimately achieved is what makes an evaluation possible in the first place" (2016, 25). How

---

[25] https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/sicherheit/praeventionsprogramm-islamismus.pdf?__blob=publicationFile&v=2; 1 March 2019.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

10

can one determine whether a program or an intervention has successfully prevented radicalization or reversed respective developments? It is obvious that there is no single standardized indicator that may be applied to the entire spectrum of CVE/PVE programs without limitations.

Specifically in the area of deradicalization, four rough target categories are being discussed: "demobilization", "exiting", "integration" and "stabilization" (Neumann 2013; Köhler 2015). The first step is the renunciation of political violence and the last step entails complete deradicalization at the cognitive and the affective level. It follows that prevention work (in this case more precisely: deradicalization work) can be deemed successful when it manages – through intensive case counseling – to help an individual shift from a higher to a lower level of radicalization.

These target categories can serve as a valuable source of orientation in the area of tertiary prevention and the evaluation thereof; however, they are not suitable for prevention measures that enter at a much earlier stage in the process of radicalization. For primary prevention, in his publication on "Fundamentals for Development-Oriented Prevention of Right-wing Extremism", Beelmann (2017) specifies reducing structures of prejudice as a goal. To this end, he recommends educational approaches such as anti-racism programs, diversity training and various measures from the area of prejudice prevention. A comprehensive amount of research has already been conducted on some of these approaches, including validated indicators that make tailored evaluations possible.

The selection and operationalization of appropriate criteria of success is a crucial step for every outcome evaluation. These criteria should be specified and carefully coordinated by the respective actors on the basis of concretely defined objectives before the measure is implemented in order to ensure beneficial results, particularly with regard to the specific research interest of the evaluation. Once the actors have determined which change processes are to be investigated more closely in the course of an evaluation, they can address the question of the study's practical research approach.

The following sections describe some of the most well-known approaches and methods that can be used in outcome evaluations. These descriptions are meant to allow the reader to observe how results are generated in an evaluation process, which conclusions can (or cannot) be drawn from them, what their scope of applicability is, and how robust they are. We first focus on the two main paradigms of (quasi-) experimental and qualitative studies and then address two approaches that do not focus on one specific method of data collection but provide a general framework for generating insights about the effects and mechanisms of social interventions.

4.1 (Quasi-)Experimental Studies and Field Tests

The objective of an experimental study is to prove a causal connection between a cause and an effect. Within academic research, this approach is often considered to be the "silver bullet to knowledge" (Döring and Bortz 2016, 194) since, under ideal conditions, it eliminates all potential interfering factors that may influence the result. In its most rudimentary form (the so-called *black-box evaluation*), this kind of experiment merely proves whether the investigated program is effective; it does not provide any further insights into why, how or for whom it works. As this is generally not a satisfying result, there are numerous variations of experimental studies (e.g., factorial designs, Döring and Bortz 2016, 194) that can generate differentiated results by comparing variations of a program among multiple groups of participants. Nagin and Sampson (2019) speak about "measuring counterfactual worlds that matter most to Social Science and Policy".

One critical disadvantage of experimental studies is the lack of generalizability of their results. Unless appropriate sampling strategies, such as those used in representative population surveys, are employed when selecting participants, the results will only be valid for the small group of participants investigated in the experiment. Options for compensating this shortcoming are provided by large-scale field experiments (Eisner et al. 2012), longitudinal or panel studies and "natural experiments" in which the differences between the experimental and the control group arise through natural circumstances. For the area of PVE, and especially in socio-pedagogical interventions, the applicability of RCTs is often limited due to small

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

11

sample sizes, non-standardized interventions and ethical concerns (see below). As a result, evaluations often take a qualitative approach.

4.2 Qualitative Evaluation Studies

Qualitative studies generally approach their object of investigation based on a different conception of impact. Essentially, they seek to understand social reality (including "effects") as a consequence of communicative and interactive processes. As such, they take as their object of analysis the perspectives of the actors, their actions (including the embedded evaluative moments), and the underlying structures and rules. These are captured using empirical surveys as well as through the reconstruction of discourses and patterns of interpretation, action and judgement (see von Kardorff 2017, 244; Flick 2006). Although concepts such as "evidence" and "effect" are not generally rejected, the qualitative paradigm critically reflects on them. Kelle (2006, 129–130) points out that standardized evaluation designs are not capable of methodologically capturing the unintended side effects and surprising impacts that are typical for social processes. He thus advocates for impact evaluations that are qualitatively oriented and open to various causalities. This is important since "every causal analysis of social action must take into account that actors operate in local contexts, according to rules, and based on existing knowledge that researchers are unaware of at first sight and that these actors often pursue their objectives in ways that the researchers cannot anticipate" (Kelle 2006: 133).

In the area of PVE, a standardized survey could ascribe the effects of an anti-aggression training to the content of such training programs – however, a partially standardized interview with the participants of the program, conducted additionally, might reveal that the *relationship work* that took place between the vulnerable adolescents and the social worker before and after the actual training sessions turned out to be the real reason for the observed effects. A simple black-box evaluation would not be able to recognize this connection and could therefore lead to erroneous conclusions.

In addition, there are field-specific reasons that may call for the application of qualitative methods. The conditions for quasi-experimental effectiveness evaluations are often absent in socio-pedagogical settings and they are also not easy to create. These settings are instead characterized by a high degree of context dependency and individual problem situations while the approaches and interventions closely relate to the respective personal and social environment (*Lebenswelten*).

The evaluation design is partially independent of the type of data collected (statistical or qualitative). In this context, Lösel (2008) recommends combining the respective advantages of qualitative and quantitative methods of evaluation, highlighting the fact that many studies do so as a matter of course: "An experimental design does not necessarily imply statistical analysis (Shadish et al. 2002). [...] Even the pioneer of evaluation recommended combining qualitative-ethnographic methods with the rigorous criteria of experiments (Campbell 1979)" (Lösel 2008, 153). The *logic model* and *realist evaluation* are two evaluation approaches that endorse the combined use of qualitative and quantitative methods for data collection and analysis in order to gain insights about the impact mechanisms of the investigated programs.

4.3 The Logic Model

Evaluations employing the logic model focus on the characteristics and processes, as well as the structural parameters, through which a prevention program seeks to make its impact:

> There is agreement on the fact that the focus should no longer be exclusively on process or structural quality but rather that the evaluation of educational work should also refer to the quality of results, to summative aspects, while also considering impacts, effects and goal attainment. [...] This demand, however, cannot obscure the fact that the question of how and for what purpose, i.e., especially, in what way educational work can be evaluated appropriately has not been adequately answered [...]. Even the sometimes inflationary use of the term evidence-basing, cannot hide this. (Yngborn and Hoops 2018, 349)

The fundamental assumption of the logic model is "that programs and projects pursue specific objectives on the basis of certain framework conditions and resources as well as through certain activities planned on this basis" (Klingelhöfer 2007, 38). Accordingly, in the course of modeling, the activities comprising an intervention, the existing framework conditions, re-

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

12

sources, and objectives are identified and put in relation to each other.

> For the context of outcome evaluations, this program theory proves interesting insofar that the outcomes – meaning the objectives achieved in the program and their connection to the activities carried out by the program – are no longer merely observable as effects but they also become visible within the program theory and can be investigated in an evaluation. (Yngborn and Hoops 2018, 352)

Especially given the diverse range of available measures in PVE, spanning various disciplines and contexts, the focus of an outcome evaluation must lie on the *impact mechanisms* rooted in the concept of a program. What (presumably) brought about the change and which mechanisms are triggered by certain activities? This implies that the theoretical assumptions about such mechanisms underlying a program are made explicit and known to the evaluators and that as many contextual factors as possible have been collected and taken into account. This includes asking practitioners about their own assumptions regarding the intervention, the conceptual design, the processes, mechanisms and impacts. In this sense, the logic model is a means through which researchers may incorporate perspectives from the field.

Logic models can uncover the (assumed) impact mechanisms of a program and thereby help the involved experts attain a "microscopic view" and an "overview of change processes" within the target group (Yngborn and Hoops 2018, 363). In order to operationalize the different components of a logic model and, e.g., identify outputs and outcomes, a variety of qualitative and quantitative methods can be used. This approach therefore appears quite promising for evaluations in the area of PVE.[26]

4.4 Realist Evaluation

Another approach for investigating the impact mechanisms of social interventions is the *realist evaluation* (or *realistic evaluation*) (Pawson and Tilley 1997; 2004). Its creators, the sociologists Ray Pawson and Nicholas Tilley, see it as an obligation of the social sciences to do their part in solving social problems by producing knowledge about the impacts of social programs. However, consistent with many other critics, Pawson
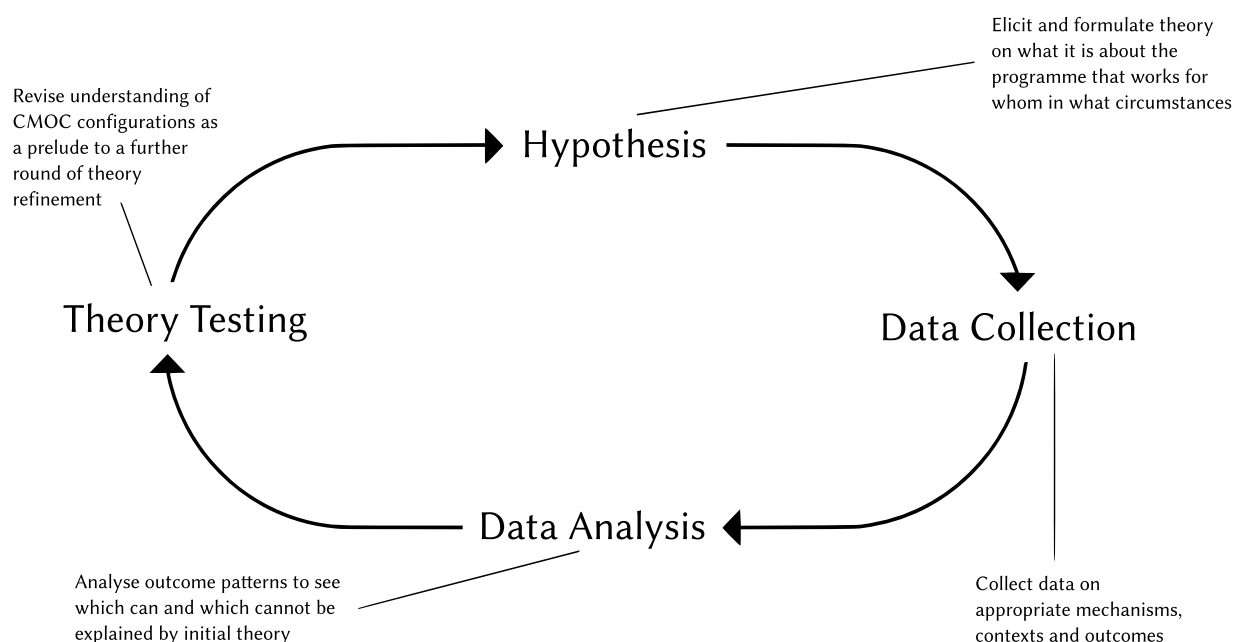
and Tilley were perturbed by the fact that the discussion around building a basis of evidence for social and prevention programs almost exclusively considered results from RCT studies (Hewitt, Sims, and Harris 2012). Tilley describes how an intervention against domestic violence in the United States that has been evaluated multiple times (using randomized control groups) reduced the risk of a relapse into domestic violence among the arrested perpetrators in some cities, while it led to increased domestic violence recidivism in others (Tilley 2010; Sherman, Schmidt, and Rogan 1992). This observation – that repeated evaluations of the same intervention in different places can generate apparently contradictory results – constitutes the basic idea of the realist evaluation. Instead of merely asking "if something works", the focus of this investigation is: "What works for whom in what circumstances and in what respects, and how?" (Pawson and Tilley 2004, 2)

Realist evaluations focus on examining the effects of social programs, thereby making them well suited for evaluating PVE interventions without being subject to the restrictive prerequisites of RCT studies and the limits of their results.

The objective of any realist evaluation is to identify the (presumed) causal mechanisms of a social intervention and to empirically test the conditions under which these mechanisms do or do not take effect. Based on these findings, researchers are able to derive generalizable assertions about functional constellations of program contents, target group characteristics and the socio-cultural context. Rather than taking an intervention or program (e.g., the counseling service of a certain organization) as the object of analysis, it is the concrete methods, resources and other characteristic features of the intervention that are analyzed. Accordingly, the evaluation results do not provide a blanket statement about whether the counseling services provided by organization X were successful or not, but rather identify which specific elements of the program, in combination with other influencing factors, have had an impact (or not) on different groups of clients. "Mechanisms are the engine" (Pawson and Tilley 2004, 18) for influencing social processes. Realist evaluations aim to uncover which con-

---

[26] For more on the logic model, also see WKKF (1998), Wyatt Knowlton and Phillips (2009) and Beywl et al. (2007).

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

13

Figure 1: The four phases of realist evaluation

Elicit and formulate theory on what it is about the programme that works for whom in what circumstances

Revise understanding of CMOC configurations as a prelude to a further round of theory refinement

Hypothesis

Theory Testing

Data Collection

Data Analysis

Analyse outcome patterns to see which can and which cannot be explained by initial theory

Collect data on appropriate mechanisms, contexts and outcomes

Source: Adapted from Pawson and Tilley (2004, 24).

ditions must convalesce for a program to be able to achieve its desired effects.

Applying this understanding of program impact to the area of PVE would entail a systematic investigation of the mechanisms that lead to the intended changes. Even though the field is marked by a vast diversity of individual interventions targeted at an even larger diversity of clients and circumstances, it is reasonable to assume that there are recurring impact mechanisms. Identifying these patterns would be one element of evidence-based PVE.

We are unable to provide a comprehensive guide for carrying out a realist evaluation here.[27] At the same time, we do not wish to overlook the fact that this type of evaluation requires considerable expertise on the part of the evaluators. The core of such an investigation lies in the empirical testing of hypotheses based on qualitative and quantitative data.

Figure 1 depicts the four typical steps of a realist evaluation: as every intervention aims for some kind of change, it is usually also based on a more or less

clearly formulated program theory and an assumption as to why and by what means the intervention should lead to the intended change (*theory of change*). It is the evaluator's task – ideally in cooperation with the designers as well as the implementers of the intervention – to transform these assumptions about an intervention's impact into testable hypotheses.

Formulating and testing hypotheses lies at the core of realist evaluations, and it requires the collection and analysis of suitable qualitative and/or quantitative data (Hewitt, Sims, and Harris 2012, 253). The hypotheses must be formulated and the data selected in such a way that allows for a possible rejection of the hypotheses based on the data. By comparing different sub-groups, realist evaluations are able to generate "a nuanced outcome pattern of successes and failures within and across interventions" (Pawson and Tilley 2004, 11).

---

[27] For a short overview, see Hewitt, Sims, and Harris (2012) and Westhorp (2014); for the original text, see Pawson and Tilley (1997).

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

14

Table 1: Selection of possible questions to and answers by a realist evaluation

The policy questions	The realist response
Did that intervention work?	It depends (in what respects?)
Should we fund X rather than Y?	Check first to see if they are commensurable
The pilot was great, should we go large?	No, play only to its strengths

(Pawson and Tilley 2004, 21)

Importantly, the results from a realist evaluation are not the be-all and end-all of the matter – they are merely provisional. They signify a momentary gain of knowledge that should be re-examined from time to time through additional evaluation cycles. This allows for findings to become consolidated in relation to questions typically asked in the context of an (impact) evaluation (see Table 1).

Evaluations modelled on the realist evaluation open up new possibilities in the, at times, deadlocked debate surrounding evidence-based approaches and the impact of PVE (Gielen 2017).[28] It has also enriched the discourse in other areas of policy-making and practice and uncovered new opportunities for action.

5 "Studies have proven that…" – The Reliability and Validity of Evaluation Studies

In order for an evaluation to deliver reliable findings, a range of prerequisites must be met when it comes to data collection and analysis. In most cases, it is not possible for every single prerequisite to be fulfilled in equal measure, meaning that restrictions are unavoidable.

This section addresses possible starting points for ensuring the quality of evaluations. This type of expert assessment can help the target audience of evaluation findings, (e.g., practitioners, policy-makers and researchers) determine how reliable and robust the results of a study actually are.

This assessment is crucial for a number of reasons. Firstly, compliance with certain standards is meant to ensure that an evaluation contains what is expected of it: a valid, relevant and appropriate assessment of a social intervention. Secondly, although no evaluation

is perfect, high-quality studies can identify and exclude a whole series of possible "technical bias[es]" (Parkhurst 2017, 7) from the outset. Technical biases can go unnoticed and lead to *systematic distortions* of the results and, ultimately, misguided recommendations and decisions. Thirdly, in both quantitative and qualitative studies, quality criteria offer a reference point that can help ensure the integrity of the results. Robust findings are more difficult to refute, whereas studies with avoidable weaknesses are susceptible to being "torn apart" by critics.

Considering the enormous influence that research and evaluation data can have on political processes, it becomes evident just how important it is to be able to reliably assess their robustness. "Assessing the (methodological) quality of evaluation studies is indispensable if one aims to arrive at a rational and well-reflected decision regarding support measures and prevention programs" (Beelmann and Hercher 2016, 98).

In this context, the same principle applies to the (outcome-based) evaluation of a preventive intervention as to the intervention itself: without a clear and precise designation of the objectives and related methods, there will be no reference point for their evaluation. A deradicalization project that does not specify the concrete aspects of (de)radicalization that it seeks to influence, and by which means, cannot be meaningfully evaluated with regard to its impacts. In the same sense, one can only assess the quality of evaluation studies in reference to the respective research interest in combination with the methodology applied.

Evaluations must always be based on their object of inquiry, meaning that the research interest and the target object should inform the selection of the methodology, and not vice versa. As such, within evaluation research, practically any method of empirical

---

[28] Also see the description of the 21st Annual Conference of the German Association for Evaluation (DeGEval e.V.): https://www.degeval.org/veranstaltungen/jahrestagungen/dresden-2018; 14 June 2018.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

15

social research may come into play. At the same time, however, this certainly desirable plurality of methods may, at times, complicate the assessment of the academic robustness of evaluation studies. Researchers from all disciplines disagree over the appropriateness of standardized systems for assessing research findings and the purposes for which they should be applied in the first place. A handful of methodological and professional quality standards have emerged from philosophical, epistemological, methodological and methodical debates.

The Joint Committee on Standards for Educational Evaluation (JCSEE) published a set of standards for program evaluation used in the United States and Canada[29] that consists of five categories: Utility, Feasibility, Propriety, Accuracy, and Accountability. In Germany, the German Association for Evaluation (DeGEval e.V.) developed a catalog of professional standards based on the JCSEE through an elaborate consultation process with its members.[30] These criteria serve as a model for planning and implementing evaluation studies and, at the same time, they serve as a benchmark for assessing the findings of concluded evaluations. They claim to be valid "across policy areas" (DeGEval 2017, 26) and "independent of the concrete evaluation approach employed" (DeGEval 2017, 28).

Along with these general professional standards for evaluations, there are also standards for research methodology. The accuracy and reliability of the results and conclusions from evaluation studies ultimately depend on the degree to which the researchers have abided by the rules of data collection and analysis. There are separate systems for this kind of assessment, including the *Maryland Scientific Methods Scale* (SMS) (Sherman et al. 1998; Farrington et al. 2002). Beelmann and Hercher (2016) have expanded the SMS by adding supplementary characteristics relevant to the context of evaluations, such as the generalizability of findings and the operationalization of variables for measuring success.

For the area of quantitative research, there is a relatively high degree of consensus among experts as to which factors influence the quality of data and results. Criteria such as internal, external and construct validity, reliability, and sampling errors, to name only a few, can be identified by appropriately trained experts without much room for interpretation and judged accordingly.

For qualitative studies, there is, of course, a greater amount of leeway for assessment using quality criteria, such as inter-subjective verifiability. Some qualitative researchers have come up with recommendations for "standards for non-standardized research" (Flick 2014, 416; Steinke 1999), while others find this to be fundamentally problematic (Hannes and Pearson 2011, 27). One commonly held view claims that standardized requirements for the quality of qualitative research compromise its diversity and ways of thinking and ultimately lead to the opposite of what they actually aim to achieve (Sandelowski and Barroso 2006, 134–137).

Broad agreement does exist among the various lines of qualitative research as to the general importance of quality standards within qualitative research (Tracy 2010). With reference to the *Cochrane Qualitative & Implementation Method Group*, a German "Handbook for Research Methods and Evaluation" speaks of "over one hundred different criteria catalogs" and asserts that "a standardized canon of core criteria capable of achieving consensus has not been created to date" (Döring and Bortz 2016, 107).

Generally, the analysis of qualitative data is based on interpretive paradigms and various methods are applied. As such, the diversity of quality standards to a large extent reflects the existing demand for context-appropriate research. At the same time, this poses significant challenges for qualitative evaluation research in terms of the respective justification of the results.

6 Conclusions and Recommendations

This article has offered insights into the state of evaluation of PVE in Germany, outlining some of the challenges and debates that characterize this field. While some aspects are certainly specific to the German context, other elements reflect situations also found internationally. While the following recommendations mainly stem from the authors' analysis and experi-

---

[29] https://jcsee.org/program/; 29 June 2020

[30] https://www.degeval.org/degeval-standards-alt/archiv/revisionsprozess; 2 May 2018.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

16

ence of the circumstances in Germany, they should be at least partially relevant to practitioners, policy-makers, donors and researchers in other parts of the world, as well.

Clearly formulate evaluation tasks

Mandates that are not clearly defined can result in poor evaluations. As such, the client commissioning an evaluation (usually policy-makers, donors or practitioners themselves) should clearly articulate which type of evaluation they are requesting and what they intend to find out. Simply commissioning an "evaluation" is akin to requesting "something to eat" at a restaurant.

Avoid automatisms

For the area of PVE, there is not simply a need for "more evaluation" but rather for more *good* evaluations. Therefore, if donors and policy-makers decide to make evaluation an obligatory component of a project or program, they should bear in mind the fact that this comes with certain requirements. Outcome evaluations are only meaningful under certain preconditions. Firstly, the conceptual framework of the intervention needs to precisely define the target object, the prevention objectives and the project logic (impact mechanisms). If this is lacking, conducting an outcome evaluation may end up doing more harm than good for the involved actors. Secondly, all participants need to agree on the research interest of the evaluation. Thirdly, the donor must allocate sufficient financial resources to enable an evaluation suitable for its target object (benchmark: ten percent of the requested funding amount).

Patience

Quick results should neither be expected in prevention work nor in its evaluation. Measures aimed at prevention, disengagement and deradicalization generally entail medium- and long-term socio-pedagogical processes as well as individual casework over several years. As such, donors should plan prevention programs and evaluations with a long-term orientation as well as allocate sufficient financial resources. A hasty evaluation (e.g., at the end of the project period)

does not benefit anyone and can frustrate the actors involved.

Ensure access to all of the relevant data

The quality of an evaluation study largely depends on the data being analyzed. Project leaders should make relevant project data available for the evaluation while considering all applicable requirements (e.g., data protection regulations). The decisive data needed for carrying out the evaluation are often only available once the project has concluded and financing has run out. Donors should therefore provide funding for evaluations running parallel to the respective program and beyond the end of the intervention period in order to ensure that all relevant project data are considered and that long-term impacts can be observed.

Plurality of methods

Determining the empirical methods most suitable for an evaluation depends on the specific research interest and the object of the evaluation – there is no one-size-fits-all solution for evaluating PVE measures. Researchers should choose the evaluation approach and the methods that are most appropriate for the object of inquiry and the context, considering the entire available repertoire.

Obligation to deliver and collect

For efficient knowledge transfer to occur, researchers, policy-makers and practitioners are each respectively under an obligation to deliver and collect. Researchers need to deliver the findings from evaluation studies in a way that is appropriate for the target audience and specifically identifies their added value. At the same time, policy-makers, donors and practitioners have a responsibility to collect, i.e. proactively request, the knowledge they need.

References

Abay Gaspar, Hande, Christopher Daase, Nicole Deitelhoff, Julian Junk, and Manjana Sold. 2020. Radicalization and Political Violence: Challenges of Conceptualizing and Researching Origins, Processes and Politics of Illiberal Beliefs. *International Journal for Conflict and Violence* 14 (2): 1-18. doi:10.4119/ijcv-3802.

Anderlini, Sanam Naraghi, Eric Rosand, and Melinda Holmes. 2019. *The SDG 16 and PVE Agendas: Different Currencies or Two Sides of the Same Coin?* ICAN Policy

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

17

Brief. http://www.icanpeacework.org/wp-content/up-loads/2019/09/SG-PVE-Policy-Brief.-Sept-2019.pdf, accessed 3 February 2020.

Attree, Larry. n.d. *Shouldn't YOU be Countering Violent Extremism?* https://saferworld-indepth.squarespace.com/shouldnt-you-be-countering-violent-extremism, accessed 3 February 2020.

Baaken, Till, Judy Korn, Maximilian Ruf, and Dennis Walkenhorst. 2020. Dissecting Deradicalization: Challenges for Theory and Practice in Germany. *International Journal for Conflict and Violence* 14 (2): 1-18. doi:10.4119/ijcv-3808

Baruch, Ben, Tom Ling, Rich Warnes, and Joanna Hofman. 2018. Evaluation in an emerging field: Developing a measurement framework for the field of counter-violent-extremism. In *Evaluation*, 475–495.

Beelmann, Andreas. 2011. The Scientific Foundation of Prevention. The Status Quo and Future Challenges of Developmental Crime Prevention. In *Antisocial Behavior and Crime: Contributions of Developmental and Evaluation Research to Prevention and Intervention*, eds. Thomas Bliesener, Andreas Beelmann, and Mark Stemmler, 137–164. Cambridge, MA: Hogrefe Publishing.

Beelmann, Andreas. 2017. Grundlagen einer entwicklungs-orientierten Prävention des Rechtsextremismus. *Gutachten im Rahmen des Wissenschafts-Praxis-Dialogs zwischen dem Landespräventionsrat Niedersachsen und der Friedrich-Schiller-Universität Jena.* http://lpr.niedersachsen.de/html/download.cms?id=2596&datei=Gutachten-LPR+Niedersachen-Version3.pdf.

Beelmann, Andreas, and Judith Hercher. 2016. Methodische Beurteilung von Evaluationsstudien im Bereich der Gewalt- & Kriminalitätsprävention: Beschreibung und Begründung eines Methodenprofils. In *Entwicklungsförderung & Gewaltprävention – Aktuelle Beiträge aus Wissenschaft und Praxis*, ed. Stiftung Deutsches Forum für Kriminalprävention, 97–116.

Beywl, Wolfgang, Jochen Kehr, Susanne Mäder, and Melanie Niestroj. 2007. Planung von Evaluationen. In *Evaluation Schritt für Schritt*, 25 f. Münster: hiba.

Biene, Janusz, Christopher Daase, Julian Junk, and Harald Müller, ed. 2016. *Salafismus und Dschihadismus in Deutschland: Ursachen, Dynamiken, Handlungsempfehlungen*. Frankfurt a.M.: Campus Verlag.

Biene, Janusz, Priska Daphi, Maik Fielitz, Harald Müller, and Irene Weipert-Fenner. 2015. Nicht nur eine Frage der Sicherheit. Salafismus in Deutschland als gesamtgesellschaftliche Herausforderung. *HSFK Standpunkte 2015/1*. Frankfurt a.M. https://www.hsfk.de/fileadmin/HSFK/hsfk_downloads/standpunkt0115.pdf.

Birtsch, Vera, Maria Kurz-Adam, Roland Merten, and Karsten Speck, ed. 2016. *unsere jugend,* 68 (11+12).

Bischoff, Ursula, Carsten Gehne, Frank Greuel, Susanne Johansson, Frank König, Tabea Schlimbach, Diana Zierold, and Eva Zimmermann. 2011. Abschlussbericht der Programmevaluation der Bundesprogramme „VIELFALT TUT GUT. Jugend für Vielfalt, Toleranz und Demokratie" und „kompetent. für Demokratie – Beratungsnetzwerke gegen Rechtsextremismus" Berichtszeitraum 01.07.2007 bis 31.12.2010. Halle (Saale). https://www.dji.de/filead-min/user_upload/bibs/686_14317_PEAbschlussbericht2010.pdf.

BMFSFJ (Bundesministerium für Familie, Senioren, Frauen und Jugend). n.d. *Über "Demokratie leben!".* https://www.demokratie-leben.de/bundesprogramm/ueber-demokratie-leben.html, accessed 10 October 2018.

Bohn, Irina, and Richard Münchmeier. 1997. *Dokumentation des Modellprojekts. Das Aktionsprogramm gegen Aggression und Gewalt AgAG, Vol. 1.* Münster: Votum.

bpb (Bundeszentrale für politische Bildung). 2018. *Nationales Präventionsprogramm gegen islamistischen Extremismus (NPP): Diese Schwerpunkte fördert der Bund.* http://www.bpb.de/politik/extremismus/radikalisierungspraevention/271943/nationales-praeventionsprogramm-gegen-islamistischen-extremismus-npp, accessed 10 October 2018.

Brown, Jennifer, Jyoti Belur, Lisa Tompson, Almuth McDowall, Gillian Hunter, and Tiggey May. 2018. Extending the remit of evidence-based policing. *Police Science & Management.* 38-51.

Bullock, Karen, and Nick Tilley. 2009. Evidence-Based Policing and Crime Reduction. *Policing: A Journal of Policy and Practice* 3 (4): 381–387.

Campbell, Donald T. 1979. Degrees of Freedom and the Case Study. In *Qualitative and Quantitative Methods in Evaluation Research*, eds. Thomas D. Cook, and CS Reichardt, 49-67. Beverly Hills, CA: Sage.

Caplan, Gerald. 1964. *Principles of Preventive Psychiatry*. 5th ed. New York, NY: Basic Books.

Ceylan, Rauf, and Michael Kiefer. 2018. *Radikalisierungsprävention in der Praxis: Antworten der Zivilgesellschaft auf den gewaltbereiten Neosalafismus.* Wiesbaden: Springer VS.

Chen, Huey T. 2015. *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective.* 2nd ed. Thousand Oaks, CA: SAGE Publications.

Christmann, Kris. 2012. *Preventing Religious Radicalisation and Violent Extremism: A Systematic Review of the Research Evidence.* http://www.safecampuscommunities.ac.uk/uploads/files/2016/08/yjb_preventing_violent_extremism_systematic_review_requires_uploading.pdf.

Coester, Marc. 2018. Das Düsseldorfer Gutachten und die Folgen. In *Evidenzbasierte Praxis in der deutschen Kriminalprävention - Ein Leitfaden für Politik und Praxis*, eds. Maria Walsh, Benjamin Pniewski, Marcus Kober, and Andreas Armborst, 37–53. Wiesbaden: Springer.

de la Chaux, Marlen, Marcus Kober, and Mitra Moussa Nabo. 2018. Überlegungen zur Schaffung einer Evidenzgrundlage für die Präventionsarbeit im Bereich islamistischer Extremismus. In *Evidenzorientierte Kriminalprävention in Deutschland: Ein Leitfaden für Politik und Praxis*, eds. Maria Walsh, Benjamin Pniewski, Marcus Kober, and Andreas Armborst, 489–509. Wiesbaden: Springer Fachmedien.

Commission for Countering Extremism. 2019. *Challenging Hateful Extremism.* https://www.gov.uk/government/publications/challenging-hateful-extremism

DeGEval - Gesellschaft für Evaluation e.V. 2017. *Standards für Evaluation, Erste Revision 2016.* Mainz: DeGEval - Gesellschaft für Evaluation e.V.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

18

DeGEval - Gesellschaft für Evaluation e.V. 2018. *Evidenz und Evaluation.* Positionspapier. https://www.degeval.org/images/stories/Publikationen/Positionspapier_Evaluation_-_Evidenz_-_Effekte.pdf, accessed 3 February 2020.

Döring, Nicola, and Jürgen Bortz. 2016. Evaluationsforschung. In *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, eds. Nicola Döring, and Jürgen Bortz, 975–1036. Heidelberg: Springer.

Eisner, Manuel, Tina Malti, Denis Ribeaud, and Barbara Müller. 2012. Groß angelegte Feldversuche in der kriminologischen Präventionsforschung. Das Zürcher Projekt zur sozialen Entwicklung von Kindern. In *Evaluation von Programmen und Projekten für eine demokratische Kultur*, eds. Rainer Strobl, Olaf Lobermeier, and Wilhelm Heitmeyer, 69–97. Wiesbaden: Springer Fachmedien.

El-Mafaalani, Aladin, Alma Fathi, Ahmad Mansour, Jochen Müller, Götz Nordbruch, and Julian Waleciak. 2016. Ansätze und Erfahrungen der Präventions- und Deradikalisierungsarbeit. *HSFK-Reportreihe „Salafismus in Deutschland", Report Nr. 6/2016.* Frankfurt a.M. http://www.hsfk.de/fileadmin/HSFK/hsfk_publikationen/report_062016.pdf.

Farrington, David P., Denise C. Gottfredson, Lawrence W. Sherman, and Brandon C. Welsh. 2002. The Maryland Scientific Methods Scale. In *Evidence-Based Crime Prevention*, eds. Lawrence W. Sherman, David P. Farrington, Brandon C. Welsh and Doris L. MacKenzie, 13–21. London: Routledge.

Flick, Uwe, ed. 2006. *Qualitative Evaluationsforschung: Konzepte, Methoden, Umsetzung.* Reinbek: Rowohlt-Taschenbuch-Verlag.

Flick, Uwe. 2014. Gütekriterien qualitativer Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung*, eds. Nina Baur, and Jörg Blasius, 411–423. Wiesbaden: Springer Fachmedien.

Frazer, Owen, and Christian Nünlist. 2015. Countering Violent Extremism in der Extremismusbekämpfung. *CSS Analysen zur Sicherheitspolitik* (183).

Frindte, Wolfgang, Brahim Ben Slama, Nico Dietrich, Daniela Pisoiu, Milena Uhlmann, and Melanie Kausch. 2016. Wege in die Gewalt. Motivationen und Karrieren salafistischer Jihadisten. *HSFK-Reportreihe „Salafismus in Deutschland", Report Nr. 3/2016.* Frankfurt a.M. https://www.hsfk.de/fileadmin/HSFK/hsfk_publikationen/report_032016.pdf.

Frindte, Wolfgang, and Siegfried Preiser. 2007. Prävention gegen Rechtsextremismus. *Aus Politik und Zeitgeschichte* 11: 32–38.

Gielen, Amy-Jane. 2017. Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How? *Terrorism and Political Violence*: 1–19.

Gielen, Amy-Jane. 2015. Supporting families of foreign fighters. A realistic approach for measuring the effectiveness. *Journal for Deradicaliazion*. 21-48.

Gordon, Robert S. 1983. An Operational Classification of Disease Prevention. *Public Health Reports* 98 (3): 107–109.

Greuel, Frank. 2018. *„Das (zu) weite Feld der Prävention oder: Wo Prävention beginnen und enden sollte."* 3/5/2018. https://blog.prif.org/2018/04/16/das-zu-weite-feld-der-praevention/.

Gruber, Florian, and Saskia Lützinger. 2017. *Extremismusprävention in Deutschland – Erhebung und Darstellung der Präventionslandschaft. Modulabschlussbericht.* Wiesbaden. https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Publikationsreihen/Forschungsergebnisse/2017PueG_ExtremismuspraeventionInDeutschland_Erhebung.pdf?__blob=publicationFile&v=4.

Hannes, Karin, and Alan Pearson. 2011. Obstacles to the Implementation of Evidence-Based Practice in Belgium: A Worked Example of Meta-Aggregation. In *Synthesizing Qualitative Research*, eds. Karin Hannes, and Craig Lockwood, 21–39. Chichester: John Wiley & Sons, Ltd.

Helmus, Todd, Miriam Matthews, Rajeev Ramchand, Sina Beaghley, David Stebbins, Amanda Kadlec, Michael A. Brown, Aaron Kofner, and Joie Acosta. 2017. *RAND Program Evaluation Toolkit for Countering Violent Extremism.* Edited by RAND Corporation, *RAND tools.* Santa Monica, CA: RAND Corporation.

Hewitt, Gillian, Sarah Sims, and Ruth Harris. 2012. The Realist Approach to Evaluation Research. *International Journal of Therapy and Rehabilitation* 19 (5): 250–259.

Kelle, Udo. 2006. Qualitative Evaluationsforschung und das Kausalitätsparadigma. In *Qualitative Evaluationsforschung: Konzepte, Methoden, Umsetzung*, ed. Uwe Flick, 117–135. Reinbek: Rowohlt-Taschenbuch-Verlag.

Kiefer, Michael. 2017. Radikalisierungsprävention in Deutschland – ein Problemaufriss. In *Salafismus in Deutschland: Jugendkulturelle Aspekte, pädagogische Perspektiven*, eds. Ahmet Toprak, and Gerrit Weitzel, 121–134. Wiesbaden: Springer Fachmedien.

Klingelhöfer, Susanne. 2007. Das Programm „Entimon": Spezifika, Potenziale und Herausforderungen einer induktiv-rekonstruierenden Evaluation anhand Logischer Modelle. In *Evaluation präventiver Praxis gegen Rechtsextremismus - Positionen, Konzepte und Erfahrungen*, eds. Michaela Glaser, and Silke Schuster, 32–52. Halle a.d.S.: Deutsches Jugendinstitut e.V.

Kober, Marcus. 2017. Zur Evaluation von Maßnahmen der Prävention von religiöser Radikalisierung in Deutschland. *Journal for Deradicalization* 11: 219–256.

Köhler, Daniel. 2015. De-Radicalization and Disengagement Programs as Counter-Terrorism and Prevention Tools. Insights from Field Experiences Regarding German Right-Wing Extremism and Jihadism. In *Countering Radicalisation and Violent Extremism Among Youth to Prevent Terrorism*, eds. Marco Lombardi, Eman Ragab, Vivienne Chin, Yvon Dandurand, Valerio de Divitiis, and Alessandro Burato, 120–150. Amsterdam: IOS Press.

Köhler, Daniel. 2016. *Strukturelle Qualitätsstandards in der Interventions- und Präventionsarbeit gegen gewaltbereiten Extremismus.* Vol. 1. Stuttgart: KPEBW.

Lindekilde, Lasse. 2012. Introduction: Assessing the Effectiveness of Counter-Radicalisation Policies in Northwestern Europe. *Critical Studies on Terrorism* 5 (3): 335–344. doi: 10.1080/17539153.2012.723522.

Litmanovitz, Yael, David Weisburd, Badi Hasisi, and Michael Wolfowicz. 2017. *What Are the Social, Economic, Psychological and Environmental Risk Factors That Lead to Radicalization and Recruitment to Terrorism?* https://campbellcollaboration.org/library/social-economic-psychological-environmental-risk-factors-radicalization-terrorism.html.

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

19

Lösel, Friedrich. 2008. Doing Evaluation Research in Criminology: Balancing Scientific and Practical Demands. In *Doing Research on Crime and Justice*, eds. Roy King and Emma Wincup, 141–170. Oxford: Oxford University Press.

Lützinger, Saskia, and Florian Gruber. 2017. *Extremismusprävention in Deutschland – Herausforderungen und Optimierungspotential. Modulabschlussbericht.* Wiesbaden.https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Publikationsreihen/Forschungsergebnisse/2017PueG_ExtremismuspraeventionInDeutschland_Herausforderung.html.

Lynen von Berg, Heinz, and Roland Roth, eds. 2003. *Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Madriaza, Pablo, and Anne-Sophie Ponsot. 2015. *Preventing Radicalization: A Systematic Review.* Montreal: International Centre for the Prevention of Crime.

Mastroe, Caitlin. 2016. Evaluating CVE: Understanding the Recent Changes to the United Kingdom's Implementation of Prevent. *Perspectives on Terrorism.* 50-60.

Möller, Kurt. 2003. Aktuelle politische Programme und Aktivitäten der pädagogischen und sozialarbeiterischen Bekämpfung von Rechtsextremismus und Fremdenfeindlichkeit — eine kritische Zwischenbilanz. In *Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet: Aufgaben, Konzepte und Erfahrungen*, eds. Heinz Lynen von Berg, and Roland Roth, 27–50. Wiesbaden: VS Verlag für Sozialwissenschaften.

Möller, Kurt, and Florian Neuscheler. 2018. *Bericht über zentrale Ergebnisse der Evaluation der Beratungsstelle Hessen - Religiöse Toleranz statt Extremismus (Managementfassung).* Esslingen. http://www.violence-prevention-network.de/de/publikationen/evaluationsberichte.

Nagin, Daniel S., and Robert J. Sampson. 2019. The Real Gold Standard: Measuring Counterfactual Worlds That Matter Most to Social Science and Policy. *Annual Review of Criminology.* 123-145.

Neumann, Peter. 2013. Radikalisierung, Deradikalisierung und Extremismus. *Aus Politik und Zeitgeschichte* 29–31 (2013). http://www.bpb.de/apuz/164918/radikalisierung-deradikalisierung-und-extremismus?p=0, accessed 27 April 2018.

Neumann, Peter. 2017a. *Countering Violent Extremism and Radicalisation that Lead to Terrorim: Ideas, Recommendations, and Good Practises from the OSCE Region.* http://www.osce.org/chairmanship/346841?download=true.

Neumann, Peter. 2017b. Was wir über Radikalisierung wissen – und was nicht. In *'Sie haben keinen Plan B'. Radikalisierung, Ausreise, Rückkehr – zwischen Prävention und Intervention*, ed. Jana Kärgel, 42–56. Bonn: Bundeszentrale für Politische Bildung.

Nordbruch, Götz. 2017. Staatliche Kontrolle statt zivilgesellschaftlichem „Wildwuchs"? Präventions- und Ausstiegsarbeit in Deutschland. https://www.ufuq.de/staatliche-kontrolle-statt-zivilgesellschaftlichem-wildwuchs-praeventions-und-ausstiegsarbeit-in-deutschland/, accessed 6 June 2018.

Organization for Security and Co-operation in Europe (OSCE). 2014. *Preventing Terrorism and Countering Violent Extremism and Radicalization that Lead to Terrorism: A Community-Policing Approach.* Vienna: OSCE.

Parkhurst, Justin. 2017. *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence.* New York, NY: Routledge.

Pawson, Ray, and Nicholas Tilley. 1997. *Realistic Evaluation.* London: Sage.

Pawson, Ray, and Nicholas Tilley. 2004. Realist Evaluation. British Cabinet Office.

Perrez, Meinrad. 1998. Prävention und Gesundheitsförderung. In *Lehrbuch Klinische Psychologie - Psychotherapie*, eds. Urs Baumann and Meinrad Perrez, 366–391. Bern: Verlag Hans Huber.

Pfahl-Traughber, Armin. 2015. Der Islamismus ist kein grüner Faschismus, sondern ein religiöser Extremismus. Eine kritische Prüfung einschlägiger Kriterien anlässlich einer öffentlichen Debatte. In *Salafismus in Deutschland. Ursprünge und Gefahren einer islamisch-fundamentalistischen Bewegung*, ed. Thorsten G Schneiders, 149–167. Bielefeld: Transcript Verlag.

Pisoiu, Daniela, Andreas Zick, Fabian Srowig, Viktoria Roth, and Katharina Seewald. 2020. Factors of Individual Radicalization into Extremism, Violence and Terror – the German Contribution in a Context. *International Journal for Conflict and Violence.* 14 (2): 1-13. doi:10.4119/ijcv-3803.

Quent, Matthias. 2016. *Rassismus, Radikalisierung, Rechtsterrorismus. Wie der NSU entstand und was er über die Gesellschaft verrät.* Weinheim Basel: Beltz Juventa.

Robert Koch-Institute. 2012. Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? *Beiträge zur Gesundheitsberichterstattung des Bundes.* Berlin. http://edoc.rki.de/documents/rki_fv/req9cSaGLflA/PDF/29PjAJd8GC8w.pdf.

Said, Behnam T., and Hazim Fouad. 2018. Countering Islamist Radicalisation in Germany: A Guide to Germany's Growing Prevention Infrastructure. *ICCT Policy Brief.*

Salabarría-Peña, Yamir, Betty S. Apt, and Cathleen M. Walsh. 2007. *Practical Use of Program Evaluation among Sexually Transmitted Disease (STD) Programs.* Atlanta, GA: Centers for Disease Control and Prevention.

Sandelowski, Margarete, and Julie Barroso. 2006. *Handbook for Synthesizing Qualitative Research.* New York, NY: Springer Publishing Company.

Schalock, Robert L. 2001. *Outcome-Based Evaluation.* 2nd ed. New York, NY: Springer US.

Schmid, Alex P. 2013. Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review. *ICCT Research Paper.* The Hague. https://www.icct.nl/download/file/ICCT-Schmid-Radicalisation-De-Radicalisation-Counter-Radicalisation-March-2013.pdf.

Schroer-Hippel, Miriam. 2018. *Beratungsstelle Kompass — Toleranz statt Extremismus. Evaluation eines Projekts zur Radikalisierungsprävention.* Berlin. https://www.berlin.de/lb/lkbgg/landesprogramm/kompass_zwischenbericht-der-evaluation-19102017.pdf.

Schuhmacher, Nils. 2018a. Ein neues Bild der Prävention? Zur Tendenz der „Versicherheitlichung" im pädagogischen Feld. In *Gewaltorientierter Islamismus im Jugendal-*

IJCV: Vol. 14(2)/2020
Nehlsen, Biene, Coester, Greuel, Milbradt, Armborst: Evident and Effective?

20

ter – Perspektiven aus (Jugend-)Forschung und Jugendhilfe. 2. Sonderband Sozialmagazin, eds. Michaela Glaser, Anja Frank and Maruta Herding, 158–166. Weinheim: Beltz Juventa.

Schuhmacher, Nils. 2018b. Evaluation der Hamburger "Fach- und Beratungsstelle für religiös begründete Radikalisierung - Legato". Hamburg. https://www.wiso.uni-hamburg.de/ fachbereich-sowi/professuren/hentschel/forschung/lega- to/legato-abschlussbericht-voe.pdf.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA: Houghton Mifflin.

Sherman, Lawrence W., Denise C. Gottfredson, Doris L. MacKenzie, John Eck, Peter Reuter, and Shawn D. Bushway. 1998. Preventing Crime: What Works, What Doesn't, What's Promising. Research in Brief. National Institute of Justice. DC. https://files.eric.ed.gov/fulltext/ ED423321.pdf.

Sherman, Lawrence W., Janell D. Schmidt, and Dennis P. Rogan. 1992. Policing Domestic Violence: Experiments and Dilemmas. New York, NY: Free Press.

Sischka, Kerstin. 2018. Was können Psychotherapeuten zur Radikalisierungsprävention und Deradikalisierung beitragen? http://www.ufuq.de/was-koennen-psychothera- peuten-zur-radikalisierungspraevention-und-deradikali- sierung-beitragen/, accessed 27 April 2018.

Steinke, Ines. 1999. Kriterien qualitativer Forschung: Ansätze zur Bewertung qualitativ-empirischer Sozialforschung. Weinheim: Juventa.

Strobl, Rainer, Olaf Lobermeier, and Wilhelm Heitmeyer, eds. 2012. Evaluation von Programmen und Projekten für eine demokratische Kultur. Wiesbaden: Springer.

Tilley, Nicholas. 2010. Realistic Evaluation and Disciplinary Knowledge: Applications From the Field of Criminology. In Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences, ed. Jos Vaessen, 203–236. New Brunswick, NJ: Routledge.

Tracy, Sarah J. 2010. Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research. Qualitative Inquiry 16 (10): 837–851. doi: 10.1177/1077800410383121.

Trautmann, Catrin, and Andreas Zick. 2016. Systematisierung von in Deutschland angebotenen und durchgeführten (Präventions-) Programmen gegen islamistisch motivierte Radikalisierung außerhalb des Strafvollzugs. kriminalpraevention.de. https://www.kriminalpraevention.de/ files/DFK/dfk-publikationen/2016_systematisierung_isla- mismusprävention_langfassung.pdf, accessed 27 April 2018.

Trautmann, Catrin, Wolfgang Kahl, and Andreas Zick. 2017. Prävention von islamistischer Radikalisierung und Gewalt. Eine Systematisierung von Präventionsansätzen in Deutschland. forum kriminalprävention 2017 (1): 3–9.

Uhl, Katrin, Susanne Ulrich, and Florian M. Wenzel, ed. 2004. Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung.

Uhlmann, Milena. 2017. Evaluation of the Advice Centre on Radicalisation. Final Report Research Report 31. Nuremberg. http://www.bamf.de/SharedDocs/Anlagen/EN/Pu- blikationen/Forschungsberichte/fb31-evaluation-bera- tungsstelle-radikalisierung.pdf?__blob=publicationFile.

Unger, Christian. 2016. Die mühsame Suche nach den Wegen aus der Radikalität. Der Westen. https://www.derwes- ten.de/politik/islamismus-die-muehsame-suche-nach- wegen-aus-der-radikalitaet-id12314633.html, accessed 6 June 2018.

United Nations, General Assembly. 2015. Plan of Action to Prevent Violent Extremism, Report of the Secretary-General, A/70/674 (24 December 2015). http://undocs.org/A/ 70/674.

von Kardorff, Ernst. 2017. Qualitative Evaluationsforschung. In Qualitative Forschung, eds. Uwe Flick, Ernst von Kardorff and Ines Steinke, 238–250. Reinbek: Rowohlt-Taschenbuch-Verlag.

W.K. Kellogg Foundation (WKKF). 1998. Logic Model Development Guide. Battle Creek, MI. https://www.bttop.org/ sites/default/files/public/W.K.%20Kellogg%20LogicMod- el.pdf.

Westhorp, Gill. 2014. Realist Impact Evaluation: An Introduction. Methods Lab Publication. London. https:// www.odi.org/publications/8716-realist-impact-evalu- ation-introduction.

Widmer, Thomas, Wolfgang Beywl, and Carol Fabian, eds. 2009. Evaluation. Ein systematisches Handbuch. Wiesbaden: Springer.

Wyatt Knowlton, Lisa, and Cynthia C. Phillips. 2009. The Logic Model Guidebook: Better Strategies for Great Results. Los Angeles, CA: SAGE Publications.

Yngborn, Annalena, and Sabrina Hoops. 2018. Das Logische Modell als Instrument der Evaluation in der Kriminalitätsprävention im Kindes- und Jugendalter. In Evidenzorientierte Kriminalprävention in Deutschland: Ein Leitfaden für Politik und Praxis, eds. Maria Walsh, Benjamin Pniewski, Marcus Kober and Andreas Armborst, 349–368. Wiesbaden: Springer Fachmedien.