Improving evaluations of programs for prevention of radicalization and violent extremism.

AN EXPLORATORY INTERNATIONAL STUDY

PABLO MADRIAZA, DAVID MORIN, SARAH OUSMAN, CATELINE AUTIXIER,
GHAYDA HASSAN ET VIVEK VENKATESH

JANUARY 2022

HOW TO CITE THIS REPORT

THE AUTHORS

Pablo Madriaza

Associate Professor at the University of Quebec in Montréal, Scientific Coordinator of the UNESCO Chair in Prevention of the Prevention of Radicalisation and Violent Extremism (UNESCO-PREV) and General Coordinator of the Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV).

David Morin

Full Professor at the School of Applied Politics at the University of Sherbrooke and co-holder of the UNESCO Chair in Prevention of Radicalisation and Violent Extremism (UNESCO-PREV).

Sarah Ousman

Doctoral student in psychology (Psy.D./Ph.D.) at the University of Quebec in Montréal (UQÀM) and FRQSC scholarship holder. Her thesis focuses on the assessment of the risk of violence related to radicalization and extremism. She is the Mapping Coordinator of the CPN-PREV.

Cateline Autixier

Holder of a master's degree in law and a master's degree in criminology, Cateline was a research assistant and coordinator at the International Centre for Comparative Criminology (ICCC) in Montréal. She then worked as an analyst and project manager at the International Centre for the Prevention of Crime (ICPC).

Ghayda Hassan

Clinical psychologist, full professor of clinical psychology at the University of Quebec in Montréal (UQÀM) and co-holder of the UNESCO Chair in Prevention of Radicalisation and Extremist Violence (UNESCO-PREV).

Vivek Venkatesh

Full professor in the Department of Education at Concordia University in Montréal, he is the director of Project SOMEONE (Social Media Education Every Day), a filmmaker, the creator of the Grimposium festival and conference series and co-holder of the UNESCO Chair in Prevention of Radicalisation and Extremist Violence (UNESCO-PREV).

Table of contents

EXECUTIVE SUMMARY

The overall objective of this study was to document the experiences lived, challenges faced and lessons learned by researchers and practitioners who have conducted evaluations of programs for prevention of radicalization and violent extremism (PRVE) in various Western countries.

Highlights

We define program and practice evaluation broadly as assessing the design, implementation or outcomes of an initiative as systematically and impartially as possible, for purposes of learning or decision-making.[1] The more specific objectives of this study were as follows:

01

Identify lessons learned from conducting evaluations of PRVE programs, in particular lessons about the challenges and obstacles faced in this process;

02

Identify the needs of Canadian practitioners and policymakers with regard to PRVE program evaluations.

To meet these objectives, we interviewed 54 individuals in this study: 32 researchers/ evaluators, 19 practitioners and three representatives of the governments of nine different countries (Belgium, Canada, Finland, France, Germany, the Netherlands, Norway, the United States, and the United Kingdom).

The rest of this executive summary presents the six main findings from this study, which represent a large share of the knowledge gained from it. A list of detailed recommendations appears at the end of this report.

---

1   A more complete definition is presented in the methodology section of this report.

FINDING 1: PRVE program evaluations should be a collaborative process.

Evaluations of PRVE programs should be a collaborative process in which all stakeholders participate, although that is too seldom the case at present. For example, the persons interviewed for this study recommended that the program-evaluation team be composed of both program staff and external evaluators, while drawing on a network of outside partners for any competencies needed for a successful evaluation but not available on this team. Within this constellation of actors, the roles and guidelines for the evaluation should be well defined so as to facilitate their efforts.

In a collaborative process such as evaluation, the quality and diversity of the relationships among the actors involved constitute an essential factor for success. Throughout this process, priority must therefore be given to establishing and maintaining trusting relationships among the evaluators, the team members of the program being evaluated, the program's users and its funders.

The people interviewed in this study also stressed the need for the evaluators to take a flexible methodological approach and exercise adaptability from the start of the evaluation. It is recommended that the evaluators employ a participatory, co-creative approach that incorporates the views of all stakeholders, including the program practitioners.

FINDING 2: An evaluation culture is a success factor for the evaluation process.

The presence or absence of an evaluation culture among the various stakeholders represents a key factor that can facilitate or impede the evaluation process. When an evaluation culture is present, the various stakeholders tend to better understand the usefulness of the evaluation, its place in and impact on their work, and the role that they can play in the evaluation process. But where there is no such culture, and the stakeholders have no specific knowledge about evaluation, particularly in the field of PRVE, they may misunderstand the evaluation process. Evaluation may then be regarded as a lower priority than the actual program activities and become a source of tension within the organization.

FINDING 3: Planning the evaluation process and developing theories of change are primary facilitating factors.

Thorough planning is probably one of the most important factors for the success of a PRVE program evaluation. This planning should ideally be done while the program is still being designed, but failing that, then when it is first being implemented. This planning must also incorporate, from the very outset, the development of theories of change to guide both the program's activities and the evaluation process. Theories of change are fundamental for establishing cause-and-effect relationships, in particular in fields such as PRVE, where evidence can be hard to come by. But our findings show that very few PRVE programs actually plan for their evaluations and develop such theories of change[2]. These programs also often lack clear objectives, which in turn limits their ability to define clear indicators for their evaluations.

---

2   Theories of change can be used to explain how a PRVE program's planned activities are supposed to produce the desired outcomes, and, if need be, to identify the mechanisms underlying these actions.

FINDING 4: Sufficient, dedicated funding for PRVE program evaluations is indispensable.

Several of the evaluators whom we interviewed said that they were not given enough time and money to do their evaluations by the methods they thought best. When funding for an evaluation is not allocated as part of the planning for the program, the evaluation is subjected to many constraints and becomes a source of tension. In many cases, programs focus their resources on the program activities themselves and treat evaluation as a secondary responsibility. When funding is earmarked for an evaluation but is insufficient, its quality is diminished and its objectives become harder to achieve. Thus every PRVE program evaluation should be allocated sufficient resources, independent of those allocated for the program activities themselves, and these resources should be allocated at the very start of planning for the program.

FINDING 5: PRVE program evaluations should apply methodological compromises and mixed designs.

Differences in evaluation cultures and in the evaluation approaches preferred by the various stakeholders may sometimes resemble "paradigm wars," leading to methodological conflicts that cause tensions in the evaluation process. Examples of such conflicts include the choices between impact evaluations and process evaluations, quantitative and qualitative methods, and in-house and external evaluation teams. In general, such conflicts relate to the perceived degree of "objectivity" or "subjectivity" that will be tolerated in the evaluation. For people who favour more "objective" approaches, quantitative impact evaluations by external teams represent the "gold standard." But many researchers and practitioners question the feasibility of this approach, because of time and budget constraints and gaps in knowledge in the field of PRVE (the factors explaining radicalization, for example).

In addition, some practitioners perceive "objective" approaches as inhuman and mechanistic. In particular, some practitioners regard the use of control groups as unethical, because the members of the control group would be denied any program benefits that the members of the experimental group might receive. Also, although external evaluation teams may offer better assurance of independence, they sometimes lack in-depth knowledge of how programs actually operate in the field, which can make it harder for such teams to do a good job. In light of all these considerations, PRVE program evaluations should apply a methodological compromise, in which the evaluation team includes people from inside and outside the organization and uses a mixed design to evaluate both the process by which a program is implemented and the effects that it has in the long term.

FINDING 6: Ethical issues are a central concern in PRVE program evaluations.

This study identified two types of ethical issues involved in evaluating PRVE programs: 1) issues regarding the evaluators' independence from political and financial pressures, and 2) issues regarding the ethics of the research or the evaluation as such. Regarding the first type of issues, a number of the people whom we interviewed for this study said that they had been subjected to pressure from their funders and/or government agencies—usually pressure to omit from their evaluation reports any findings that did not show the programs in a favourable light. Such pressures clearly conflict with the need for an independent evaluation process. The second type of ethical issues are widely debated in the worldwide literature on evaluation of PRVE programs. These issues relate mainly to the potential for such programs and their evaluations to stigmatize members of vulnerable populations, in particular by associating radicalization with specific groups. Another aspect of these issues is how to reconcile the need to access sensitive data with the need to protect participants' privacy.

01

Introduction

One of the few points of consensus about programs for the prevention of radicalization and violent extremism (PRVE programs) is that the processes by which such programs are implemented and the effects that they have are not evaluated often enough (Baruch, *et al.*, 2018; Bellasio *et al.*, 2018; Feddes and Gallucci, 2015; Gielen, 2017; Hirschi and Widmer, 2012; Mastroe and Szmania, 2016; Romaniuk, 2015). tudies on PRVE program evaluation have tended to treat it as a technical, scientific subject, focussing solely on methodological issues such as what to evaluate, how to conduct the evaluation and what information sources to use. These studies often try to determine how many evaluations have been conducted and, if possible, what their quality has been like and what their shortcomings have been.

Assessments of the state of PRVE program evaluation in the literature are usually very negative. The literature generally finds that there are not many such evaluations and that they are insufficient to measure the programs' impact on radicalization and violent extremism in the real world. Most of the existing evaluations **make very scant use of primary data** (Baruch *et al.*, 2018; Marret *et al.*, 2017; Romaniuk, 2015). They employ **methodologies that are unsystematic, inconsistent and unharmonized** (Davey *et al.*, 2019; Feddes *et al.*,2015; Lindekilde, 2012; Marret *et al.*, 2017). They **evaluate PRVE programs over very short terms and often fail to use control groups** (Marret *et al.*, 2017; Romaniuk, 2015). Lastly, these methodologies display very little transparency about their sources of information, especially in the case of tertiary prevention programs, and many evaluators are not independent (Horgan and Braddock, 2010; Mastroe and Szmania, 2016; Williams and Kleinman, 2014).

The explanations for these problems are related to the scientific dimension and focus mainly on conceptual and methodological issues, such as problems with conceptualizing radicalization and prevention (Lindekilde, 2012; Mastroe and Szmania, 2016; Ris and Ernstorfer, 2017); problems with the small number of cases and access to information (Lindekilde, 2012; Pistone, *et al.*, 2019); problems of methodology and analysis as such, in particular causal relationships (Holmer, *et al.*, 2018; Lindekilde, 2012; Madriaza and Ponsot, 2015; Mastroe and Szmania, 2016; Ris and Ernstorfer, 2017) and the lack of clear, coherent, harmonized indicators for evaluating these programs (Baruch *et al.*, 2018; Davey *et al.*, 2019; Feddes and Gallucci, 2015; Horgan and Braddock, 2010; Lindekilde, 2012; Mastroe and Szmania, 2016; Romaniuk, 2015).

Many of these assessments of the state of PRVE program evaluation have been made through literature reviews and evaluative studies, but few have gathered primary data — for example, by interviewing the people involved in the evaluations — so these assessments have remained largely a discussion among researchers. Those literature reviews that have gone further have more systematically analyzed the problems identified by the authors of the studies reviewed. This was especially true of two reviews. In the first, Bellasio *et al.* (2018) identified such problems by analyzing the discussion sections of evaluation studies published between 2013 and 2018 on strategies, policies and interventions to counter terrorism or to prevent or counter violent

extremism. In the second review, our own systematic review of PRVE program evaluations conducted through 2019 (Madriaza *et al.*, 2022), we inventoried the limitations that their authors had identified, as shown in Table 1.

Table 1. Limitations identified by authors of PRVE program evaluation studies (Madriaza *et al.*, 2022)

1. Methodological limitations

    a. Design
        i. Exploratory design of studies
        ii. Lack of measurements taken before and after intervention
    b. Indicators
        i. Limited
        ii. Inadequate
    c. Data collection
        i. Data based on perceptions
        ii. Credibility of information obtained
        iii. Inadequate or limited data collection tools
        iv. Access to data
        v. Incomplete information
    d. Participants
        i. Small samples
        ii. Unrepresentative samples
        iii. Lack of control groups
        iv. Homogeneous samples
        v. Social desirability bias
        vi. Reticence

2. Analytical limitations

    a. Analytical capacity
    b. Depth of analyses
    c. Generalizability of findings
    d. Type of analysis conducted
    e. Type of results obtained
    f. Causal relationships
    g. Sensitive information

3. Limitations regarding evaluators

    a. Lack of independence
    b. Tendency to apply subjective judgments in their analyses
    c. Translation problems

4. Limitations regarding programs and practitioners

    a. Programs
        i. Short periods in which programs are carried out
        ii. Lack of clear definitions
        iii. Program funding
    b. Practitioners
        i. Lack of commitment to the evaluation process
        ii. Security issues
        iii. Role of external practitioners

5. External limitations

    a. Time limitations
    b. Contextual and security limitations
    c. Budget limitations

There are thus very few primary data concerning evaluation issues encountered in the field or the views of actors from outside the research community. For example, in our systematic review, we found only one practitioner identified as the author of a study dealing with the evaluation of a PRVE program (Madriaza *et al.*, 2022). In the rare assessments where practitioners and even researchers have been interviewed directly, the information supplied and the analyses performed remained fairly anecdotal. In 2013, Chowdhury Fink and her team analyzed discussions with researchers and practitioners at a symposium in Ottawa, Canada to draw some lessons about the evaluation of programs for countering violent extremism. Romaniuk (2015) used this same information and additional interviews to assess the situation regarding evaluation in this field. Chowdhury Fink, Romaniuk and Barakat (2013) identified the conceptual challenges mentioned above and other, more operational challenges, such as the limited availability of evaluation expertise, funding, the lack of a culture of transparency,

receptiveness to the findings of evaluations and the incorporation of evaluation at the program design stage—all issues that we see again in the present study.

Other studies have addressed the subject of evaluation only tangentially, while dealing with extremism-prevention issues in general. For example, Davey *et al.* (2019) interviewed 19 practitioners involved in online and offline programs for prevention of right-wing extremism and found that in both kinds of programs, evaluations were not always done and that particularly in the case of offline programs, success was hard to define. The study on evaluation that preceded the present report (Clement, *et al.*, 2021) mainly re-analyzed data from another study that focused on issues of practice in prevention of violent extremism at the international level (Madriaza, *et al.*, 2017).

> In the present study, one notable finding was that practitioners often give little consideration to evaluation, perceiving it primarily as a constraint, but that it can also be reconceptualized as an opportunity for the practice community (Clement *et al.*, 2021).

But evaluation is not solely a methodological or conceptual issue or a concern for researchers and evaluators alone. On the contrary, many other actors, such as practitioners, program users, and funders, are involved in the evaluation process too. Although methodological issues remain a central concern, they arise in a specific practice and a specific context that make the process far more complex than the academic debate alone. Evaluation must also be understood as a process of co-ordinating these various realities and, in order to address this complexity, these actors must be interviewed to understand in more detail how evaluation is actually done in the field. That is why, in parallel with our systematic review of PRVE program evaluations, we also interviewed the actors from the field—the people who have performed such evaluations and the practitioners whose programs have been the subject of them. We did so to identify not only the lessons learned from carrying out evaluations of this kind (and especially the challenges and obstacles encountered in the evaluation process), but also the needs of practitioners with regard to program evaluations.

This study is part of the PREV-IMPACT Canada project (https://chaireunesco-prev.ca/en/projects/prev-impact-canada/) led by the UNESCO-PREV Chair and supported by the Community Resiliency Fund of the Canada Centre for Community Engagement and Prevention of Violence and by Public Safety Canada. The PREV-IMPACT Canada project aims to develop and implement Canadian models for assessing practices in primary, secondary and tertiary prevention of violent radicalization (VR) and, ultimately, to build the capacity of key stakeholders in VR prevention in Canada. The first component of this project concerns research, and its objectives are to:

a. document and compare strategies and tools used to evaluate VR prevention programs in Canada and elsewhere, on the basis of evidence and existing practices;

b. develop distinct, innovative evaluation models (logic models, strategies, tools, indicators, methodology) adapted to local primary, secondary and tertiary prevention programs in order to guide VR prevention policies and programs in Canada;

c. test the evaluation models on three Canadian VR prevention programs.

02

Methodology

1. OBJECTIVES

The overall objective of this study was to document the experiences lived, challenges faced and lessons learned by researchers and practitioners who have conducted evaluations of programs for prevention of radicalization and violent extremism (PRVE) in various Western countries.

We adopt the definition of evaluation provided by the United Nations Evaluation Group:

> An evaluation is an assessment, conducted as systematically and impartially as possible, of an activity, project, program, strategy, policy, topic, theme, sector, operational area or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency, impact and sustainability. An evaluation should provide credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations and lessons into the decision-making processes of organizations and stakeholders. (UNEG, 2016, p. 10)

The more specific objectives of this study were as follows:

a. to identify lessons learned from conducting evaluations of PRVE programs, in particular lessons about the challenges and obstacles faced in this process;

b. to identify the needs of Canadian practitioners and policymakers with regard to PRVE program evaluations.

The central questions examined in this study were therefore as follows:

- What lessons have been learned by the people who have conducted such evaluations?

- What challenges, facilitating factors and obstacles did these people encounter in conducting these evaluations?

- What are the needs in the field with regard to PRVE program evaluations?

- What recommendations do the researchers, practitioners and government representatives whom we interviewed offer for conducting such evaluations?

2. INTERVIEWEES

To answer these questions, we interviewed three kinds of people: 1) researchers who had conducted evaluations of PRVE programs, 2) practitioners who had worked in PRVE programs that had undergone evaluations, and 3) government representatives. As Table 2 shows, we interviewed a total of 54 people from nine different countries: 32 researchers/evaluators, 19 practitioners and three government representatives. Note that in selecting Canadian researchers and practitioners, we applied the criterion of experience with PRVE program evaluations more flexibly, because very few such programs have been evaluated in Canada.

Table 2. Interviewees, by country and role

Country	Researchers/ evaluators	Practitioners	Government representatives	Total
Belgium	3	1		4
Canada	4	14	1	19
Finland	2	3		5
France	2	1	2	5
Germany	4			4
Netherlands	5			5
Norway	1			1
United Kingdom	7			7
United States	4			4
Total	**32**	**19**	**3**	**54**

3. DATA COLLECTION

To select our interviewees, we used three strategies for our initial data collection:

a. We identified the authors of PRVE program evaluations included in two systematic reviews of such evaluations, both conducted by the Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV). One of these reviews covered evaluations of primary and secondary prevention programs, while the other covered evaluations of tertiary prevention programs (Hassan, *et al.*, 2021a; Hassan, Brouillette-Alarie, *et al.*, 2021b).

b. We consulted some experts to identify the most relevant interviewees.

c. To select Canadian interviewees, we also used the cartography that the CPN-PREV team had developed to identify secondary and tertiary prevention programs in Canada (Hassan, *et al.*, 2020).

d. Using these three strategies, we prepared a preliminary list of researchers, practitioners and government representatives. We then gathered data from these individuals in semi-directed interviews. Originally, we had planned to make four research trips to Europe, the United States and other parts of Canada, during which we would conduct these interviews as well as make field observations. But we were able to complete only two trips—one to the United Kingdom, the other to France and Belgium—before the COVID-19 pandemic broke out. To gather the rest of our data, we held an online focus group with Canadian practitioners and conducted online interviews remotely with our other interviewees. In total, we conducted 40 interviews and held one focus group, the recordings of which were transcribed verbatim and anonymized.

4. DATA ANALYSIS

Our team's researchers and their research assistants used NVIVO software to perform descriptive thematic analyses of the anonymized content from the interviews and the focus group. This content was coded in the following three steps (Saldaña, 2013).

a. **First coding cycle:** We organized the data into thematic categories and codes and analyzed the general topics addressed on the basis of our research questions: a) an overall description of the evaluation methods and procedures that our interviewees had used, b) the lessons that they had learned in carrying out the evaluations, c) the challenges, facilitating factors and obstacles that they had encountered in the course of the evaluations, d) the needs that they had identified concerning the evaluations and e) their recommendations.

b. **Second coding cycle:** In the second coding cycle, we analyzed the general topics in greater depth, reorganizing and reclassifying the information coded in the first cycle into more specific categories within the chosen topics. During this cycle, we used an open, "in vivo" coding scheme to reflect the concepts used by the interviewees. We then reorganized these specific categories into a coding tree that is represented by the titles and subtitles in the present report.[1]

b. **Integration:** In this third step, we examined the conceptual categories in relation to our research objectives and conceptual framework.

---

[1]   In this report, we could not cover all of the topics about which we had questioned our interviewees and that we included in the data analysis. Instead, we have concentrated on the challenges, facilitating factors and obstacles that our interviewees had encountered in conducting PRVE program evaluations. This was the case, in particular, for the overall description of the methods and procedures that they had used in their evaluations. However, the lessons that they had learned in carrying out these evaluations, the needs that they had identified concerning them, and their recommendations were merged with the first categories, because the interviewees often used the same ideas in responding to our questions.

5. LIMITATIONS OF THIS STUDY

This study does have certain limitations that should be mentioned. First, researchers and evaluators are overrepresented in comparison with practitioners and especially with government representatives. For the most part, the analyses that included the role of funders and policymakers were a reconstruction from the experiences of the researchers, evaluators and practitioners. A second, similar limitation is that this study does not include the viewpoints of the users or beneficiaries of PRVE programs. How to include users and beneficiaries in the evaluation process is a whole other subject that might be useful to explore in future studies.

A third limitation is that we often did not differentiate the roles of researchers and evaluators. Many of the people who do PRVE program evaluations are researchers, but that does not mean that these two roles should be confused. Professional researchers and professional evaluators follow very different kinds of logic, particularly in how they use results. The former use results mainly for purposes of scientific publication and academic discussion, but that is not necessarily the case for the latter. In this study, however, we did not draw this distinction, because our goal was to present the views of the people who had done the evaluations, whether or not these people were professional researchers. The issues specific to research on evaluation as such were secondary here.

A fourth limitation of this study concerns its geographic representativeness: all of our interviewees and focus-group participants came from Western countries, and Canadians were overrepresented, especially among the practitioners, because of the Canadian focus of the PREV-IMPACT project. However, the number of interviewees from other countries was large enough to make meaningful, useful comparisons, which was one of the objectives of this study. Moreover: although we defined our research framework from the outset so as to include Western interviewees only, our systematic review shows that a substantial share of all PRVE program evaluations worldwide have been conducted in non-Western regions. The usual precautions should therefore be taken if any attempts are made to extrapolate our research findings to non-Western settings.

Lastly, as in any qualitative study, the findings from this study are partly dependent on the subjective perceptions of the interviewees who contributed to it. Nevertheless, the number of interviewees was large enough to reveal significant trends in their perceptions of certain specific issues.

Box 1. What benefits does evaluations of programs for the prevention of radicalization and violent extremism (PRVE) provide?

According to our interviewees, evaluations of PRVE programs have many different kinds of benefits.

1. Benefits for PRVE practitioners

Program evaluations let practitioners obtain client feedback and think about the strategies and practices that will improve their knowledge and the services that they provide and make their work easier. Program evaluations also provide practitioners with some of the information that they need to meet policymakers' and funders' requirements in a field of practice that is difficult and constraining because of political pressures, the demand for results, and limited funding.

2. Benefits for PRVE program users (beneficiaries, recipients, clients)

Evaluating PRVE programs regularly serves as a form of quality control over the services provided to program users and thus ensures that these services improve their well-being. Program evaluations also give program users a voice by letting them share their experiences and suggest solutions by which the programs can better help them along their life paths.

3. Benefits for PRVE researchers and PRVE program evaluators

Evaluation of PRVE programs is still an emerging field, so the scientific literature on it remains limited, although it has been growing since 2016 (Madriaza *et al.*, 2022). The empirical data available in this literature is still too scarce to show what does and does not work in prevention programs and thereby better inform prevention practices. By conducting a growing number of evaluations in the field, evaluators are gathering relevant data both on the factors that may improve chances for success and on the risks associated with prevention efforts of this kind. Analysis of this data as well as production and transfer of knowledge to practice settings will make it possible to share the lessons learned and build the capacities both of evaluators and of practitioners, who are currently facing numerous challenges.

4. Benefits for PRVE program funders and policymakers

Funders and policymakers often use evaluation results to justify their decisions on funding new PRVE programs, as well as to guide their approach when developing PRVE policies and local, national and international PRVE strategies. Evaluations let funders and policymakers base their decisions on evidence and on programs that have demonstrated their value and will have a better impact on society.

5. Benefits for society

PRVE program evaluations also provide benefits for society as a whole. Whether or not evaluation results are conclusive, when made available to the general public, they let people keep abreast of the initiatives that are being taken to counter radicalization and violent extremism, as well as of the effectiveness of PRVE programs. Through this informative function, program evaluations contribute to a certain accountability and transparency in PRVE efforts, which are sometimes criticized for being closed and inaccessible, despite the major impact that radicalization and violent extremism have on society.

03

Obstacles and Facilitating Factors in PRVE Program Evaluations

Evaluating programs for prevention of radicalization and violent extremism is highly complex, for many reasons. For one, PRVE programs have come into existence only fairly recently. PRVE is an emerging discipline that deals with sensitive problems affecting many aspects of society, politics and psychology, as well as religion and national security. In addition, because PRVE is such a new field, empirical data on PRVE program evaluations are still scarce, which makes evaluators' work harder. They must apply their knowledge and skills as well as they can to meet this challenge and overcome the obstacles that they face in doing their evaluations. The present study has identified many such obstacles. Our findings concerning them are discussed in the following section.

1. OBSTACLES TO PRVE PROGRAM EVALUATIONS

Evaluating PRVE programs involves its fair share of challenges. Our interviews with researchers/evaluators and practitioners from Canada and other countries revealed the many obstacles that evaluations face. These obstacles vary in their nature and complexity. Some are methodological, while others have more to do with interpersonal relationships, management, finance or logistics. This section describes the obstacles that we identified in the course of this study, as well as the solutions that our interviewees proposed for overcoming them.

1.1 Organizational and logistical obstacles

The organizational and logistical obstacles that we identified related mainly to planning, management, operationalization and implementation of evaluation protocols. These obstacles consisted of elements and circumstances that prevented evaluations from proceeding as planned, for reasons such as problems with planning, or administrative delays, or lack of qualified staff. The seven main obstacles identified by the researchers/evaluators and practitioners whom we interviewed in this study were as follows: 1) failure to plan for evaluations as early as possible (ideally, while the PRVE program is still being designed, but no later than when it is first being implemented; 2) lack of understanding of what evaluation consists of and related problems in assigning roles; 3) lack of knowledge regarding program evaluation in general, and PRVE program evaluation in particular; 4) giving program evaluation lower priority and less importance than the organization's operational activities; 5) bureaucratic "red tape" involved in communicating with government

representatives and completing the forms required by certain funders; 6) time constraints that severely limit the opportunities to conduct evaluations and impair the quality of those that are done; and 7) high staff turnover in the organizations that deliver PRVE programs.

1.1.1 Failure to plan for program evaluation from the outset

Many of the researchers/evaluators and practitioners whom we questioned about the obstacles that they had encountered in evaluating PRVE programs said that having become involved in the project too late greatly complicated their evaluation efforts. One British researcher stated that having become involved too late was one of the main challenges that she had had to face.

> Yes. So that's the main challenge, being asked for an evaluation when basically the program's already finished or has already been implemented or already been going on for a year or a year and a half. [...] That has consequences for the quality of your evaluation, which I'm always very, very clear about from the outset. It means that you can't do a proper outcome evaluation; what you do then will almost certainly be a process evaluation.[2]

As this comment indicates, some organizations do not start making any plans to evaluate their programs until they are already under way or even nearing completion. As one Dutch researcher put it, "[...] the difficulty with evaluations is that often you're called [in to do them] halfway through [the program] or you have to do [them] afterwards [...]". Apparently, managers are more naturally inclined to recognize the need to evaluate programs once they are already under way.

Waiting until a program is operational before evaluating it does have some advantages. For one thing, it lets the organization devote its scarce human and financial resources to its operational activities. But waiting also has fairly sizable negative impacts on the quality of the evaluation that can be performed. In contrast, considering evaluation requirements while the program is still being designed provides an opportunity to clearly define initial assumptions regarding the program's underlying theories of change and its goals and objectives. Involving evaluators in the process too late limits the change indicators that they can measure and the time that they have to do their jobs and can also negatively affect the methodological design of the evaluation.

Box 2. Theories of change

Connell and Kubisch define the theory of change approach as "a systematic and cumulative study of the links between activities, outcomes, and contexts of the initiative" (1998, p. 2). This approach was designed to evaluate and accommodate the multi-level, multidimensional impacts of comprehensive interventions in which the task of linking actions to outcomes is extremely complex, at a time when existing evaluation approaches were considered inadequate or inappropriate (Sullivan and Stewart, 2006). This evaluation model is part of the theoretical approaches to evaluation and is based on the idea that the evaluator must help to identify the theory of action implicit in an intervention in order to define what should happen if the theory is correct (Sullivan *et al.*, 2006). Part of this task is to identify the indicators of short-, medium-, and long-term change that will let the evaluators determine what elements they need to form an evaluative judgment. The theory of change approach is helpful for improving program planning, facilitating decisions about evaluation methods, and reducing the difficulties of causal attribution that are often the bane of evaluations of interventions of this kind (Mackenzie et Blamey, 2005).

Source: Madriaza *et al.*, 2021, p.13

These obstacles are also reflected in the following comments by a Canadian practitioner:

> [...] definitely in one case, we just got brought in way too late in the program and so, there was only so much evaluation we could do, so a lot of our evaluation ultimately amounted to [...] just kind of picking up some of the assumptions that they've made around how the project was designed and some of the [...] questionable ideas that they had [...] underpinned the project with. So, in one case, [...] the whole project had been...the theory of change was related to making young Muslims feel more integrated into society... Well, okay, that's kinda dumb, though (both laugh)... that's not complete... we can take out more, like, principled and philosophical critiques—whether or not that's accurate—and then substantiate them with more information. But if you come in that late, you can't really correct that assumption at the outset. If you're brought in at the program design stage, and you're able to at least raise the concern that, "You know, that might not actually be a complete or particularly sound theory of change, it has a lot of cultural assumptions built into it," then you might have a fighting chance of actually evaluating the program in a more meaningful way [...].

---

2   In this report, all excerpts from interviews have been condensed and edited for clarity.

Lastly, if an evaluation is done too late, some potentially interesting data (for example, data on pre-measurements or on the first cohorts of program users) may be lost, and the program may end up relying on erroneous assumptions or erroneous theories of change.

1.1.2 Lack of understanding of the program evaluations purposes

One Canadian program director whom we interviewed said that it had been hard to develop the evaluation for that program, because first, the team had had to be educated as to what an evaluation consists of, how it can be useful, and what implications it would have for their work. Such lack of understanding is seen mainly when an organization has no in-house expertise in evaluation, but has such a limited budget that it must perform its evaluation in-house anyway, rather than engage external evaluators with the relevant experience.

> And, I think, as a team, I think getting everybody to understand this project and get on board with the way it works, things were difficult. But then, I think in the first couple of years, it was kind of like getting the machine to start working, so there was a lot of back and forth in terms of evaluation, who's doing what? Why is this important? What are we missing? All of that back and forth as a team, it took a while for us to kind of get into the habit. Now that we're in the habit, it has become a lot easier.

Understanding of the purposes of a particular evaluation may also be lacking when an outside evaluation team that is not familiar with an organization's field of practice is called on to observe and evaluate its activities. This situation may cause mistrust, tension, resistance and misunderstanding of the evaluators' role among members of the organization to be evaluated. As one Dutch researcher put it:

> Sure, well, I mean, innately with some people there was clearly some hesitancy to speak with an outside party. It's all very sensitive work, in general, probation work, especially with these types of clients, so they didn't want any outsiders looking over their shoulders, so to speak. I think it helped a little when they found out that we weren't journalists, and that we weren't there looking for a story, and that we were there to try and come up with something that was as objective as possible. But I think in general, [many people thought that] the word "evaluation" meant that they were going to be criticized.

In addition, one Canadian program director said that it had been complicated to define and assign the roles within the evaluation team so that every member's role matched their competencies and interests and they clearly understood what it was.

> Basically, I think that when the team started, there were a lot of gaps in terms of who's doing what, like who's taking care of the referral sheet, who's taking care of the intake sheet, who's taking care of the consent forms, assent forms, the M&Ms there are so many forms. Right?

This same manager also mentioned that within the evaluation team, there had been differences of opinion about how the program should be evaluated, what indicators should be used, what the evaluation design should be, and other issues. We will return to this subject in section 1.2.

1.1.3 Lack of theoretical and applied knowledge regarding PRVE program evaluation

Even when PRVE program staff embrace the idea of program evaluation in principle, they may not have any experience with it in practice. They do not necessarily know what factors to consider, how to operationalize indicators, and so on. Evaluation of PRVE programs is still an emerging field, so if staff are left to develop their evaluation protocols themselves, there will be very little in the way of resources and scientific literature for them to rely on, and the quality of their evaluations may suffer as a result. As one Canadian practitioner put it:

> [...] what I didn't see when I was in graduate school was much of anything related to evaluation of extremism and radicalization programming. [...] I think that's probably the most challenging thing: if you want to get into the field of evaluation, realistically, I don't know how much support you're going to find or how much literature and academic backup you're gonna have to do it, so that's probably a unique challenge, really doing a comprehensive look at what's available and what you can build your practice and career around, because I don't know, I'm not sure. It certainly didn't exist when I was in graduate school, but then again, that was a while ago.

Also, according to another Canadian practitioner, some members of the teams that conduct in-house evaluations (for example, researchers who work for program organizations) have only a theoretical knowledge of program evaluation and/or radicalization and violent extremism. This poses a problem when the time comes to conduct the actual evaluation in the field, where the evaluators may encounter realities that they did not consider when developing their protocol.

> [...] they had a scientist working on that team who was more used to working with theories—an anthropologist—but the document that we got for evaluation was still very broadly stated, so we really had to make it theory-based, the grounding of the theory had to actually be there...we had to make it very explicit.

1.1.4 Evaluation receiving lower priority than organization's operational activities

In some cases, evaluation is perceived as a lower priority than the operational activities through which organizations deliver their programs. The director of one Canadian program director said that its evaluation tended to be treated as secondary, or even forgotten, because it was a lower priority than working with the individuals and families who had been referred to the program. He also thought that his organization already had so much work on its hands just delivering the program—making sure that interventions proceeded smoothly, ensuring families' safety, providing users with legal assistance, making sure that people showed up for their court dates and complied with their probation conditions, and so on—that it had little time and few resources left for evaluating it.

> So... there have been multiple challenges, I think, one of them being that evaluation or documentation tends to be placed on the back burner of things where it doesn't necessarily seem that important, because there's so many overwhelmingly important things going on, like making sure that the groups are running, making sure that the families are safe, whether or not they're in safety focus, whether or not they're in regulation focus [...] and on top of that, like for a couple of our consultations, we also help them with legal issues, so making sure that their court dates are met and their probation is not broken. The list could honestly go on and on [...].

The fact that evaluation is often regarded as secondary to intervention might partly explain why, in many cases, evaluation is not planned from the start and evaluators are brought in after programs are already under way.

1.1.5 "Red tape" involved in the evaluation process

Within program organizations, evaluation places additional burdens on administrative processes that are already complex. One Canadian program manager mentioned the countless documents that funders required the program to fill out for evaluation purposes, as well as the length of these documents, which could take hours to complete. This manager suggested shortening and combining some documents to optimize the process, so that organizations would have less incentive to avoid evaluating their activities.

> [...] In one sense, I think, the documentation of the evaluation aspect of things that [funder's name] requires us to do... a lot of people had said that it was very extensive and very tedious to complete, right? I didn't believe it until I saw all those forms, and I went "Wow!" (both laugh), "that's a lot! ... Obviously, we spoke with [funder's name] about condensing them or making them shorter [...].

One researcher from Germany complained about the slow, cumbersome process required to communicate with and obtain responses from the justice system with regard to certain files.

> [...] Something that we really couldn't change but which was not a very good experience was the time it takes to request the data about... for the police record data, and the investigation data, and from the courts. And those took over a year and from a project-management point of view, this is very complicated if you wait one year to get the data you need. But there is nothing we can really do about it.

These administrative obstacles make work harder not only for evaluators but also for practitioners. Because evaluators have limited time to complete program evaluations, solutions should be found to improve communication with government bodies and reduce the number of forms that have to be filled out.

1.1.6 Time constraints

The evaluators whom we interviewed in this study felt that they were given too little time to conduct their program evaluations. Many of these evaluators said that having been pressed for time had inevitably affected the evaluation designs that they could choose and the quality of the evaluations that they could do. Tight timelines thus impose a significant additional challenge for evaluators. One Canadian practitioner felt that his evaluation team was given far too little time to conduct high-quality evaluations of the impact of programs. This practitioner was given only one year to evaluate a particular program, but to determine whether the program's users had successfully re-entered society after completing it, the evaluation would have had to continue for longer, so that post-intervention data could be collected and the program's impact could be properly assessed.

> I mean, it was quite clear that this was a program that was being rolled out, so we couldn't really have any results yet because for most of these people, they were sentenced to quite a long probation period as well, because of the nature of their offences. So we're talking about 5 years or so, so we wouldn't really know whether they were reintegrated or not for at least another 5, so I didn't really have a sense that we would have any kind of real results in terms of impact for this 1-year evaluation.

Some Finnish researchers added that the limited time available to them to evaluate a particular program had prevented them from recruiting more participants and making certain observations that would have been relevant to their research.

> Other barriers were time and resource constrained. I mean, it would have been better if we had done observation to actually see what happens in the sessions, but we didn't have enough time to do that. And with my data, I had trouble getting participants [...].

Another Canadian practitioner reported that lack of time limited his opportunities to verify his data and revise his evaluation protocols. As he described it, because of significant time constraints, he often had to develop evaluation protocols very quickly. But to ensure that they were really suited to what had to be evaluated and to the realities in the field, some verifications were also necessary. This meant that if he were doing things properly, he would be making adjustments iteratively after the initial data collection, to ensure that the method he was using was appropriate. Unfortunately, the timelines were too tight to allow for such verifications and adjustments, which may have reduced the quality of his evaluations.

> The other is timeline again [...] we wind up with sprints where we have to develop an initial version very, very quickly sometimes, because we have to build in room to iterate. Because otherwise, you wind up in a situation where you do the evaluation and you get the data back but you don't really have time to do revisions so you have to cross your fingers and hope that the evaluation will basically confirm that what you're doing already works because you don't have time to change it.

Lastly, one Dutch researcher stated that organizing and implementing his evaluation had taken so much time that he had been unable to recruit a control group, which he felt represented a major limitation of the evaluation.

> [...] So we did a bit of research and did find some people of the same age and with other similar demographic characteristics, but unfortunately, we were under a lot of time pressure and ... had our hands full organizing the rest of the study, so unfortunately we did not have a control group, and that is an important limitation.

All of these experts criticized the discrepancy between the limited time devoted to program evaluations and the amount of time that would be needed to really evaluate changes, especially among individuals who were in prison or on probation for long periods. Under these circumstances, it becomes hard to gather evidence-based data on the impact and effectiveness of PRVE programs over the medium and long terms.

1.1.7 Staff turnover

The seventh main organizational/logistical obstacle to PRVE program evaluations that our interviewees reported was high staff turnover both in the organizations whose programs are being evaluated and in their partner organizations. In the words of a Canadian practitioner with an organization that has many partners:

> [...] So staff come and go, and then we have to kind of restart the capacity-building in those organizations. So, this, the staff turnover in different organizations, right? Not just one organization, we are working with so many partners. So that's one challenge.

These organizations lose employees who are already familiar with and trained in the protocols and tools used for PRVE program evaluations, as well as in issues of radicalization and violent extremism. These organizations then have to hire new staff who must learn everything from zero, and to whom the evaluators must explain the evaluation protocols and tools all over again. This slows down the evaluation process, especially when the organization works with numerous other partners who have their own respective teams.

1.2 Methodological dilemmas and "paradigm wars"

When the time comes to design the evaluation of a PRVE program, many questions arise. Many issues must be considered to ensure that the design meets the identified needs and objectives effectively while taking the realities of the field into account. In addition to methodological issues, the evaluation designers must consider the budget allocated for the evaluation, the amount of time that the evaluators will be given to conduct it, the specific characteristics of the program, and issues specific to PRVE.

One of the main challenges in developing the evaluation protocol is to strike the right balance among the methodological recommendations in the scientific literature, the requirements that must be met to achieve the evaluation objectives, and what is actually possible given the realities in the field and of the program to be evaluated. One British researcher whom we interviewed raised the importance of proportionality in choosing an evaluation research design. In his opinion, it is essential for the protocol to accomplish the evaluation objectives in a proportional way, without trying to do too much more or too much less, and for the chosen approach to be suitable for meeting these objectives.

> I think that maybe one challenge we sometimes had was around proportionality. Like what is necessary for a certain project? How do you design the evaluation in a way that serves the purposes for which you're doing it in the first place? If you want to do a process evaluation, then you don't necessarily need to survey thousands of participants. It's more than you need, and once you've got it, you have to analyze it. So that's been a challenge, making sure that the overall approach in design, monitoring and evaluation is appropriate for what you're going to use it for.

As stated above, in choosing a methodology for evaluating a PRVE program, many factors must be considered, such as the budget, the local context, the nature of the program and its clients, and the purposes of the evaluation. But some of the decisions concerning the approach and the paradigm underlying the evaluation may be based on schools of thought that differ or even conflict with one another. The approach preferred by evaluation stakeholders and evaluation cultures may vary from one sector to another, leading to methodological dilemmas or "paradigm wars" that can cause tensions during the evaluation process. The following sections discuss the dilemmas that may arise in choosing the approach to follow in evaluating PRVE programs.

1.2.1 Impact evaluations versus process evaluations

Regarding the choice between impact evaluations and process evaluations, the opinions of the researchers/evaluators and practitioners whom we interviewed varied widely. In making this choice, many variables must be considered, and certain methodological compromises must be made. Ideally, the evaluators should have enough time and resources to conduct a process evaluation and an impact evaluation simultaneously. But in the real world, that is far from the case. One French researcher whom we interviewed recommended a mixed approach, for two reasons. First, it is not always possible to determine a program's effectiveness from its impact alone. Second, it is important to consider the experience and opinions of the field practitioners, ideally without getting in the way of their work:

> Today, I couldn't tell you whether I would do an impact report and whether or not that would be effective. So when they asked me what we should do, I explained that the current literature recommends mixed approaches. So in addition to the attitude questionnaires, I developed a protocol that let the practitioners participate as well, so that they could consider their own day-to-day practices and tell me what was implementable and what was not, in a realistic way, without interfering with their direct work with the program users.

However, most of the practitioners whom we interviewed recommended impact evaluations rather than process evaluations, possibly because impact evaluations are more quantitative and are perceived as "objective." But process evaluations are preferable when the evaluators are brought in late in the program or after it has ended, as it is often the case with PRVE programs. As one Dutch researcher described:

> Though that was also the case with the City of **X**. I didn't get asked till the program was already finished. So, all the practitioners had already been trained, and all the parents had already received sort of parenting classes, which seriously impeded me from doing a proper evaluation. So it was more like a process evaluation. [...] I couldn't evaluate the activities anymore, I could only talk to the people who had organized them, I could talk to the members of the community and to certain practitioners, which I did. I held focus groups and conducted many, many interviews to get an impression of what had happened, what have they had learned, what could go better and what could be improved.

Process evaluations should therefore be regarded mainly as an alternative to impact evaluations or a source of additional information in evaluations that take a mixed approach. Process evaluations are also considered valuable because they gather information that may facilitate the delivery of existing programs and the implementation of new ones. As one French researcher put it, "That's why I say that a process evaluation is important, because it lets you talk about psychosocial risks." Likewise, a Canadian practitioner stated:

> Specifically I think one thing that that would be helpful, from a practitioner perspective, and I think that there's maybe one project under way in Canada on this, but some sort of more projects around, you know, process or design evaluation. So I think we've all agreed that tracking outputs is likely a poor substitute for tracking outcomes. But no one is necessarily at the point where we can really effectively measure outcomes, particularly over a longer time horizon. So a good process evaluation might tell us how to build something that works effectively, which would obviously be optimal for all of us on the front line.

The above paragraph raises the point that while evaluators tend to prefer impact evaluations, there are still many questions about their value—in particular, about whether a program's impact can really be measured. Both researchers/evaluators and practitioners have some doubts about how realistic impact evaluations can actually be. One French researcher argued that realities in the field, in particular time and budget constraints, made the idea of evaluating impact illusory. "And besides, there is also the way that people want to measure the real impact, which at this stage is an impossible, illusory framework—it would require extensive work over a far longer period."

Much has been written about the methodological conflict between impact evaluations and process evaluations of PRVE programs. According to one German researcher, this conflict may cause tensions among the evaluators, and such tensions are also seen internationally: some countries seem to prefer one approach rather than the other. This same researcher also introduced the idea of a "paradigm war," which eloquently expresses the divergences in methodological opinions in the field of PRVE: "I see that there's a paradigm war around methods, and I think it's not so much about the methods as about the ideologies behind them, ideologies about how you should work with people and whether you should have impact evaluations at all."

He went on to discuss the situation that he has observed in Germany, where, according to him, the value of impact evaluations and quantitative methods using evidence-based data is widely questioned and disputed, although this debate occurs mainly in academia rather than in practice settings.

> And, then we also operate here in a climate which is very difficult, because in Germany there are many people who are very skeptical about impact evaluation, and in particular about the methods that we apply and use. [Our organization] has no dogmatic standpoint on any methods. We mix methods as it is pragmatic to do so. But there are also some skeptics who say that quantitative research is entirely useless in this area. And these actors, they have a strong voice, and so we operate within this climate of this, really, a paradigm war, although I had thought it had ended a long time ago, which is mainly waged in academia. [...] And there is a lot of skepticism about the evidence-based approach, and we try to do impact evaluations and get internal validity and to some extent external validity, and we try to achieve this as good as possible. But we do not say that the randomized control trial is the only method that can make statements about impact and causation. There are realist approaches and other things, but at the same time, people sometimes use the methodological divide as an argument for political issues.

Our interviewees also underscored the shortcomings of impact evaluations. First, it is hard to conduct an objective impact evaluation if the evaluators do not have access to the program users, but this access is very often limited, because of the users' vulnerability. Here one Dutch researcher describes how limited access to program users undermined the quality and objectivity of the impact evaluation that his team conducted:

> And lastly, we were also trying to do a limited impact assessment, and I say limited because we couldn't measure [this impact] in any objective way. And I think that's the biggest downside of the project methodologically, because we were only given access to speak with the probation service staff members and with people from the Dutch counterterrorism coordinator. So if we wanted to know whether people were actually reintegrating, all we had to go on was the certainly educated but still, you know, kind of, I wouldn't say guesswork, but the opinion of the program staff. We couldn't speak with the actual clients; we couldn't use a control group and stuff like that.

Also, even when access to program users is possible, the users who meet with the evaluators are often chosen by the organizations being evaluated, which can cause biases in the evaluation. To make its program look good, the organization might select those users who had made

the most progress, and not let the evaluators speak with users who had not done so well. The following comments by a Dutch researcher illustrate this phenomenon:

> I think just the main issue for us was the sample that we were dealing with, given that, of course they were selected by the prison environment, which makes it very hard for us to know like, okay, it could very well be the case that these people were selected because they all had a very positive experience. Because I think the NGO, again, I'll send you the summary, but they said that they had helped hundreds of people since they were first founded. Well, we interviewed, compared to the number that they had helped, only a percentage of like, maybe 20%. All of those beneficiaries, so yeah, it's very hard for us to know whether these people were selected because they were actually satisfied with their results.

This researcher also noted that it is hard to determine what parameters should be measured in order to make an impact evaluation objective.

> Well, I learned a lot about evaluations in general, and about the work that probation services do in general and how very difficult it is to come up with any kind of measures of objective impact and not. I think most of my lessons were in that regard, and also the stuff that we did, I mean, I liked it, and I learned a lot, but it is also in a sense very limited. That's what we told our financial sponsors as well. Look, we would love to do another 2-year evaluation, but I think we have reached a limit with this kind of method. If you want to do more, then you need to give access to clients, you need to do a more quantitative approach, you know, pre- and post- any particular intervention.

Another problem with impact evaluations is that they can be hard to complete in a short period or a single funding cycle. A British researcher described how his team had encountered this problem with an evaluation that they had done:

> The other big problem is that because of program funding cycles, we very rarely get to the point that we can actually prove the impact, the outcome. ... even when we might be able to show impact at the community level or prove that it wasn't just a contribution, that almost always happens after our funding has ended, so unless we're going to get a sort of Phase 2 funding that would let us re-evaluate something in another place, then that's just not feasible for us, which I think does again limit us, which is unfortunate. Explaining that to the funders has been a bit of a challenge.

Still another issue that researchers cited was that measuring a program's impact on a larger scale was complicated. They wondered how it might be possible to determine whether the program being evaluated had really had an effect on the community. One Canadian practitioner questioned how well the social impact of a program can be evaluated, given the statistical infrequency of crimes associated with radicalization and extremism and the difficulty of obtaining relevant databases:

> The area where I struggle evaluation-wise is tying more measurable tasks to larger social impacts, which is what we want to see. I think that part is obviously harder to evaluate, but I do think that our hope is when you develop an objective, it's something broad, and then we can break it into pieces that are more measurable but with achieving a larger objective.

> [...] it's hard to say whether those task-oriented evaluations connect with broader potential changes or impact on the prevention of violent extremism in **X**. Especially because we're looking at something that happens, statistically, quite infrequently, so it's hard to have a base to compare to.

Lastly, a Norwegian researcher's comments on a particular issue regarding the evaluation of multi-sector, multi-agency programs in Norway may shed some light on the issues of evaluating hub/situation-table programs in Canada. This program model makes it much harder to assess program impact, by making it almost impossible to determine the source and the cause of any change observed. It thus presents an additional challenge for impact-evaluation efforts.

> I'd like to say that because this Nordic multi-agency approach isn't... that they are not fixed programs or projects, they are a way of working, it's very flexible, a lot of variation, so it's very hard to sort of evaluate the program as such [...] you have a person with a problem, a youth with a problem, you can always throw a lot of different measures at him or her, and some of it will have an impact, and some will not. And in the end, you don't know which one will work. So, it's very hard to know... to sort out, separate or isolate the impact of one particular measure... because there's always a lot of other things going on, and it's against the logic of the Nordic welfare state to try to capture the effect of one measure on one individual at one point in time. We had some issues with that when we were going this multi-systemic therapy (MST) in Norway, and people would attempt to do an American-style evaluation where they said to all the other professionals, stay away from this guy

> for now; he's ours. That's totally against the Nordic logic that says "now we have to work together and do our best together," and maybe it turns out that it was one youth worker, or one psychologist, or one very empathetic schoolteacher who had the real impact, but they'll never know it, because there are so many.

The researchers/evaluators and practitioners interviewed for this study, while recognizing that process evaluations do provide a wealth of information, widely prefer impact evaluations instead. But high-quality impact evaluations take more time to do. They also require more resources, as well as access to program users, which can be especially hard to obtain. In addition, the evaluation activities must begin at the very start of the program, or even while it is still being designed, which seems impractical in many cases. Thus, realities in the field push evaluators to opt for mixed evaluation designs or force them to conduct process evaluations, for lack of any indicators or variables to measure. It is also hard to assess the impact of a program on a larger scale, such as a community, a city or a country. This raises the question of whether it is possible to really measure the impact that a PRVE program has had both on the program users and on the community.

1.2.2 Quantitative versus qualitative evaluation design

The decision whether to conduct an impact evaluation or a process evaluation also indirectly entails choosing between a quantitative and a qualitative research design. Every researcher or practitioner has their own preferred method for conducting evaluations, guided both by their theoretical orientation and by the circumstances in which these evaluations will take place. According to one researcher whom we interviewed, Germany seems to favour qualitative designs over quantitative ones. In contrast, many other evaluators strongly prefer quantitative designs, which they consider more objective because they are based on numerical measurements. But taking such measurements can sometimes be impossible because there is not enough time, or the evaluators are brought in too late, or appropriate indicators are hard to find, or access to program users is limited. Hence it is not unusual for evaluators to opt for mixed or qualitative designs instead. Each type of design has its own benefits but also its own limitations, which need to be considered.

There do seem to be some prejudices that qualitative designs are too subjective. Our interviews revealed these prejudices, as well as a debate about the quality of qualitative and quantitative designs. Are evaluations made using quantitative research designs really better than those made using qualitative or mixed designs?

Some of our interviewees asserted that when the right resources to do them are available, quantitative designs provide the best evaluations. The following two statements, the first from a French researcher and the second from a Canadian practitioner, exemplify this point of view.

> [...] But what would be the optimal conditions that would let a large research team with a wide range of skills and lots of money produce an evaluation that was really significantly better than what could be produced by a smaller team with a more qualitative methodology under clearly suboptimal conditions, but you do what you can when you have a limited sample and are dealing with government programs addressing sensitive problems that involve political and security issues [...].

> I mean we've had that in our project, we've done workshops and [...]there is a way, I suppose, that you could "evaluate" a workshop, you know, getting participant feedback and that kind of thing, but as compared to the large N evaluation that we are doing for the rest of the project, it's like apples and oranges. So, to a certain degree, we're not worried about evaluating the success of the workshop, I mean, it's obviously useful to say that they're useful, but compared to doing some sort of statistical analysis, it's not as important. So, I think every project probably has to make these decisions, so figuring out the things that are going to actually work and be useful in terms of finding effects is important.

According to these two interviewees, qualitative evaluations are of poorer quality because they use smaller samples and do not include numerical measurements and statistical analyses, and they should therefore be used only when the budget and other resources for an evaluation are limited.

On the other hand, some of our interviewees believe that qualitative designs are far more human and less mechanistic than quantitative ones. They argue that PRVE programs are basically focused on human beings, and that this fact should be reflected in their evaluation. One French researcher raised the idea of humanity in research:

> I'm not there to provide quantitative results, and I want to say that this smacks of a very capitalist mindset, this desire to produce something without worrying about the human side of professions that are essentially focused on human beings.

A Belgian researcher also expressed a certain skepticism about quantitative designs, just as another researcher, based in Germany, had observed earlier:

> So in general, in the political, administrative and professional culture of the French Community of Belgium, almost all of the actors have backgrounds in the humanities and social sciences—teachers, social workers, and so on. In particular in the French-speaking community, compared with the Flemish-speaking community, there is an extreme mistrust, a rejection of any attempt at objectification or formalization, any use of grids and indicators, even on this rather sensitive issue, because after, some people were saying yes, but we should conduct risk evaluations with risk scales, some people, are they radicalized or not—no, grids and all that, it's far too mechanical.

We did not, however, see this same tendency to favour qualitative designs among researchers and practitioners based in North America.

Also, one French practitioner offered an interesting caveat about the perceived objectivity of quantitative data. She said that even if the numbers are objective, they do not necessarily get interpreted objectively, "because you can really make the numbers say anything you want."

One German researcher suggested that evaluators' preferred research approaches depend on their profession and their training. For example, he said that evaluators with backgrounds in psychology are more familiar with statistical methods, whereas evaluators with backgrounds in social work lean toward more qualitative research designs. Their training instils a certain culture or attitude about evaluation.

> [...] there is a different culture of evaluation, let's say in forensics, when it comes to the treatment of sexual offenders, for example, where most of the practitioners have a background in psychology and have been educated in statistical methods. You have a different evaluation culture there than in an area where most of the practitioners are social workers who have been taught in university that quantitative methods are evil.

Lastly, evaluators seem to prefer qualitative designs and process evaluations when they lack time and resources but still want to try to supplement their qualitative data with quantitative data as much as possible.

> Ideally, ... you can do additional evaluation, such as interviews with participants and trainers, and many more things, and if possible, get quantitative data to supplement the qualitative data, but again, from my experience, this is very hard to do, and to do it

in time. It also requires some investment to start before the training starts ... We were lucky, but that's not always the case. Usually in an evaluation, you just jump in halfway, so that was one of the lessons learned, that doing evaluations is tough, tough work... really... yeah (laughs).

> I always prefer to combine quantitative and qualitative forms of data collection.

The choice of a quantitative or a qualitative design for a PRVE program evaluation will thus depend on how much time and money the evaluators have to do their job, what their theoretical orientation is and, potentially, what country or region they come from.

1.2.3 Internal versus external evaluations

Another methodological dilemma discussed by the researchers/evaluators and practitioners whom we interviewed was choosing between internal and external evaluations, which in these interviewees' minds seems to mean choosing between subjectivity and objectivity. They regard external evaluations as the "gold standard" and internal evaluations as less objective. The practitioners' organizations believe that from a methodological standpoint, to the extent possible, it is preferable to have their programs evaluated by external evaluators.

On the other hand, some of the interviewees regard internal evaluations as more "human," given the relationship that already exists between the evaluators and the people participating in the evaluation. But as one Canadian practitioner indicated, this can lead to certain biases:

> ... We do our best... it's hard to do internal evaluations from the perspective of non-bias, you know, we try to remove ... our sort of organizational pride, how we think about... how things should be, so that's the risk of doing internal evaluations, in terms of the oversight, the external oversight, so that's where we enjoy the external ones, because I think that's more of a pure way of doing things, but given our circumstances, challenging to do that at times [...].

A German researcher shared this view, especially regarding situations where the organization's funding depends on the results of the evaluation:

> [...] they are just project reports written by the programs themselves, there's no really evaluatory part of it, there is no checking, not of any quality, there's nothing negative. They usually don't find anything that can be improved, they just ... celebrate themselves and sell that as an evaluation. So, yes, this is problematic.

Some of the researchers whom we interviewed firmly opposed internal evaluations, on the grounds that they make the evaluators both the judge and a party to the case. According to one French researcher, even allowing an organization to choose the third party who will evaluate its programs constitutes a source of bias:

> The organization that was going to be evaluated had to find and sign a contract with an evaluator itself directly, which seems very strange to me as a researcher. I go, that can't be, it makes the evaluator both a judge and a party to the case. But sometimes that's how it's done, so then everything depends on how things are arranged by the institution. And when I've raised this issue with the prison administration or with the organization itself, they have not been especially shocked by this arrangement. They say that there are lots of arrangements like this, in particular for medical and social-service institutions.

Internal evaluations are also favoured when the evaluation budget is limited, because they are generally less costly. As one Canadian practitioner put it:

> Yeah, basically ... internal evaluations are really done, because, you know, we are mandated to. But number 2, most of the time, it's because of financial resources. When we have the opportunity to hire an evaluator externally, we do that, but that's not always the case. So that can be a barrier, because we're really a primary direct service organization, we're not a research organization and we don't have a research budget, so we've built internal processes. That's why we're strong in that area but it's not always best to evaluate your own program [...].

Indeed, as another Canadian practitioner explained, the cost of an external evaluation can sometimes exceed a program's entire budget for its operations: "So, like for the **X** piece, we had more investment going into the research and evaluation fees than into the actual programming, right? So... we're very careful about that."

Not all organizations have the knowledge required to conduct an internal evaluation. They are therefore encouraged to rely on the expertise of partners or consultants who can guide them through the process, help them to think about the evaluation, and add a touch of neutrality to it. As one Canadian practitioner described:

> [...] it seems that it's been helpful even to the main evaluator to have other like-minded folks to collaborate with, and it's not just one person trying to figure out: "OK, so how do I capture the nuance

> of this CVE space with this particular evaluation?" So having that expertise to rely on, to consult, and having more minds to collaboratively look at this evaluation, has been good so far.

The choice between an internal evaluation and an external one seems to depend largely on financial factors and the organization's capacities. The other key element seems to be the need to properly balance the objectivity and the subjectivity of the evaluation. Having an outside observer or mentor could contribute to this balance. For example, one British researcher told us that his organization had a third person validating the coding who was not involved in the program.

1.2.4 Longitudinal versus cross-sectional design

According to the practitioners and researchers/evaluators interviewed in this study, lack of time and financial resources is what determines whether an evaluation design will be longitudinal or cross-sectional. These interviewees agreed that there is a dire shortage of longitudinal data about preventing radicalization and violent extremism, and that more longitudinal evaluations would be helpful: "So I guess the trouble is that we just don't have enough longitudinal data on this." "So that would be good to have longer longitudinal data, to be able to collect longitudinal data; it's essential for prevention."

However, longitudinal evaluations require measurements to be taken at several different points in time (t0, t1, t2, etc.) and are therefore costly and complicated to carry out.

First of all, a longitudinal evaluation requires substantial resources simply to find program users again at various times in the future, which is often very hard complicated to do and hence poses a major challenge. As one Dutch researcher related:

> It was already a challenge finding these youngsters again after a couple of months. But we did actually reach most of them, which is also kind of part of the outcome, I think, that we can also talk to them three months later. [...] They were all over the place, and yes, you have to sit at home with them, you meet up at the coffee bar or somewhere; pretty challenging to find them again.

Decisions also have to be made about the optimal duration of a longitudinal evaluation and the times at which measurements should be taken in the course of it. This latter decision seems to get made arbitrarily, according to the total duration of the evaluation, rather than any evidence-based criteria.

> And I think that's part of that longitudinal data

problem: how do we really figure out what the experiences of those people who have gone through the entire program have been, and what does that really look like? [...] Some people report that they're still going through the radicalization process after 20 years. I mean, how do we really know when they're done, *per se*?

In reality, by default, cross-sectional evaluations are preferred in many cases, in particular for budgetary reasons. In longitudinal studies, the participants are followed for extended periods during which data is gathered at several times to observe the effect of the intervention over time. In contrast, cross-sectional studies are conducted over shorter periods and take measurements at a single point in time (Levin, 2006). So whereas a longitudinal evaluation lets the evaluators assess the effect of a PRVE program over time, a cross-sectional evaluation lets them observe that effect at one point in time only. Cross-sectional evaluations therefore make it far harder, if not impossible, to deduce causal relationships between a program intervention and its results among program users (Levin, 2006; Setia, 2016). Cross-sectional evaluations do, however, let researchers find correlations among the variables studied and measure the prevalence of these variables in the target population (Levin, 2006; Setia, 2016).

1.2.4 The control-group paradox

Many of the researchers/evaluators and practitioners whom we interviewed identified the lack of control groups as an important issue in PRVE program evaluations. Here a Dutch researcher explains how the lack of a control group, due mainly to a lack of time and resources, influenced the quality of an evaluation by his team.

Unfortunately, one of the very important shortcomings of the methodology is that we had no control group. We had only had an experimental group, so we could only look at the changes among those participants. There are several ways that you could get a control group. For example, you could get youth [who had participated in] similar of kinds of surveys or conduct interviews with people who were waiting to do the training in the future. Or you could do a bit of research on people of the same age and with similar demographic characteristics. But unfortunately, we were under a lot of time pressure and there was no possibility [of doing] that at the same time [...].

[...] A control group was not possible. The ideal randomized control trial intervention is not going to happen unless you have a large budget and unless it's really integrated into the training that you're planning to do. Then maybe it could happen, but I think that currently, it would be very difficult to do it in a way that meets high scientific standards.

The use of control groups also raises ethical issues. As one Canadian practitioner explained, it would be unethical to offer potentially beneficial resources to one group of at-risk individuals, but not to others:

[...] But the nature of it is that we can't really control for time that easily, because it just happens over time, and we can't really run experiments and have control groups over the same time period. One reason is ethical: ... we want to make sure that all at-risk users see the advertisement, we don't want to control the amount of it.

To overcome this obstacle, this practitioner opted for sequential treatment groups:

[...] We ended up doing more of a time-series analysis where rather than treating our audiences separately, we treated them in ABA sequence. In the A phase, we stuck to the normal approach. In the B phase, we made changes to enhance the approach in particular ways. And in the second A phase, which we're in now, at the very end, we're just back to normal? So that gave us enough statistical power during that B phase to tell whether we were seeing any significant change.

In addition, as the aforementioned Dutch researcher pointed out, realities in the field sometimes make it impossible for evaluators to establish a control group, so they must exercise creativity in adapting their protocols instead.

For researchers, it's always good to go out into the field and see how it's really going and to find out what the limitations of your methodology are. I think that was a very, very good lesson for us. You'd always like to do a randomized control trial, but if you do that in the field, you're out, so you have to be very creative [instead].

Box 3. Online evaluations of PRVE programs

Because the Internet can contribute to radicalization of individuals, PRVE programs are themselves making increased use of this technology. The issue of online evaluation of PRVE programs is all the more relevant because this is a new field of evaluation. But only one of the practitioners interviewed in this study offered an opinion about this evaluation modality. Given the growing importance of online technology in PRVE programs and hence in PRVE program evaluations, we considered it appropriate to dedicate a section of this report to this emerging issue. Although it is still in its infancy, it reflects the current state of the evaluation sector as well as the challenges that it will have to face in the near future. As one Canadian practitioner told us:

> So, one thing I would say as a general point is that evaluation online or of online data is just not very mature. There's a lot of evaluation of stuff that happens on the Internet in general, you know, like brand monitoring, and companies finding how many impressions and clicks and stuff like that that they have on their websites, but there's not a lot of work being done or enough work being done on online activity related to extremism… like how to evaluate it, and how to establish what it really means. There's a lot of hype, and a lot of numbers get thrown around, but there's not a lot of certainty around actual impact.

The first concern here is how to protect program users' privacy and confidentiality. The second is how to interpret data gathered online so as to measure changes in attitudes. Because online data is not a direct information source, they are hard to interpret with any certainty. As this same Canadian practitioner went on to say:

> So one of the things that people often want to be able to do with online work related to radicalization is to measure and evaluate behaviour change. They actually want to see people changing their behaviour, changing their perspective on the

online space as a consequence of a particular intervention or a particular treatment, but that's nearly impossible for a variety of reasons, one being privacy. A lot of the time, even if it would be technically possible to view people's activity online in such a way that you could understand something about their change in behaviour, the ethical limitations there and the limitations of privacy are quite high, particularly for people who go in and out of high-risk environments. So, there's that. The other thing is that online data is always mediated, right? It's always coming through the online environment, so it's always to review you from an oblique angle. The data itself is not a direct representation of that person, and so you have to be cautious about how much you interpret it and what your interpretation means, that sort of thing, because it's not coming at you straight ahead. I suppose even when you're sitting face to face with someone and talking to them, there are still biases and things like that that you need to be careful about.

Lastly, this practitioner pointed out that program users can be subject to various social pressures and want to project a certain image of themselves that is not necessarily genuine and could falsify the evaluation data.

> So, like I said, one of the challenges I've always found with online data is that there's a performative aspect to it. Often people are going on social media and saying things in order to be perceived in a particular way; they're trying to perform and create an image for themselves, and it may not be a particularly accurate picture of who they actually are. But that's an ongoing issue with online data.

To conclude this section on methodological dilemmas, whatever the evaluators' preferred methods may be, the budgetary and other resources available for evaluations within PRVE program organizations are what really determines which evaluation designs are used. Most of the researchers/evaluators and practitioners whom we interviewed seemed to think that the ideal evaluation would be an external impact evaluation or employ a mixture of quantitative methods supplemented by qualitative interviews. It would also be longitudinal and include a control group. Because such a design is somewhat utopian, evaluators are forced to choose methods that do let them evaluate the parameters that interest them but also entail some compromises, notably in terms of objectivity.

1.3 Disagreements over what indicators and parameters to measure

As discussed above, agreeing on what approach to take and what roles everyone will play in evaluating PRVE programs can be difficult within a single organization, never mind across an entire country. One Canadian PRVE program manager told us that these issues had caused so much debate on his team that their evaluation had taken longer to carry out. There is also much debate about what factors should be evaluated, because the evaluators want to make sure to measure everything that might prove important and relevant.

Another central question is how evaluators should quantify the changes that they observe, which assumes that someone has already decided what changes the PRVE program will try to achieve. And answering this latter question requires the construction of a theory of change that is appropriate to the type of program and activities planned. In practice, however, PRVE programs are often designed without any theory of change, and the evaluators have to reconstruct the programs' implicit theories of change after the fact. The need to do so not only can generate discord within the evaluation team but also raises an important methodological concern, because the theory of change influences the evaluation approach that is chosen and the quality of the evaluation that results. It is therefore important for PRVE program organizations to have access to program evaluation specialists who can help them to choose evaluation methods that support their evaluation objectives. (Even if the evaluation is going to be conducted in-house, access to a consulting service can be very helpful.)

But agreeing on a methodological approach for the evaluation protocol is only the first part of designing an evaluation. Regardless of what approach is chosen, a decision must also be made about what indicators and parameters will be measured in the evaluation. This decision too can be controversial and entail its share of challenges, as discussed in the following sections.

1.3.1 Translating vague program objectives into clear, specific ones

In order to design an effective PRVE program evaluation, the program's objectives must be specified. For most of the researchers/evaluators and practitioners whom we interviewed, regardless of what country they came from, this was no simple matter.

Many of these interviewees said that the objectives of the programs that they were asked to evaluate were often too vague, and that one of their main challenges was to translate these vague objectives into specific, measurable indicators that would let them evaluate these programs. One Dutch researcher described this challenge vividly:

> Something we already spoke about is the lack of indicators. Someone says "**X**, can you come and evaluate our program?" And then you look at the program and find out that there are no indicators, only this broadly formulated aim of preventing and countering violent extremism. And so [I end up having] to formulate indicators [myself so that I can] know what I'm going to research. So, that's how I overcome it, by formulating indicators myself: as you know, that's not what a good program should abide by or look like.

If the evaluators were not involved in conceptualizing, designing and implementing the PRVE program, then to define the indicators for evaluating it, they will, of course, have to collaborate with the people who did that original work. One U.S. researcher described this process as follows:

> So before developing the survey, we did … a literature review trying to identify existing survey instruments that were measuring similar constructs or somehow aligned with what this campaign was focusing on. So there was a lot of discussion with the campaign developers to figure out what they were trying to achieve. You know, everybody is going to tell you "we don't want kids to hate each other, we want kids to be nice to each other," but that's not measurable.

Team members may disagree about which specific indicators should be used in an evaluation, so many discussions may be needed to reach a consensus.

1.3.2 Choosing and formulating indicators: deciding what to measure, and how

Choosing and formulating indicators also seems to be a real headache for organizations, because so many factors must be considered in making this choice. One Dutch researcher deplored the lack of specificity in PRVE programs' theories and indicators of change. She pointed out the difficulty of measuring "prevention of radicalization," and said that the indicators should be more concrete:

> ... I saw a problem: they have a tendency not to formulate proper, smart outcome indicators. So yeah, "the goal of the **X** program was to prevent and counter radicalization and to create more resilience," but how do you measure that, how can you ... create this theory of change: that training practitioners such as police officers youth workers and school teachers contributes to preventing and countering violent extremism?

Many of the researchers preferred to use indicators that they considered objective. For example, one British researcher talked about measuring the knowledge that program users had acquired, rather than their confidence in that knowledge, which is more subjective.

> We have been increasingly moving away from using sort of softer subjective indicators, such as knowledge-confidence, which we used a lot initially. So yeah, harder things, like actual knowledge questions or trying to use more established measures for attitudes rather than the bespoke things that we were using in the past.

In any case, these indicators are not always adequate to measure PRVE program outcomes. Sometimes the chosen indicators turn out to be inappropriate for detecting the changes that the evaluators are looking for, so they have to go back and redefine them, which costs additional time, money and resources. A researcher from the U.S. described one such experience:

> [...] We haven't seen a change in behaviour, which could mean two things: either the intervention is not working to change behaviour, or our infrastructure is not sophisticated enough to detect a change. So now we're working on the index and also on the training to figure it out; I think it's a measurement problem. I think that the index does not capture exactly what the training is supposed to change, the kind of behaviour that the training is supposed to change.

Indeed, the process of choosing indicators often takes a great deal of time, as one Canadian practitioner described:

> In building evaluation metrics, we initially struggled with what we were trying to evaluate—it was our big question. And we have answered it through our six-month round of public engagement and now our awareness session.

Questions arise about how to quantify a PRVE program's success, and more particularly its larger-scale effects. For example, how does one quantify the impact that a program had on a community? What indicators should one use to determine whether a program does effectively reduce crime or recidivism? Everyone seems to have their own views about what things to measure, and there does not seem to be any consensus on this issue. One Canadian practitioner described how this issue arose in one program evaluation:

> So we had done a lot of groundwork for the evaluation, and that's why it was a bit difficult in the beginning to come up with the right indicators. And we had the conceptual issue that the [program] goal was to increase social and physical assets in those particular networks, so the question was, how could we set up some tangible targets so that we could measure the increase in social capital? What formula would let us turn whatever data we collected into some kind of tangible numbers so that we could set that target and measure it? For example, we could make the target a 10% increase in the social and physical capital, but how were we going to measure that 10%? So that was the challenge, but when it comes to this new project, I'm thinking that the measurement challenge is the data around hate crime and extremism. Currently, there's no standard way of collecting that data because the legislation is vague, and all those pieces, right? So the challenge this time is how we'd measure the progress... we suppose there has been a reduction in hate-motivated crimes in the city, because if you go by the crime statistics, different agencies... different organizations measure differently, so the question for us is how we'd measure the change in hate-motivated crimes and also the change in behaviour and perceptions out there in the community.

Another issue concerns the use of certain indicators to measure PRVE programs' success. For example, recidivism rates are often used to measure such success, but according to one German researcher, they are not really suited to this purpose:

> Of course, 95% or 99% of evaluators in the field look for impact, but in fact they look for recidivism rates, they look for case numbers, they look for any kind of metrics. But the longer I have been working in the field, the more I have moved away from these metrics. They do not tell me anything about the quality of the program. I have seen it myself that many of these metrics can be and are falsified.
>
> They want to have quick results, they want to have something to show for it, and they are interested in hard-core impact metrics, especially recidivism rates, even though recidivism rates make no sense at all. Currently, in the deradicalization field, they are not moving away from it. And the public demands it, and policymakers demand it. They want recidivism rates, even though they know, for example, in the U.K., if you are following the discussion, there are hundreds of former terrorists who had been released, and six of them have been re-arrested or have recommitted any terrorist crimes. That's a recidivism rate of about 3%, which is phenomenal for any kind of disengagement or deradicalization program, but it's still enough to almost or completely shut down a whole deradicalization program in the U.K.

Evaluation of PRVE programs is still an emerging field, and in the absence of evidence-based data, such controversies are hard to resolve. Faced with complex, rapidly changing circumstances, evaluators have no choice but to proceed on the basis of what has been done before and has rarely been measured in the field, and on what has been done in related fields, and to learn from their mistakes so as not to repeat them in future. In this way, some guidelines for evaluating PRVE programs are developing very gradually.

1.3.3 Causality

Many of the researchers/evaluators and practitioners whom we interviewed had asked themselves how they could attribute observed successes to a PRVE program or program activity when they might have resulted from external factors instead. This issue is especially relevant to programs that apply a hub or situation-table model, but arises in more traditional programs too. One German researcher nicely summarized this issue of causality as follows:

> So if a [program participant] changes certain behaviour … , the program claims success or claims causality, but they have no idea whether it was actually they who were causing the change or having the positive influence or whether it was the person's mother or their friend or whatever. So it's really difficult to establish causality between your

> program's intervention and a change in one of your clients.

Thus, many factors may explain observed changes, and causal links between an intervention and effects are hard to establish, because the vast majority of interventions and evaluations do not take place in controlled environments. Also, some researchers have their doubts about the truthfulness of the information gathered from program users:

> I think the downside is, just as you might argue in the case of other kinds of criminals, it's not easy to detect whether or not they're lying to you as a researcher, and I found that very problematic. […] I have basically never experienced this level of uncertainty as to whether I can trust what someone says.

To conclude this section, the following excerpt from an interview with a Canadian practitioner summarizes the issues raised:

> […] People kinda know that what they're doing can't really be meaningfully evaluated, but they feel like that there's no… not no risk but like… the risk of inaction is greater than the risk of doing something that can't be evaluated. I think that's often how people look at it. And I think there's some truth to that. … But I think more broadly that just getting [your] theories of change straight and getting a clear picture of what you can actually expect from the intervention work and what can be meaningfully measured, I think that helps. And it seems obvious, but it certainly helps in terms of program design … If you can say, "Well, yeah, I want to do these 15 things, but only these three of them can be meaningfully evaluated", then maybe you can still do all 15, but those three can kind of have a special priority in the project, because they're are the ones that you'll be able to meaningfully evaluate.

The ideal evaluation does not exist. But some people think that the risk of doing nothing may be far greater than the risk of taking some steps to try to prevent radicalization and violent extremism, even if some of the factors involved cannot be evaluated. Not everyone shares this view; some people believe that when it comes to such prevention efforts, in the absence of evidence-based data, the risk of taking actions that are ineffective or even counter-productive is just as real. One of the best documented examples in the literature involves programs that, by targeting Muslim communities in particular, have contributed to stigmatizing them and may thus in some cases not only have failed to prevent radicalization, but

also have aggravated the situation and the tensions with these communities.

1.4 Obstacles related to funding and funders

One of the largest obstacles to PRVE program evaluations that our interviewees identified concerned funding. The two preceding sections showed the close connection between budget limitations and limitations on the methodological options for evaluation designs. Instead of being able to base their evaluation protocols on the methods that they consider most appropriate, evaluators are forced to make compromises to stay within the tight budgets allocated for evaluations. These compromises entail certain sacrifices in the quality of the evaluations. One Canadian practitioner explained that only a tiny fraction of their program's budget is earmarked for evaluating it, so that they really cannot do a proper evaluation, even though they have to do some kind of evaluation in order to receive funding and preserve their accreditation:

> [...] So we're not a research organization. We get money for direct services and then maybe 3% or 5% of the budget for evaluation, so if we get $100,000, that will be like $5000 [...] Yeah, it's funny, because they never give enough money for an evaluation, but they make it a requirement for accreditation as a health centre. Also, with all of our funding they want some outcomes, but it's always a battle, because it is like, here we go again, for $5000, I don't know what we can do [...].

A British researcher underscored how hard it was for his team to conduct high-quality evaluations that met funders' requirements, because the resources needed to do so were not provided:

> [...] Depending on the funder, it means that we don't always get enough time and budget to do the sort of monitoring and evaluation that we could or would ideally like to do, and obviously we also have to meet the funder's stated requirements for the evaluation. We try to do the best possible with our funding [...] We can't do things without the resources.

Many of the interviewees also mentioned that certain funders seemed to have unrealistic expectations about evaluations, which might be due to a disconnect between these funders and the realities in the field. Some evaluators reported that funders tended to want to see immediate, broad-based results, whereas the program interventions were carried out within the community gradually would take time to produce any observable large-scale changes. Hence it makes sense that program evaluations do not reflect such changes immediately,

especially since most evaluations are conducted over short periods. This places the program managers in a delicate position, as described in the following two statements, the first from a Canadian practitioner, and the second from a German researcher.

> It's much longer, you know ... it's long-term work, and unfortunately, even with the evaluations, what our funders are looking for is immediate impact, so how do you give funders what they're looking for as far as, you know, that this program is having an impact but it's not life-altering.

> I've been doing this kind of work for over 10 years now, and I haven't seen any significant developments in this field. Most policymakers are not interested in evaluating or improving the quality of the program. They want quick results, they want something to show for it, and they're interested in hard-core impact metrics [...].

A British researcher related the following experience that she had had with a government funding source, which reflected the observed disconnect between PRVE practitioners and realities in the field, on the one hand, and government and non-government funding agencies on the other.

> Let me tell you one thing, one particular government that I will not name ... "Very good **X**, can you do that in three hours and can you change their brains permanently?" And that's not only not possible, but it also wouldn't be ethical [...] I think the mindset in government is that there is one thing called radicalization, and that you can reverse it with some other thing, but there is no such thing... It's not that these people aren't intelligent; they're profoundly intelligent, but this is not their field.

It would therefore be helpful to find a way of adjusting funders' expectations about the programs that they are funding, or at the very least, to engage them in a conversation on this subject. But the position of authority that funders enjoy gives them a degree of power both over programs and over evaluators, which sometimes leads to the kinds of abuses that we discuss in the section of this report that deals with ethical issues. Thus, as one Belgian researcher suggested, the desire to see their funding renewed sometimes prevents organizations and evaluators from resisting funders' demands:

> So there's no direct pressure, but it's a small world, and this is true for everyone: when someone is awarding your contracts and providing your funding, are you going to take the risk of disagreeing with them?

Some evaluators also reported that they felt pressure, both from funders and from program organizations, to make positive findings in their evaluations. Funders tend to fund those programs whose results are perceived as the most promising and to want the programs that they have already funded to perform well. This places the evaluators in a delicate position, as one Finnish researcher described:

> [...] But it's hard... the program is in a very sensitive position: it has specific funds, and it has to perform well to keep those funds. You have to be honest about what results the program is getting, but you don't want those results to lead to its fundings being taken away [...].

Because any rigorous evaluation will usually find at least some things that are negative or that could be improved, the frequent dependence of funding on evaluation findings can drive some organizations to conduct evaluations that are biased or of poor quality. (This issue too will be discussed in the section on ethical concerns.) One German researcher suggested that to correct this situation, policymakers should fund organizations that provide high-quality evaluations: "Theoretically, the policymakers, where the money comes from, the politicians and the ministries, should tie the funding to high-quality evaluations."

But that would create a vicious circle, in the sense that, as discussed in the section on methodological obstacles, the quality of evaluations depends very heavily on the budgets available to organizations. Now, in this precise situation, the obtaining of funding would depend on the methodological quality. Thus, for this proposal to work, the initial budget provided to the organizations would have to be better calibrated to enable quality evaluations to be carried out.

On the other hand, one German researcher stated that certain organizations were sure of securing funding from their funders regardless of the results of their evaluations or even the quality of their programs:

> And I personally discussed it with some practitioners, and they simply laughed at the idea of being evaluated, because they knew they would get the money anyway. They get the money handed out, right? So they don't see the need to engage in any form of evaluation. They don't see the need. And they get money, they get more money than they could actually spend, they get calls from ministries saying "Just hand in any kind of application with your name on it, and we'll give you money. It doesn't even matter what you do with it; we need to spend the money!"

This certainty of secure funding raises other issues, in particular regarding control of the quality of the services provided by the funded organizations. To a certain extent, from the standpoint of implementing and evaluating a program, being sure that one's funding will be renewed can be just as harmful as not receiving enough funding.

2. FACILITATING FACTORS IN PRVE PROGRAM EVALUATIONS

On the basis of their experience both in delivering and in evaluating PRVE programs, the researchers/evaluators and practitioners whom we interviewed also identified a number of factors that facilitate such evaluations. In conducting PRVE program evaluations, these individuals had learned a wide range of lessons that provide some insight into how PRVE program evaluation practices might be improved both in Canada and internationally.

The following sections discuss these facilitating factors and these lessons. Facilitating factors associated with collaboration, such as relational facilitating factors, will be discussed in the section on collaboration issues.

2.1 Organizational and logistical facilitating factors

Some facilitating factors concern aspects of the evaluation that should be considered at an early stage in the process. For such factors to be effective, they must be put in place and used by the program organization when the evaluation is being planned, or even when the program itself is being designed. There are many such factors. They depend largely on the administration, management and co-ordination of the programs and of their evaluators, when these evaluations are performed by an outside team.

> Start building a well-designed program first, before you do anything else—before you work with any clients or start talking about recidivism and impact rates. Your program is only as good as the effort you put into designing it, and the better and more detailed your structural integrity, the easier it will be in the end to interpret any kind of impact data or metrics. ... And your day-to-day work will be much easier if your staff is well trained, if you have risk assessment, if you have emergency protocols, if you have standard intake procedures—all of that just makes it a damn sight easier to do that kind of work and to do it more effectively.

Because of their strong logistical and organizational component, we have grouped these facilitating factors into a single category and discuss them in the following section.

2.1.1 Planning for evaluation as soon as program design or implementation begins

One facilitating step that all of our interviewees agreed on was to start planning the evaluation of a PRVE program as soon as you start designing the program, or at the latest, as soon as you start delivering it. These researchers/evaluators and practitioners from Canada and other countries reported that, in their experience, the evaluations that had gone the best were those in which the evaluators had been involved very early in the process and the evaluation protocol and/or the evaluation indicators to be measured had been defined either before the program became operational or at its very start. This experience was reflected in the following recommendation by a researcher who works in the Netherlands: "My recommendation is to include researchers during the policy-design process, or at least before implementation, instead of after the project has already been implemented or is even finished."

When they plan the evaluation at the very start of a program, the program evaluators, practitioners and managers have more time to design a protocol that will meet the specific needs of the evaluation, that will be sounder methodologically, and that will satisfy the organization's time and budget constraints. Such planning also gives the evaluators enough time to revise their protocol if necessary to adjust their methodology and the change indicators that they will be measuring. Involving the evaluators at an early stage opens more doors from a methodological standpoint. For example, it lets evaluators take pre- and post-measurements and establish control groups. The researchers whom we interviewed who had been involved from the very start also said that their theoretical knowledge regarding radicalization, violent extremism and terrorism was greatly appreciated by the organizations and had helped them to build their programs on evidence-based theoretical frameworks. This also provides an opportunity to exchange knowledge on the subject, which can be very helpful, because some practitioners, by their own admission, do not have knowledge in the specific field of PRVE. Here is what one Canadian researcher had to say on this score:

> Being in there from the start was incredibly useful. And they were very happy to have me there, because I also brought the academic kind of literature background. They know their day-to-day business, but they're not steeped in the broader history or broader context of terrorism.

It is also important to determine the theory of change underlying a program's interventions, along with its specific objectives, at an early stage in the planning process. This includes, for example, the changes that the organization wants to see in its clients. Ideally, the evaluators should be involved at this stage, so that they can provide the organization with support and guidance in choosing theories and objectives that are clear and compatible with the evaluation. The evaluators can also help the organization to develop change indicators that are both measurable and appropriate to the specific field of PRVE. This step lays the groundwork for the evaluators and greatly facilitates the evaluation when the time comes to carry it out, as the following two interviewees explained:

> [...] right from the beginning of the program, I got them, they were very open to it, but I was the voice in the room always saying "What are we going to keep track of, what are we going to monitor, what are our evaluations going to look like, what are our outcomes?" Because when you talk to the **X** or the police, their metric for success, of course, is rearrests, right? Or recidivism, and obviously that's part of it. I didn't mention it in our evaluation, but looking at it, checking to see the follow-up, and what kind of further contact they had with the criminal justice system is an element of this. But right from the beginning I was able to say, that's fine, but there's got to be more than that. It has to be, in a subjective way, did we help this person? Did we make things better for them? Did we, at least in the short term, make or appear to make violent extremism a less popular or less alluring choice?

> That's also a goal, to increase young people's resilience, but that's way too broad, so to develop the questionnaire, I also have to develop a theory of change and indicators of how we're going to measure it. [...] So tomorrow I'm going to be interviewing the person who designed the program (well, in her head, not on paper) and who also delivers it in all the classrooms. I'm going to ask her what her idea is behind the program and what she's aiming for, and to tell me in concrete detail, so that I can formulate indicators.

Another advantage of planning the evaluation at an early stage of designing the program is that the evaluators can then compile some basic data, which becomes harder, if not impossible, if the evaluation is not planned or carried out until the program is already under way or has already been completed. Planning early gives the evaluators a wider range of options.

Both the researchers/evaluators and the practitioners whom we interviewed mentioned two financial factors that facilitate evaluations: 1) receiving an initial amount from the funder that is sufficient to cover the evaluation expenses, and 2) forecast the expenses and budget for evaluation accurately enough to avoid having to redirect funds to this purpose later on. When organizations and evaluators receive enough funding to cover their evaluation expenses, they no longer need to worry about having to choose between keeping their staff and continuing their program activities on the one hand and doing a high-quality evaluation on the other. Such compromises can also be avoided by planning the expenses in advance, but not all organizations are in a position to do so. The overwhelming majority of our interviewees said that they received very limited funding overall and that the portion of their budget allocated to evaluation was even more limited, even though their funders often made evaluation a requirement.

A few of our researchers/evaluators and practitioners raised the idea of an evaluation culture and said that developing such a culture from the start of a program can facilitate evaluation later on. They suggested that to develop such a culture within an organization, program managers, as soon as they form their teams, must clearly explain to them the importance and relevance of program evaluation and the concrete impact that it will have on their work. These researchers/evaluators and practitioners also suggested that managers take practical steps to show their staff how things that they do from day to day (such as filling out a form after every meeting) can contribute to the evaluation of their program. To conduct a high-quality evaluation, the evaluation team, whether it is in-house or comes from academia, must be grounded in an evaluation culture:

> For good evaluations to take place, we first need to work on the mindset, on creating a culture of evaluation and on implementing a monitoring and evaluation structure. Only then can we, as an academic community, start evaluating all these programs in what we can consider a proper, thorough way.

One Canadian practitioner felt that it was a lack of financial resources that had driven his organization to develop an evaluation culture. With funding in short supply, all members of his staff had had to pitch in and apply their skills and knowledge as well as they could to contribute to the evaluation process. Although this evaluation culture had arisen in response to an obstacle (the lack of funding), it turned out to be a facilitating factor for his team.

> Yeah, it's funny, because they never give enough money for an evaluation, but they make it a requirement for accreditation as a health centre.

> Also, with all of our funding they want some outcomes, but it's always a battle, because it is like, here we go again, for $5000, I don't know what we can do. So I think that constraint has been a plus in the end, because it has forced us to build an evaluation culture to which everybody has had to contribute [...].

2.1.2 Identifying potential partners at the outset and forming a team according to the needs identified for the evaluation

Another organizational and logistical step that can facilitate the program-evaluation process is to identify key people who can help this work go smoothly. The researchers/evaluators and practitioners whom we interviewed mentioned the importance of developing a network of outside partners on whom they could rely for certain expertise and services during the evaluation, and of building an evaluation team whose members had the right skills to meet the evaluation's various requirements.

Developing a network of outside partners

Many of the researchers/evaluators and practitioners whom we interviewed said that identifying partners who can help the team and defining the terms of the partnerships with them early on seem to facilitate the evaluation process. Such partners usually have expertise or experience from which the evaluators can benefit. Domestic and international partners with experience in program evaluation seem to have been of great help to the evaluators whom we interviewed for this study. For example, when asked what factors had facilitated the evaluation of his program, one Canadian practitioner immediately mentioned the ties that his team had established with two organizations that had expertise in program evaluation:

> I think certainly our partnerships with **X** and **Y,** who are always eager to provide support—either to partner with us on a project or be a consultant. That's been an amazing strength, and we've learnt a lot from them, because that's what they do, so it's always exciting when they're working with us on a particular piece. That's been a real strength. I think that we have a nice culture and that people are interested and engaged and want to do it, but those two things are most critical.

A Norwegian researcher also mentioned the crucial importance of the expertise and assistance of local partners, in this case to lend a hand with evaluation efforts in the field:

> It's a necessity. We couldn't have done it from my institution alone; it would have been impossible. So we need local partners, and we've been

fortunate to have very competent ones. And also partners with different skills. I know a little about how to do surveys, for example, but we have some very competent people for that... so, we have people with a lot of different competencies and contacts, which facilitates collaboration among the stakeholders and helps the evaluation run smoothly.

In addition to having local partners, building a network of international partners has also proven to facilitate evaluators' work, particularly in Canada. According to two Canadian practitioners, because some European PRVE programs have been around longer, they can prove valuable as international partners. Because they have more relevant experience and data, Canadian teams can talk with them about the lessons that they have learned and learn from their errors while adapting their knowledge to the Canadian context:

Not only are we doing it across Canada, but we're trying to connect with international partners and see what they're doing and what's working for them. Because ... it's unfortunate, but in Europe, in Sweden and some places, they're much further along in the development of these programs, so trying to learn lessons from them and doing things like that as well are helpful for my evaluation strategies.

It's pretty difficult to evaluate them, because you don't have enough examples to get a good idea of what's going on. Maybe that's when you need to go to some international partners and find out what worked for them, to see how they evaluated what they were doing, because maybe they have more numbers. Go to Belgium, or France or England and see how they're evaluating. I know there are other issues in those places, but we just don't have the numbers.

Because PRVE programs are very new in Canada, it makes complete sense that Canadian evaluators may need to turn to and sometimes even depend on a network of domestic and international partners who can support their evaluation efforts and share experiences with them. This explains why a larger number of our Canadian interviewees mentioned the need for partnerships to conduct their evaluations.

In addition to building a network of local, national and international partners, one Canadian practitioner mentioned the importance of having these partners come from a variety of backgrounds. PRVE is a multi-dimensional discipline and requires many different kinds of expertise to properly understand the issues that it involves. For example, one practitioner mentioned how important it had been for his organization to work with

law-enforcement partners when it first started designing its program:

From the early days before the program was even running, when we were still putting together the application for funding, **X** had higher-ranking people on our steering committee. So, right from the start, they've been involved and known what's going on, and I think that was a really good decision on the part of the program to include law enforcement partners right from the beginning, because they feel like they have a stake in it.

Thus organizations can facilitate the evaluation of their PRVE programs by working with partners who have more experience in evaluating such programs, have a variety of expertise regarding the phenomena concerned, and are prepared to help with the evaluation process. In Canada in particular, where PRVE programs are relatively new, such expertise may be lacking locally, so organizations and evaluators may sometimes want to reach out internationally, and especially to Europe, to build partnerships and discuss evaluation practices in this specific field.

Building a competent, structured, diverse evaluation team

According to our interviewees, having an in-house or external evaluation team with the experience, knowledge and skills needed to conduct a program evaluation properly considerably facilitates the evaluation process. Our interviewees told us that having team members who had conducted or participated in program evaluations before, even if the programs were unrelated to PRVE, was a big help from a logistical standpoint. For example, these team members had a clear idea of what was needed to conduct the evaluation and how to organize the process. Experienced people like these serve as guides and resources for the rest of the team, especially when the evaluation is conducted in-house (external evaluators usually already have the relevant expertise).

Clearly define the roles of the members of the evaluation team

To avoid the confusion that can arise when mandates are unclear and roles and responsibilities are not well understood, it is important to clearly define the roles and tasks of every member of the evaluation team. Being familiar with the roles, tasks and mandates of the other team members also facilitates the process and prevents misunderstandings. These roles must be defined not only for the team that will conduct the evaluation, but also for the members of the organization whose program is being evaluated (such as managers, program co-ordinators and practitioners), if they are going to contribute to the evaluation. One Canadian practitioner explained how the work can be organized more effectively when everyone involved knows what they are responsible for:

> Essentially, I think that when the team started, there were a lot of gaps in terms of who's doing what, like who's taking care of the referral sheet, who's taking care of the intake sheet, who's taking care of the consent forms, the assent forms, the M&Ms—there are so many forms, right? So, when I talk about policy and procedure, an example would be making sure that the supervisor knows that he or she is going to fill out the referral sheet, that the youth outreach workers complete the intake, consent and assent forms, that the committees do the M&Ms, that the counsellors make observations while they're counselling and things like that. So, just breaking the evaluation down into steps, assigning the various steps to the various members of the team, and making sure that everybody knows who is responsible for each step and understands their own role.

This clarification of roles also allows a better understanding of the tasks that are assigned. But a certain adjustment period should be provided so that everyone becomes comfortable with their role and their tasks.

Ensure diversity in evaluators' gender and socio-cultural background

It also makes evaluation easier if the evaluation team members are diverse in their ethnic origin, gender and religious backgrounds, to name just a few characteristics. The clientele of PRVE programs varies widely, and some programs work with clients from very specific communities. Having some members of the evaluation team who belong to these communities themselves and thus reflect the public that they serve helps to establish a relationship of trust with these clients. It also helps the evaluators to be more sensitive to the issues and realities that these communities face and to avoid missteps that could adversely affect the work accomplished by the program practitioners.

For example, one Canadian practitioner who works with clients most of whom come from a particular community observed that having at least one evaluator who can speak the clients' first language makes the evaluation easier and reduces the mistrust that some clients may feel toward the evaluators. Overcoming the language barrier also lets the team better explain the purpose of the evaluation and better communicate the essential information. This practitioner added that it also lets the team make sure they have obtained the clients' free, informed consent. He also said that some participants feel more trusting and at ease, which helps to improve their participation in the evaluation. The following interview excerpt gives an idea of the importance of having this kind of diversity on the evaluation team:

> So, one aspect of the **X** model is that we actually require one of our youth outreach workers to be from the **Y** community because it's a culturally responsive project. So these workers have been instrumental in building that relationship with parents, because with a lot of immigrant communities, there's definitely a gap between the services that they'd like to engage with and what's actually out there. By adopting a culturally sensitive approach and involving the community not only in the program itself but also in the program evaluation, we've achieved some amazing results, and the community has been very responsive for the most part [...] it's also taking the time to explain to them in [name of the language spoken by this community], "This is what we're trying to achieve, and this is why it's important," and is also confidential. Right? Having the trust to be able to say, "You know, I trust this person and I'm going to do it." It has gone well for the most part, I think.

As a Finnish researcher told us, it is also important to take the evaluator's gender into account. In many cases, clients may not feel comfortable being interviewed by an evaluator of the opposite sex. He gave the example of certain communities in which it is not customary for women to talk with men.

> We were four people doing the interviews, men and women. So I talked only with men, and my female colleague talked with the women participants. We matched the gender of the interviewer and the interviewee, and that's important.

This may also apply in other situations, such as when the client simply has a personal preference or has experienced a sexual assault. Care should therefore be taken to include both women and men on evaluation teams, to encourage gender diversity and ensure that everyone's beliefs and sensitivities can be considered.

Make sure evaluators have enough experience and the required skills

PRVE is a fairly specialized field, so it may be hard to find evaluators who combine a good theoretical and/or practical knowledge of PRVE with experience in evaluating PRVE programs. Also, because external evaluations are expensive to begin with, organizations with limited budgets may not have the financial resources to recruit teams that have such experience or to hire external evaluators from related fields (such as working with street gangs) or even sometimes completely unrelated ones (such as medicine) in addition to hiring consultants who specialize in PRVE. Consequently, some organizations will either decide to conduct an internal evaluation instead or accept an external evaluation team that has no expertise in PRVE. One factor that facilitates

the evaluation when the evaluators are unfamiliar with the theoretical foundations of PRVE or with evaluation is to build capacities and develop tools that can provide a structure for the evaluation, as one Canadian practitioner described:

> There's a need for evaluation of CVE, but there aren't enough people who know how to do it, because it's very difficult and can be very expensive. So there's a need to build capacity and to develop tools that can be used with minimal resources at the local level, because not everybody can have this level of evaluation funding.

Provide guidelines for evaluation team members

When the members of the evaluation team have little experience in evaluation or in PRVE, giving them a common guide or protocol to follow when collecting data for evaluations seems to facilitate the evaluation process. The team can then produce their evaluation in accordance with guidelines developed by people who do have expertise in these areas. Some of our interviewees mentioned the usefulness of toolkits that provide guidance for organizations that have less experience in evaluation but do not have the financial resources to hire outside evaluators. But this facilitating factor is relatively controversial, because it raises concerns about standardization. Such toolkits do in a sense standardize evaluation protocols so that they can be applied elsewhere, which many evaluators do not support. Opinions about standardized tools and methodologies are especially mixed among researchers and practitioners who work out in the field. But there is still some consensus that standardization can be beneficial so long as the evaluations do take program-specific issues and local circumstances into account. Among our interviewees, it was the practitioners who placed the greatest stress on the importance of such issues and circumstances. One German researcher described how he encountered a great deal of resistance when he presented his toolkit to practitioners who disagreed with standardization:

> There was a lot of hostility—the field pushed back, they completely rejected it. In Germany, many rejected it, because they claimed that it was just unfairly standardizing their work, and they think that they need to be open and flexible at any given time and that their programs in their country cannot really be compared with others and that deradicalization work cannot be standardized at all.

In general, the researchers/evaluators whom we interviewed perceived standardization as having more benefits for the quality of evaluations than the practitioners did. The researchers/evaluators felt that following standardized instructions systematizes the

taking of measurements and facilitates the evaluation, especially when it has to be conducted at multiple sites. One French researcher considered standardization essential for carrying out evaluations:

> [...] You have to have standardization; without standardization, evaluation is impossible. That's the truth, and that's why I want to see more specifically which tools they have, because I hope they already have some form of standardization internally to evaluate the **X** programs, and on the basis of this standardization, about which I might have some critical opinions to offer, I might draw some broad conclusions myself [...].

Ideally, the guidelines for an evaluation should be independent of the ideologies associated with the culture and the types of radicalization, so that they can be adapted and applied in a variety of contexts. For example, the guidelines should not deal solely with Islamist ideologies or ideologies that are specific to a given country, but should be more general, so that evaluators working in a variety of countries and contexts can take them as a framework and then adapt them.

> [...] This toolkit is designed in a way that basically addresses core mechanisms of deradicalization programs. It isn't culturally sensitive, it doesn't really tell you about the role of the particular ideology and so on. But any program in the world, if they hope to have a solid logic and internal integrity, they need to address the kinds of factors that this toolkit covers.

Despite the disagreement around the use of standardized, "turnkey" tools, the majority of our interviewees agreed that the local context is the most important consideration when toolkits and guidelines are used. The following excerpt from one of the interviews summarizes this consensus nicely:

> I think what works is to be super context-specific [...] As long as there is enough insistence on taking the time to look at the contexts and the specifics within the context, in terms both of the geographic area or areas that you're looking at and of the project itself, or the team itself, or the institutional culture of that institution—taking the time to do the assessment and be context-specific rather than coming up with pre-made tools.

Some interviewees also suggested using a toolkit that was more informative and less directive and left more room for the programs to shape their evaluations to meet their specific needs:

> So, I wanted to give the power back to the programs. And I don't know if you've actually seen the toolkit, but the toolkit itself doesn't tell you exactly how you have to fill each factor with content. It just describes the fields and the factors that you need to discuss [...].

This kind of toolkit tells the evaluators about the factors that are important to consider for the evaluation and in PRVE, but without being too directive. It thus serves to train the evaluators while facilitating collection and analysis of the evaluation data.

One of our interviewees said that providing orientation days, training sessions and explanatory documents to new employees also seems to facilitate evaluations. Holding "train the trainer" sessions also fosters continuous learning among team members and thus enhances the continuity of the evaluation process.

Lastly, to facilitate evaluations, one possible alternative to guidelines and toolkits would be to review evaluations that have been done in the past, identify the lessons learned from them, and develop an improved protocol based on those lessons. This review would also provide useful information on how the evaluations proceeded and how the partners worked with one another:

> That's right. We made adjustments according to the time, the group, and the fact that the project was under construction. So we made some adaptations, but basically, we always used the same research protocol that I have been using for all my other projects for the past 16 years.
>
> We're still talking about two different initiatives, but many of the same partners. So I found the previous evaluations helpful. Obviously, anytime I do evaluations, I always look closely at the past evaluations to see what the experience was, how easy data collection was, how the partners worked together, how well the program responded to the evaluator—that just gives me a sense of what I'm getting into. So I used it in a couple of different ways.

In conclusion, in an ideal world, the people who evaluate PRVE programs would be familiar with such programs and the kinds of issues that they face and hence would be able to produce high-quality evaluations well suited to the needs of this field. But in reality, for financial and logistical reasons, that is not always possible. Most PRVE program organizations have to either work with evaluators who are not trained in PRVE or conduct their own evaluations in-house. In such cases, guidelines and toolkits can be used to inform the evaluation team about the process and the specifics of PRVE and build their capacity to conduct the evaluation. However, as one Canadian practitioner described, evaluators must also give due consideration to local issues that are specific to the community or setting in which the program is delivered:

> Understanding the nuances and challenges of this particular space would help evaluators to craft appropriate evaluations for these types of initiatives. It's not just about having a foundation in this space in general, but also the specific challenges that practitioners and programs face in it. So, the fact that the sample size will always be small, that there will be a struggle with language, things of that nature—I think that evaluators should be aware of these things when they're coming in. I also hope that all evaluators will do that regardless of the initiative or the field that they're working in, but understanding the local context.

Thus, it is easier to consider the issues that are specific to the local context when some members of the evaluation team come from the community or setting where the program being evaluated is delivered. In this way, the team can place the results of the evaluation in perspective and make more appropriate recommendations.

2.2 Methodological facilitating factors

Methodological factors were the second category of factors facilitating PRVE program evaluations that emerged from our interviews. These factors relate directly to the development of the evaluation protocol and methodology, rather than the planning and implementation of the evaluation itself.

2.2.1 Flexibility and adaptability

According to our interviewees, being flexible and able to adapt one's methodology to the changing situation in the field is a facilitating factor for an evaluation. There is a difference between an evaluation protocol that is "perfect" from a methodological standpoint and one that is realistic and takes budget limitations and actual field conditions into account. These aspects are dynamic and change constantly. Hence, being flexible in one's methodology and developing an evaluation protocol that leaves room for adjustments as the work proceeds make this work easier for the evaluators in the long run.

> We were able to ... get to a consensus, and the biggest thing is, where there was, I wouldn't say a divide, that's too strong a word, but where some people thought one thing and others thought another, they recognized that the protocol was adaptable enough that it could fit into their environment and they could manage those things in-house. So to me that was a saving grace for the whole process, for sure.

Well, it means that you accept some flexibility, you accept that there is some give-and-take in the research process. Sometimes the team does some initial work, and then you, as the researcher, put the research on pause, to look a few things over or check some preliminary results. And after that, you go back to your role as the researcher/observer and let the team do their job.

In addition to this indispensable flexibility and adaptability, we also identified some other methodological facilitating factors.

2.2.2 Participatory, co-creative approach

Consult the practitioners and involve them in the evaluation process

A number of the researchers/evaluators and practitioners whom we interviewed mentioned how important it is for these two groups to consult each other both when the evaluation protocol is being developed and when the evaluation is being carried out. These interviewees stressed how much having the evaluators and the field practitioners design the evaluation protocol together facilitates the process and improves the result. Even if the evaluators theoretically have expertise in program evaluation, the practitioners can inform them about the realities in the field, the specific characteristics of their clients, and the changes that they would like to see in these clients, as well as about how realistic the evaluation schedule is. The practitioners can also take charge of some of the tasks associated with the evaluation and involve themselves in the process rather than just play a passive role. This approach makes it possible to develop an evaluation that can then be operationalized and that takes as many factors as possible into account.

So, you know the pillar for us is around engaging staff. If someone is coming out to an organization, and I think they should meet with the staff and not just with management—meet with the staff teams, give the staff roles within the evaluation. So far, our staff has had pieces of ownership and pieces of ways that they could provide support— not just giving out questionnaires to clients, but in the development and in being able to comment, or being on a committee or supporting a committee at least. Around that, I think education is really critical like, why we're doing this, what's the benefit for you as staff, what's the benefit for the client. So looking at how everyone can... what's the win-win situation is critical.

One Canadian practitioner also briefly mentioned co-creation with the program's clients:

Yeah, co-creation and also clients who have bought into this co-creation and also being part of the team to support that process... When we worked with **X** in particular, they're really good at that, they're very good with client engagement. Often, on their evaluation teams, they have a former client informing the client about the experience of it, and that's fantastic.

Involve the community in the evaluation process

One Canadian practitioner told us that in developing the protocol for evaluating his program, it had been essential for him to involve the community with which he works and to build ties with some of its key members. In his view, this approach not only provides access to populations that would be inaccessible otherwise, but also facilitates the evaluation.

I think engaging the [specific ethnicity] community around researchers was a tremendous step, because there was a lot of distrust in the community—it's not a familiar of way of operating, so there were an enormous number of barriers. I think it was impressive that it went smoothly, but that was because we did a lot of thinking before taking a community-based participatory research approach, engaging the advisory board, and having truthfully worthwhile partners, hiring community leaders, so we were putting our money where our mouths were and saying that we valued this.

This practitioner even went so far as to have members of this community review his protocols and provide feedback on them so that his evaluation would be sensitive to the cultural dimensions concerned. This effort then made it easier to engage participants and carry out the evaluation.

So they reviewed our protocols and looked at every single question. They said: "No, you can't ask that," or "You have to change the wording here" and we took it. So things like, for example, assessing sexual abuse, we didn't put that in, even though it was part of the standard trauma battery, and that question was taken out. So that was one example. They had full veto power over what was asked. A lot of what we did was in response to their suggestions—for example, conducting every interview face-to-face, which sounds logical, but someone else might have tried to do it paper-and-pencil, and we did not. They helped us to make sure that every step of the way, we were doing things in a way that was the most culturally synchronic for everyone.

As discussed in the preceding pages, the factors that facilitate evaluations of PRVE programs can be divided into two categories. The first, organizational and logistical facilitating factors, include planning evaluations at an early stage, selecting appropriate evaluation team members, building a network of partners, training the teams, and developing guidelines to be followed in the evaluation process. The second, methodological facilitating factors, involve a co-creative approach, methodological flexibility and good adaptability. These two kinds of factors facilitated the work of the evaluators whom we interviewed from Canada and other countries. The facilitating factors were relatively consistent across the interviews, regardless of what country the interviewees came from. This means that these factors that were identified by the researchers/evaluators and practitioners whom we interviewed in this study can be applied regardless of where in the world an evaluation is carried out, so long as the specifics of the local context are also kept in mind. In the following interview excerpt, a Canadian practitioner shares the recipe for facilitating a PRVE program evaluation and provides a good summary of the facilitating factors discussed in the preceding pages:

[...] having an evaluation in mind, right at the start of project development, even before project implementation, that's certainly key, for sure. Number 2, engaging the organization. In my organization, we already have a bit of an evaluation culture, but you do have to develop one and engage your staff, the people who do the frontline work, as well as management and the board of directors. I think that's really critical for it to be successful. The proper allocation of resources and funding to do that is really important. I also think that feeding information about outcomes back to staff and clients is really important. Often staff and clients provide information input to an organization but they don't get anything back from it. So if you lose that momentum and engagement of staff, I think that's such a critical piece. I think also you have to be very clear about what you're measuring and what you're evaluating, and what's the targeted intent and purpose, that's important too.

04

Interpersonal Collaboration in PRVE Program Evaluations

The preceding sections on obstacles and facilitating factors identified numerous issues with which PRVE program evaluators must deal. Although each of these issues has its own distinctive features, one common thread seems to run through most of them: interpersonal relationships. Whether these relationships are between program managers and program evaluators or between program evaluators and program practitioners, the way that they are handled can make the evaluation process easier or harder. Collaboration among the various parties involved in evaluating PRVE programs can prove complicated for many reasons. For example, conflicts may arise between the program organization team and the program evaluation team or other teams. Or there may be a lack of trust regarding the evaluators and the factors that facilitate such collaborations. In this section, we focus directly on the specific issues of collaboration and relationships among stakeholders in PRVE program evaluations.

1. TRUST AND MISTRUST

The subject of mistrust arose repeatedly in our interviews with researchers and practitioners who had conducted PRVE program evaluations—more specifically, mistrust between program organizations and program evaluators and between program users (beneficiaries, clients) and program evaluators. This mistrust tainted the relationships among these stakeholders and had many negative consequences for program evaluations, such as refusal to co-operate, poor data, and workplace conflicts. The evaluators and practitioners did suggest some solutions to this problem, however.

1.1 Mistrust between program organizations and program evaluators

Many organizations depend on their program-evaluation results to obtain the funding that they need to continue their work. This dependence can cause some tensions between the evaluators and the organizations. When evaluations are done by outsiders, program practitioners may perceive their presence as intrusive and feel that they are being judged. Some practitioners may therefore become defensive or anxious when dealing with these evaluators. This mistrust between people often leads to mistrust of the evaluation process itself. Even the word "evaluation" makes some program staff uncomfortable, because it makes them think that their performance is going to be judged in a way that might have direct, negative consequences for them. One French evaluator described this sense of unease:

> I am pretty uncomfortable with the word "evaluation" [...] That's what I saw during my first interviews too: when the evaluator comes into a room, it can tend to create a bit of mistrust and to distort—I don't know if that's the right word—the relationship that I can have with the people I've come to see. If I wanted to exaggerate, I could say it's as if people felt like the teacher had walked into the room and was going to start handing out marks. But I really don't feel like that's what I'm doing.

Such mistrust of evaluators and preconceived ideas about them might be due to a lack of understanding of their role. Some people tend to be confused about or misconstrue the purpose of evaluations, which exacerbates the lack of trust and can even generate a degree of paranoia. As one Dutch researcher explained:

> [...] It's not always clear to people who are not in academia how academia is different from journalism. And I think there was some fear ... that we were out for the sensational story, and that once we had it, that's how we would score *per se*. So one of the barriers is that in the beginning, people would tend to think that we were journalists. Well, I know that they knew that we weren't journalists, that we worked for a university, but they still thought, "Oh yeah, but what you guys ultimately want is to write an article, and they sell better if they're a little bit controversial, if they give some juicy details about government programs, stuff like that." I'm exaggerating now, but I was under the impression that this was a bit of a concern. But maybe this was because I've done some other work with government partners like the police, there it was more pronounced. But once you get trust, many more doors open.

To prevent such problems from negatively affecting a PRVE program evaluation, it is essential to properly prepare program staff for the evaluators' arrival and to explain their role and the specific purposes of the evaluation. One German evaluator told us that taking the time to reassure the organization about the evaluation seems to facilitate co-operation and help to build a trusting relationship. Other crucial requirements for building trusting relationships with program organizations are for the evaluators to be transparent about their findings, communicate these findings regularly to the parties whom they have evaluated, and take responsibility for their evaluation work. This makes the evaluators seem accountable in the eyes of the practitioners and also encourages PRVE program organizations to display similar transparency and accountability regarding their prevention and intervention activities.

> We build trust, and we say that even if the impact of your program is not as desired, this doesn't mean that the evaluation will recommend to stop the program entirely. It's a point from which you can reform and improve the program and, at the same time, you as an organization show the willingness to be transparent and to be accountable for what you're doing.

By doing these things, the evaluators adopt a more neutral posture that reassures the organizations. It can be especially hard for organizations to open their doors to evaluators whom they do not know. Our interviewees therefore recommended that evaluators exercise particular patience, tact and discretion when trying to build relationships with organizations that they are going to evaluate. One Canadian practitioner put it this way:

> [...] Opening your doors to be evaluated definitely means opening yourself up to a certain vulnerability, that's for sure. It means you have to do some introspection and question some of your own assumptions. So, for example, if we ever have to go into schools to evaluate some aspect of their activities that we have supported, well then we're certainly going to have to show a lot of professionalism and tact to make people feel comfortable talking to us.

The way that evaluators first introduce themselves to program practitioners and program users also seems to play a role in securing their co-operation. Here is how one evaluator cited the impact of appearances and perceptions:

> I made quite a big mistake in the beginning, my very first time speaking to probation services staff. I wore a suit and showed up at this meeting where they're all together and I'm being introduced, and that's not how they dress. And I think I presented myself as a very formal outsider. That didn't help. So I had the impression from the outset that they were all crossed arms looking at me like, hmmm, who is this guy, younger than they were, also maybe the personal relationships there were some hesitancies to overcome.

Thus evaluators should pay particular attention to the way that they introduce themselves and present the evaluation. It is essential for them to try to build relationships with the members of the organization that is being evaluated—for example, by sharing some slightly more personal information about themselves, their work, and so on. But this approach does depend on the evaluators' own personal styles:

> But it's also very much a matter of style. I think I'm a bit more formal than that, I know **X** is fantastic at networking, and maybe her more personal approach worked far better. So that's something to figure out.

Past collaboration between the evaluators and the program organization seems to minimize mistrust and apprehension about the evaluation and foster trust instead. As one Canadian evaluator put it: "One of the main explanations was that we had long-time experience working with this organization, so they knew me, and they knew that they could trust me."

In addition to having existing relationships with program organizations, fielding a team of evaluators from a variety of professional backgrounds also seems to help build trust. As one evaluator from the United States described it, such a team can better understand the issues raised by the organizations:

> And then we also had some folks who were very familiar with the inter-agency process at the U.S. federal level, so when we had some difficulty in getting some agencies to buy in and agree to be interviewed, it was good to have a group like that who came at the problem from different perspectives and had different contacts in different agencies and actually knew people they could talk to and build trust with. Having a few of us on the project who knew the language of law enforcement and could "speak cop," you know, that was good. It's very easy to make the argument that because efforts to prevent terrorism and counter violent extremism combine everything from law enforcement and intelligence gathering to public health services, it's valuable to have folks coming from different disciplinary and agency perspectives.

Anonymity and confidentiality also seem to be key elements in building trust between evaluators and program organizations.

> We, of course, anonymized everything to make sure that they felt they could speak freely, especially because it was a small team in the beginning. Maybe they felt that their manager was not doing a good job or maybe they felt unhappy with another close partner. For us to be effective, they had to be able to discuss that kind of stuff. So through anonymization, by stressing our independence, and giving them some kind of control and sharing raw data with them, we hoped to build some rapport.

Program staff and program users who fear that their confidentiality will not be protected may hesitate to answer evaluators' questions, because violations of confidentiality might have harmful consequences such as job loss or even legal prosecution, in some cases. Smaller organizations may be especially affected by this issue, because employees may feel more reticent to talk about negative experiences that they have had with their colleagues or managers. One French researcher described how the director of the program that she was evaluating had tried to make it harder for her to have confidential conversations with program employees:

> And I think that what was hard for the director was that in the end I managed to work things out. The practitioners appreciated being able to talk with me separately, one-on-one, knowing that I would respect and protect the confidentiality of whatever they told me.

As one Canadian practitioner summed it up: "120% confidentiality."

Lastly, an open-door policy seems to facilitate collaboration between program evaluators and program organizations, but that is possible only when there is trust. It can help the evaluators to do their work if the organization receives them in a trusting way, gives them access to all the necessary documentation and to program users, and lets them conduct field observations when necessary. As one Canadian researcher described it:

> What was very, very facilitating was the wide open door that the organization gave us so that we could observe everything. Being able to get a look not only at their tools, but also at the school—we showed up like that and sat down in the back of the class and were able to observe without any obstacles. It was highly, highly facilitating.

And as another researcher, from the U.S., told us:

> They opened their doors to us, they gave us room to meet, they often came to our multidisciplinary team meetings, so they had teachers' representation and great communication. I believe we also had the teachers fill out reports, and we did have some other levels of data, they provided us with school grades, so there was a lot of communication and support from this school, as well as from the community.

1.2 Mistrust between program users and program evaluators

Mistrust sometimes also occurs between program users and program evaluators. The people who use PRVE programs are often members of vulnerable or marginalized groups and may mistrust the people who evaluate their programs. This mistrust can pose a considerable challenge for the evaluators. As one Dutch researcher described it:

> I mean, innately, with some people, there was clearly some hesitancy to speak with an outside party. It's all very sensitive work, in general, probation work, especially with these types of clients, so they didn't want any outsiders looking over their shoulders, so to speak. [...] But I think that in general, many people think that the word "evaluation" means that they're going to be criticized.

And as a Canadian practitioner related:

> [...] There's really a big mistrust, to begin with... so I think we have to expect some challenges in evaluating this population... I think that with our other programs, it's much easier to implement an evaluation framework. But with this program, we have clients who don't even want to give their real names, and some of them don't even want to even meet with us in person, right? And I see for the field of CVE, it's really challenging ... We're working with clients who might be under investigation, or involved in court proceedings, or under surveillance ... [So] this is a very challenging area, unless someone is incarcerated, it becomes a little easier, I think, but... you know, that's going to be a challenge for this particular program, right?

Thus, because such program users are so vulnerable, in addition to building trust with them, evaluators must demonstrate their independence from the authorities and the steps that they are taking to protect the users' confidentiality. This points to the importance of the evaluators' professional ethics and the ethics of the evaluation protocol. It also raises the question of evaluators' independence from the organizations whose programs they are evaluating. If all of the evaluators come from inside the organization, they might find it harder to build trust with vulnerable program users, especially if the organization has a reputation for sharing information with the police or the justice system. Even if these in-house evaluators' objectives are clearly different from those of the program practitioners, it might still be a good idea to show the steps that will be taken to ensure the confidentiality of the users' evaluations of the program.

For evaluators, building trust with program users is essential, both to increase their participation in the evaluation and to obtain better data from them. One

Dutch researcher reported that in his initial interviews with program users, his main goal was simply to build trust, especially since he works with a population that is vulnerable and frequently marginalized. Building this trust makes the work of evaluation not only easier but also more human. Another Dutch researcher felt that trust was essential to a dialogue between evaluators and users.

> So the first time you meet a person ... the data, you don't know how trustworthy it is, it is mostly building a trusting relationship. So we were introduced by the trainers, we sat in on the meetings, so they had seen us already, and then we had the first interview, and it was more or less getting to know each other, and that is also very important for evaluations.
>
> They needed to trust me, or else it wouldn't go anywhere. Yeah, when you're doing interviews, it's always the first challenge, you know, how to make sure they open up.

It is not enough for evaluators to just show up, introduce themselves, and then ask users to fill out questionnaires or participate in interviews. Talking and building relationships with them first will make them more comfortable and make it easier for them to work with the evaluators. This may, however, take some extra time.

> A lot of times, it's difficult to meet someone and say, "Hey, I need you to fill out a pre-survey." You know? You have to build the trust, you have to get to know them, and you have to kind of build that rapport, and then you can ask for the evaluation piece as soon as you think it's comfortable to do so. But I think that there's a gap in understanding the amount of time that takes.

Having the evaluators be introduced by team members who already know the users also seems to facilitate the development of this trust. It highlights the importance of collaboration between the evaluation team and the staff of the organization that they are evaluating.

> You have to find sort of a middle person, you have to find a connecting person, otherwise people will just say, "Ah, here's another person from the government. What do you know? You're only here to make life more miserable for me."

One Dutch researcher felt that sharing some information about himself with program users helped him to open a dialogue and build a rapport that made it easier to create trust. He also said that being transparent about the purpose of the research also seemed to play an important role.

I did also tell a little bit about myself, so not only asking them for information but also telling them a bit about myself, such as "I was raised religiously, in the Christian religion," but I also told them that I went to church, and I remember that the students really appreciated that, that I also gave some private information about myself. We were very direct with them about our purposes, because one of the issues in this training program is ethics. So they had to sign an ethics form, so they knew that the project looked at radicalization, and we sat together with the trainers before, and we had to ask, "Hey guys, what do you think? How can we tell them this?" And it was clear. We decided in the end that the trainers should explain to the youngsters what this is about, and so we gave them the forms as they were approved by the ethics board here at the University of **X** so that the youngsters could read this and the trainers, while they were doing that, explained the context of the research..., etc., etc., anonymity... these kind of things... yeah, we made it crystal-clear that the results could not be traced back to them as individuals.

2. LANGUAGE BARRIERS

A Canadian program manager and a German researcher/ evaluator both stressed that a language barrier between program users and program evaluators makes it hard for them to develop trust and communicate with each other and thus impedes the evaluation process.

I would not even say that out of the entire roster of people that we have served in the community, I would say less than 5% have reported violations of their conditions for release, only because it's also taking the time to explain to them in [language spoken by this specific community], "This is what we're trying to achieve, and this is why it's important."

So recently, I had an evaluation with refugees involved, and I used my interactive method, because at first, we had used the method where they had to write something down, and it was completely difficult, because of their [limited] knowledge of German, and they found themselves in a situation where they had to present something in the workshop... and then we changed the method into a scale evaluation and brought different perspectives together, and it worked very, very well.

3. REGULAR COMMUNICATION

Communication is an important element both within PRVE program teams and between them and their partners and evaluators. One Dutch researcher considered it essential for evaluators to communicate regularly, from the earliest possible stage, both with the people who develop and implement the program and with their partners. He also emphasized the importance of providing regular updates. This facilitates the organization's activities as well as the evaluation, because these two things are closely linked.

We had a lot of support from the trainers, so we really worked closely together, we communicated very continuously with the trainers, and got updates from the trainers in return, and that was really important. [...] It helped very much that we were there from the start. So before they actually got together as groups, we had already communicated with the organization, and the first time that the people got together, the trainers really explained what the purposes were, so we really had a good preparation there. I think that if we had started halfway through, it would have taken much more effort and trouble to explain things, because the group dynamics would have started and it would have been a new start for everybody.

4. CO-CREATING THE PROGRAM EVALUATION

The findings from this study indicate that PRVE program evaluations should be shared undertakings and that their success depends largely on how effectively the stakeholders—be they individuals, teams, or organizations—collaborate. Our interviewees stressed the importance of three main aspects of co-creating and collaborating on program evaluations: working as a team to design the evaluation protocol, aligning the partners' vision and mission, and thoroughly understanding each partner's strengths and weaknesses and how each partner's activities will reinforce and/or influence the others'.

First, our interviewees stressed the importance of teamwork in designing the program evaluation. In their view, the evaluators should work together with the organization whose program they will be evaluating, so that they can develop an evaluation protocol that is as appropriate as possible for that program's particular circumstances. One researcher from the Netherlands said that the people responsible for evaluating the program should sit down with the people who deliver it to have a down-to-earth discussion about their objectives (which are sometimes vague and not operationalized), as well as about the outcomes that they want to see and what they expect from the program. This enables the team to

establish a concrete plan, clear hypotheses, and a list of expected outcomes that will serve as the basis for the evaluation. It is interesting to see the close connection between the development of the program evaluation plan and the program's activities. In a sense, the process of planning the evaluation let the program more clearly define its objectives and the steps that it will be taking in future. This planning helps to orient the program activities and to move from a vague list of objectives to a well-defined list of hypotheses and actions whose impact can be measured. This also implies close collaboration between the evaluation team and the team that manages the program.

> So, the first phase that we had to do, which is often the case, I guess, in these interventions, is that we had to really sit down with them and say: "OK, so after this first part of the training, where you discuss issues of identity and turning points, what do you say at the end of the training, about expected outcomes?" And ... a very important part of this evaluation was to come up with very clear hypotheses beforehand, like what are the outcomes, so that you can then test them. For example, self-esteem. One of the predictions would be that self-esteem would increase after the training, and that was a very important one. And another one consisted of very objective, behavioural criteria: before, none of the youngsters had a job or an education, whereas afterwards, they predicted, they would have gotten an education, an internship, or a job. So these are very concrete criteria that need to be stated beforehand.

The second important aspect of co-creation and co-collaboration that our interviewees identified was aligning the partners' vision and mission. When all of the partners participating in the evaluation have a clear idea of their vision, mission and objectives right from the start, they can align their activities on the basis of these objectives. The evaluators should also help the partners to align their programs, activities and strategies in accordance with the vision and objectives that have been established. According to one Canadian practitioner, this will also enable them to have a common framework and a common set of indicators that will then facilitate the evaluation.

> So, I think that right from the get-go, the researcher needs to help the partners in aligning those activities and strategies to ensure that the programs are designed in a way that will help them achieve the objectives, right? Because if there is no alignment right from the get-go, then the partners would probably have to set some goals, but their activities that they might have come up with during the action plan in process might not be as strongly aligned toward achieving those goals.

> Besides, before even beginning the evaluation, like development of the evaluation framework, the researcher needs to help them in aligning their programs and the activities and the strategies to the goals and vision, right? So that would probably be the most important piece to begin with, I believe. And then going from that particular point trying to come up with a common set of indicators and tools and framework, right? To measure the work.

He added that it is important to agree on a common terminology that reflects the values of the organization and its partners. This terminology should ideally be consistent between the partners and the people working inside the organization. It should also be inclusive and non-stigmatizing.

Third, to co-create a better evaluation protocol and encourage better collaboration, a good knowledge of the partners involved in the process has proven beneficial. According to one Canadian practitioner, it is important to have a thorough knowledge of the strengths and weaknesses of each of the organization's partners and of how each partner's activities will reinforce and/or influence the others'. The consequences and ramifications–positive as well as negative–that each partner's work will have for all of the others' (especially in hub-type organizations) is an element that must be incorporated into the evaluation in order to organize it more effectively.

> When an evaluator will be assessing a collaborative effort involving multiple partners with multiple programs, rather than a single project, it becomes especially important to know the strengths and weaknesses of all the partners and the ability of their programs and activities to reinforce one another. Without that broader understanding, the evaluator won't be able to help the partnership in a realistic way.

Thus, to design a better evaluation and encourage collaboration, it is essential to consult the stakeholders, and to do so throughout the evaluation process. The evaluators should avoid working in silos and should regularly consult the organization that they are evaluating, the practitioners and the other partners concerned and invite them to discuss how they perceive the objectives of the evaluation, the indicators that should be used and the methodology that should be employed. These consultations among the evaluation stakeholders also provide an opportunity to clarify the partners' vision, mission and values. Collaboration is one of the keys to the success of program evaluations, and consultation and mutual knowledge of the partners' missions, strengths and weaknesses are indispensable tools in this regard.

05

Ethical Issues in PRVE Program Evaluations

When questioned directly about the ethical issues involved in PRVE program evaluations, the researchers/evaluators and practitioners interviewed in this study did not have much to say, except that it was critical to follow the rules set by their research ethics committees and to obtain free, informed consent from the evaluation participants. But the interviewees did raise some ethical issues of great importance either directly or indirectly without always associating them with the ethics of program evaluation as such. Some of these issues were specific to a particular geographic area or to whether the person was a researcher or a practitioner, while others concerned all of our interviewees, regardless of their country or profession.

These ethical issues fell into two categories: 1) issues related to the evaluators' independence from political and financial pressures and 2) issues related to research ethics. The first category of issues arise, first of all, in terms of politics and the media and have ethical repercussions both on program evaluations and on program users. The issues related to research ethics mainly concern transparency in recruiting the individuals who participate in the evaluation and access to the participants' data (such as their medical and criminal records). The following two sections discuss these two types of ethical issues.

1. EVALUATORS' INDEPENDENCE FROM POLITICAL AND FINANCIAL PRESSURES

The evaluators' independence is one of the ethical issues of greatest importance and concern. When questioned about their evaluation reports, many of the practitioners and researchers whom we interviewed said that they had been subjected to pressure from government agencies and/or from their funders—usually, pressure to exclude from their reports findings that depicted the evaluated programs in an unfavourable light. Many of the practitioners said that they had been subjected to pressure mainly from funders, whereas the researchers said that they had faced

pressure both from government and from funders (in many cases, the funders too are ministries or other parts of government). The independence of program evaluators, and of researchers more broadly, is of fundamental importance for ensuring the quality of the knowledge generated by the scientific community. Researchers' objectivity and their independence from government and politics is what enables them to advance science and produce useful knowledge that can be translated into practice. Funders and government agencies that exert pressure on evaluators are violating the ethical principle of independence. Some of our interviewees even equated such behaviour with censorship. In some cases it might also be regarded as a form of coercion, in light of the power that these actors have over the evaluators, who often depend on them for funding with which to operate, as the following interview excerpts illustrate:

> [...] When you've been awarded a contract to do an evaluation, are you really going to take the risk of disagreeing with whoever awarded the contract, knowing that your funding depends on them?
>
> But we actually had a more difficult time—I realize this is being recorded—but a more difficult time with our funders, because they kind of wanted us to keep some things out of the report because they would reflect poorly on some of their counterparts. And we made it clear that that's the last thing we would do, because it would amount to censorship. It would be horrible for us if we did that and it became apparent, and we also said that it would be horrible for them if it became apparent that they even tried to do that.
>
> So, there is very high donor push for exactly the results that you try to obtain. Almost ignorance, sometimes, about complexities, even though they know, but they also have very high political pressure from their respective governments, who just want to have it reported that everything has been delivered and implemented and it's all fine and dandy. Which is ridiculous. And for honest reporting and good evaluations, I also think that's the main issue: it has a very detrimental effect to the extent that, because donor-funded organizations are so dependent on their donors, they tend to just do positive reporting. And I'm trying to like, push our own government to say well, maybe exclusively positive reporting should be considered a warning sign in and of itself, because it's just not realistic.

This raises broader questions about how funding affects evaluations. Many funders try to focus on funding programs that produce tangible, positive results, or at least appear to do so. When organizations receive program evaluations that contain some negative elements, they worry that their funding might be cut off and that they might ultimately have to lay off staff or stop providing services. Such concerns feed the apprehension, anxiety and in some cases outright hostility that some organizations display toward program evaluations.

However, in the interviews, all of the practitioners and researchers who mentioned experiencing pressure also said that they had refused to give into it, regardless of the consequences. The reasons that most of them cited were the importance of the principle of independence in evaluations and research, and their own sense of integrity. They also proposed some potential solutions that we discuss later in this report.

Some of the researchers whom we interviewed said that in some organizations delivering PRVE programs, it is unfortunately still common practice to falsify or "cherry-pick" the findings from evaluation reports. These researchers said that this problem should be addressed rapidly. Beyond its ethical consequences for research findings, it also creates unfair competition for funding and tends to reinforce this bad practice. Even more problematically, it leads to a renewal of funding for programs that may look effective on paper because they have received biased evaluations, but that do not really produce benefits in practice. This exposes the users of these programs to consequences that may be harmful not only to their psychological health, but to other aspects of their lives as well.. It also places society and public safety at risk, because in the most extreme cases, if users who are often vulnerable do not receive proper supervision, bad programs could accelerate the process leading to acts of terrorism or violent extremism. As one researcher underscores:

> [...] The program would do anything they can to appear to be highly effective and ethical, and if you don't talk to the clients themselves, you'll never find out if the program is actually a huge scam. So this is really, really problematic, and in that case, the researchers actually hate this kind of situation because unwillingly or subconsciously, they're basically giving them a clean bill of health, or, you know, they say "You have been evaluated" and with that evaluation, the program goes out, gets more funding, gets more clients, and does more of their parasitical work. So, I've seen some evaluations make these bad programs even stronger, which is absolutely problematic in the field of countering radicalization, terrorism and extremism. I can't say this often enough: bad programs are not just a waste of money and resources, they are actually dangerous. They create greater risks.

Thus this ethical issue has implications that go well beyond research. If political and financial pressures successfully undermine the evaluators' independence, that could ultimately have an impact on program users and on society as a whole.

Box 4. Media, political and financial pressures on PRVE program evaluations

Radicalization and acts of terrorism receive heavy coverage in the media. According to our interviewees, the combination of media influence and political pressure can have major consequences for the independence of researchers doing PRVE program evaluations and hence ultimately on the health and safety of program users and the public. Of course, the relationship between the press and politics is neither one-way nor linear. Many other actors are also involved in it, such as businesspeople, interest groups, social movements, and moral entrepreneurs. This has been widely discussed in the literature, particularly in the context of the construction of social problems (Neveu, 2015). This relationship is also affected by numerous, complex factors. The statements related in this report reflect only the way that this relationship and its impacts on evaluation of PRVE programs are perceived by the people whom we interviewed.

The issues associated with evaluators' independence from political and financial pressures thus appear to be closely related to one another. On the basis of the interviews that we conducted, we have developed a dynamic model that reflects the perceptions of some of them. We recognize that this model does not provide a sufficiently broad perspective on the issues involved or consider all of the factors that might be relevant to understanding them. But it does afford a more general way of visualizing these ethical issues, and it does show that a combination of media influence and political pressure can have major consequences for the independence of researchers doing evaluations and hence ultimately on the health and safety of program users and the public. This element therefore deserves particular attention, because of the importance of its impact on PRVE program evaluations.
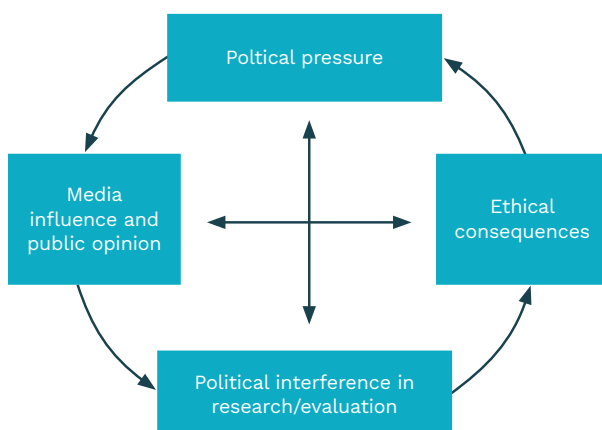
1. Media influence and public opinion

The way that the media report on events that occur in society can influence public opinion. For example, if the media in a given country give an especially sensationalist slant to their coverage on the threat of Islamist terrorism, the public might get the impression that violent radicalization associated with Islam is the greatest danger even if, in fact, radicalization associated with the extreme right is the main threat in that country.

> [...] There is a lot of consciousness that needs to go into the division of attention given to the different topics in violent extremism generally. Because we also don't want to focus too much, and I think the UK does a significant portion of their CVE work around Islamic jihadism. So that is something that we really take into account, which I don't know if it could skew the evaluation, because some people think we should focus more on some topics than others. But yeah, that's often a challenge.

2. Political pressure due to public opinion

Media influence and public opinion in turn exert pressure on politicians. This pressure may incentivize governments to implement programs, take actions and communicate about them to show that they are keeping their promises and doing something (with the ultimate goal of remaining in power). Thus, instead of concentrating on the problems that evidence-based data show merit special attention, governments focus on addressing the public's priorities and getting results in the short term. Some of the researchers and practitioners in this study had seen this pattern and believed that their governments tended to focus on programs to prevent Islamist radicalization rather than including other types of radicalization that are also present in society. But as described in the following interview excerpt, there are some indications that this situation may be starting to change:

> Because at the time, when we started, it was still much more about Islamic, and much less about right-wing. But even at that point, we had to have stark conversations about "Look guys, we have to make sure that, I strongly believe that these kind of programs have to not be focused on any particular kind of ideology, but broad".

3. Political interference in program evaluations:

Under the circumstances just described, politicians often feel intense pressure for their initiatives to be effective and produce tangible results that they can then report to the public. This pressure leads some governments to interfere in PRVE program evaluations so that the findings show the programs to have been effective and to have yielded positive results. Because government is in the vast majority of cases the main source of funding both for the organizations that deliver PRVE programs and for the people who evaluate them, they can easily be subjected to financial pressure. The fear that their funding may not be renewed may induce organizations and evaluators to yield to political pressure. Some of the researchers/evaluators and practitioners whom we interviewed said that certain agencies of the European Union had asked them to completely remove the negative findings from their evaluation reports, so as not to look bad to the public. In other cases, they showed a desire to control the language used in evaluation reports, and in particular the wording of their conclusions, so as to depict the results in a positive light. As one researcher described it:

> They would always be concerned about how things were phrased, and at some point I felt this concern about how things were phrased was going too far, and we had to make sure we had our boundaries and that if we are factually incorrect, you are always welcome to show us where that's the case. But how we write about things, and the judgments we make, those are ours and cannot be influenced.

Lastly, in addition to financial pressure, the position of authority that politicians occupy vis-à-vis the program organizations also influences the situation.

4. Ethical consequences for program evaluators, program users and the public

The first consequence of pressure from funders and government is to compromise evaluators' independence, so that they and/or the organizations whose programs they evaluate yield to this pressure and alter their evaluation findings. Other organizations, in their evaluations, will manipulate their measurements so as to show only the positive aspects of their programs—for example, an organization might choose user satisfaction with its program as a metric, but then make sure that the evaluators could interview only users who were already known to be satisfied with the program.

> It could very well be that these people were selected because they all had a very positive experience. Because I think the NGO, again, I'll send you the summary, but they said that they had helped hundreds of people since the time that they existed. Well, we interviewed and compared to the number that they had helped, and the percentage was only maybe like 20%. All of these users, so it's very hard for us to know whether these people were selected because they were actually satisfied with their results [...] Some of them also weren't very able to tell us even the name of the NGO that had helped them, or didn't seem all very aware. But we also had one or two inmates who'd be like, "I want to start by saying how great this organization is. This is not just some organization, it is absolutely great and should be funded."

Thus a program that was actually ineffective or even potentially harmful might secure a good evaluation by deliberately measuring the wrong indicators. Failure to report or evaluate the negative aspects of a program can have ethical consequences. For example, a harmful program might threaten its users' mental health or their lives, or even increase the risk they posed to public safety. To repeat an interview excerpt cited earlier:

> [...] The program would do anything they can to appear to be highly effective and ethical, and if you don't talk to the clients themselves, you'll never find out if the program is actually a huge scam. So this is really, really problematic, and in that case, the researchers actually hate this kind of industry because unwillingly or subconsciously, they're basically giving them a clean bill of health, or, you know, they say "You have been evaluated" and with that evaluation, the program goes out, gets more funding, gets more clients, and does more of their parasitical work. So, I've seen some evaluations make these bad programs even stronger, which is absolutely problematic in the field of countering radicalization, terrorism and extremism. I can't say this often enough: bad programs are not just a waste of money and resources, they are actually dangerous. They create greater risks.

This can potentially have harmful consequences that would not have occurred otherwise, such as successful or failed attempts to commit acts of violence. These events are then reported in the media, thus influencing public opinion and perpetuating the vicious cycle.

2. RESEARCH ETHICS

The second category of ethical issues involved in evaluating PRVE programs concerns research ethics more specifically. Here, three main issues arise: evaluating PRVE programs that work with vulnerable groups, achieving transparency while avoiding stigmatization, and accessing participants' personal data while protecting their privacy.

2.1 Evaluating PRVE programs that work with vulnerable groups

PRVE programs that focus on secondary and tertiary prevention work with vulnerable groups, such as incarcerated people, minors and marginalized communities, which raises ethical issues for program evaluators. This is why, according to the researchers/evaluators whom we interviewed, research ethics committees take so much longer to approve their evaluation protocols when they involve interviewing program users. Some researchers appreciate the reasons behind such delays, but are so pressed for time to complete their evaluations that they instead choose protocols that do not require interviewing program users from vulnerable groups. One researcher told us that he found such decisions regrettable, because having access to program users lets evaluators produce far higher-quality program evaluations than if they could speak only to program staff.

Several researchers/evaluators reported that meeting with members of vulnerable groups as part of a program evaluation could be very delicate and required a great deal of caution, because any misstep on the evaluators' part could wipe out all the progress that the program had achieved and thus undermine or destroy the trust that the program practitioners had built up with these groups.

Lastly, one researcher who had conducted program evaluations in developing countries said that it was hard to obtain free, informed consent from people who were illiterate:

> For example, almost all of the participants whom we interviewed were illiterate, [...] I interviewed a lot of terrorist prisoners. So same there, you write these lengthy ethical paragraphs on how to deal with sensitive populations and consent procedures, but I always feel like, in different power constellations, it's very hard to get a grasp on how "voluntary" voluntary really is, and specifically when it relates to consent. [...] You never know what they have been told, or what they maybe think.

The same researcher also observed that prison officials in these parts of the world can sometimes exercise coercion in recruiting inmates to participate in program evaluations, so that researchers cannot always have control over the recruitment process or be sure that the rules of ethics have been followed.

> Because I would say, "Okay, can I go to the prison cells and explain to the prisoners what I'm here to do and ask who wants to be interviewed?" and they would say, "No, no, no, no, no, it's too dangerous. We'll just bring you the ones who want to participate." And yeah, then they throw a prisoner in your room, and you have no idea whether a gang leader told them to come, or what the prison staff basically told them. And even if they don't tell them to come, there are always unspoken power arrangements, right?

2.2 Achieving transparency while avoiding stigmatization

The Canadian practitioners whom we interviewed pointed out the sensitivity and stigmatizing potential of the language associated with radicalization and violent extremism. In delivering and evaluating PRVE programs, practitioners and evaluators have to explain the objectives of their programs and evaluations to the participants, so that they can give their free, informed consent. But these practitioners said that if these explanations include certain terms associated with radicalization and extremism, they may tend to stigmatize the participants, and hence should be replaced with less threatening terms. This creates a conflict between the need for clarity and transparency and the need to protect program and evaluation participants from negative feelings and further marginalization.

> It's just not helpful to talk in language of, you know, "You've been referred to this program that deals with extremism or deals with radicalization," it doesn't add anything to the intervention, but it does take away from it, as far as continuing to stigmatize the client, continuing to place them in this particular category of extremists or terrorists. So, I think it does a disservice to us, but at the same time, it's important that we're able to be transparent. So, and I say this because, even in my work with clients, it's not that I absolutely won't use the word "extremism," because that is what we are funded to work on, but I try to add context to it, because it's important that folks know what they're signing up for... we have been funded to address or counter radicalization and violence, and without transparency, we can't gain informed consent from our clients, so that is a hurdle that we have to overcome. But we rely on the rapport that we have built with our clients to be able to carry us over that hurdle.
>
> [...] Here the challenge is that people are very,

> very sensitive, even about terminologies, like the term radicalization, the perception of this term is that it is exclusively used for Muslim communities, right? So don't even use this term, imagine the outcome of the evaluation, right? So, you have to be very, very careful about the language, the terminologies and the way you prepare the narrative and the communication and the messaging so that when we have the outcome report, it doesn't stigmatize any particular demographic or any particular race or gender or ethnicity. And that is something that would be a challenge.

But what new terms should be used to replace those that are perceived as stigmatizing? If the participants are to give truly informed consent, these new terms must be just as clear as the old ones.

> I stressed that to our evaluator, to take that same route of the language that you use, not only when you're interviewing clients or stakeholders but also when you are writing your evaluation, I think… we can add new language to the lexicon around CVE and move this conversation forward. I think language has consistently been a challenge even when we say we don't want to use the language of radicalization and extremism, then the challenge becomes: "What do you replace it with?" So things like "destructive ideology" or "targeting of identifiable groups," things like that, but still I think it's been a struggle.

Practitioners and researchers/evaluators must thus strike a balance and try to use language that is both transparent and non-stigmatizing.

2.3 Accessing participants' personal data while protecting their privacy

Accessing participants' personal data was another important ethical issue identified by our interviewees—more specifically, by certain researchers/evaluators. In their view, accessing participants' personal data, such as their medical and criminal records and in some cases the data that they generate on the Internet, creates a number of dilemmas because of the risks of breaching their privacy. The researchers/evaluators who raised this problem identified two main issues. The first concerns making secondary use (for evaluation purposes) of the data that PRVE programs maintain on their users. The second concerns submitting requests to the appropriate authorities to access program users' medical and criminal records.

Regarding secondary use of PRVE program data on users, these interviewees pointed out that these users have not necessarily given their consent for the information

that the program has compiled on them, much of it in the course of program interventions, to be used for program evaluation purposes. In fact, this constitutes a secondary use of this data to which, according to some of the interviewees, the participants would have to provide separate consent, given the sensitivity of the information in their program files.

> Well, of course, there are data-protection issues at stake that you might as a researcher or as an evaluator come across data that is ethically hard to handle because the clients might not have agreed to their data being used for evaluation.

With regard to protection of privacy, one researcher noted that it is very hard to access participants' criminal, legal and medical records, even if one makes official requests to the proper authorities through the proper channels. According to him, access to such data would enhance evaluations of programs and facilitate assessment of their impact. But complicated, time-consuming procedures, data-protection legislation and authorities' reluctance to share these records all make accessing such data difficult.

> You need to get the funding for the evaluation, and you need to get the co-operation of the programs to be evaluated. In Germany and the European Union, data protection is a massive headache, with the European data protection legislation, all of that is complicating the matter further.

Interestingly, one Canadian practitioner told us that it is harder to access program users' data in Canada than in European countries, because the Canadian data protection legislation is even stricter:

> So, there's this one big study, I think it's being done in Sweden, like a life course analysis of people who have gone through these programs, and there the researcher just got access to basically every client's entire history, like from primary school through to medical records, through their involvement with the program, and looked at what worked. That's obviously not practical or legal in Canada. So one of the challenges of evaluation is it's not only the ethics of it, but just the legal constraints on what kind of information can be shared and with whom, as it relates to general human services data.

The evaluators whom we interviewed seemed to perceive the data- and privacy-protection rules in their own countries as more restrictive than those elsewhere, but this ethical issue was not mentioned often enough to provide a very representative picture of the situation. It would be worthwhile to conduct studies to investigate

this question more objectively.

Lastly, the use of digital data that PRVE program users generate when using the Internet raises its own ethical problems. Deployment of online PRVE programs is increasing dramatically, and access to the associated data and protection of users' privacy are among the greatest ethical issues that such programs raise.

So one of the things that people often want to do with online work related to radicalization and that kind of thing is that they want to be able to measure and evaluate behaviour change... they want to see people actually changing their behaviour, changing their perspective in the online space as a consequence of a particular intervention or a particular treatment, and that is nearly impossible for a variety of reasons, one being privacy. A lot of the time, even if it would be technically possible to

see people's activity online in such a way that you could understand something about their change in behaviour, the ethical limitations there and the limitations of privacy are quite high, particularly for people who move back and forth between low-risk and high-risk environments.

The thin line between potential research benefits and privacy protection thus creates ethical dilemmas for program evaluators, who may be forced to adjust their protocols to strike a compromise between these two concerns.

06

DISCUSSION

How Can We Involve All PRVE Actors in Evaluating PRVE Programs?

Viewed more broadly, this study has shown that PRVE program evaluation is a dynamic process in which various forms of applied reasoning, often associated with the professional backgrounds of the various actors (funders, researchers, evaluators, practitioners, etc.) interact and sometimes clash. Each of these groups of actors has its own distinctive ways of perceiving, legitimizing and validating PRVE interventions, so what one group perceives as legitimate, valid, or of high priority may differ from or even contradict what is so perceived by another. For example, for program practitioners, protecting the confidentiality of users' information is a central concern, whereas for researchers and other program evaluators, accessing this information is essential for conducting successful evaluations. Meanwhile, for public servants in government agencies that fund PRVE programs, it is just as fundamental to consider sound management and efficient resource allocation.

These differentiated positionings are also expressed in a differentiated approach to what program evaluation should be, what purpose it serves, and what results it is expected to provide. These elements consequently determine the definition of program evaluation for each

professional group, the role that it can play, and the validity of this process and its results. The various actors involved in PRVE program evaluations thus come to the process with various preconceived ideas about it. These ideas must be levelled, clarified, and then negotiated among these individuals so that the evaluations can be carried out and regarded as relevant, satisfactory and useful for all parties concerned.[3] Ideally, the need or obligation to conduct a program evaluation will be only the starting point for the mobilization of these preconceived ideas, which will evolve so that differing views are taken into account. But this process will not always go smoothly: clashes among these varying ideas will occur in the course of evaluations, causing tensions among the actors involved. Their ability to resolve these tensions often constitutes a factor in the success or failure of an evaluation. A program evaluation is a success when all of the parties concerned see the usefulness of its findings from the standpoints of their respective professions. The program practitioners will regard the evaluation as useful if it helps them to improve their practices. The program funders will regard it as useful if it helps them to make good decisions about allocating resources. The researchers/evaluators themselves will

---

3   This reasoning is inspired largely by the ethnomethodological approach, in particular as applied by Cicourel (1995) and Vargas-Diaz (2021).

be satisfied with their methodology if the other actors perceive their evaluation findings as useful. When one of these viewpoints predominates, or when the evaluation is seen as satisfying only two of these sets of stakeholders at the expense of the third, then the evaluation will be perceived, in particular by this third group, as irrelevant or unimportant.

One final element to consider is the dynamics of power.[4] Throughout the program evaluation process, the various actors are positioned very differently along the scale of decision-making and power. When all of the actors are present, the funder, at least in theory, has all of the power to require a particular type of evaluation that will meet the funder's specific needs (see Box 5). The researchers/evaluators, for their part, have greater latitude in deciding on the evaluation methodology. The practitioners are still further removed from decision-making power in the evaluation, but paradoxically play one of the most central roles in its success. For example, in the present authors' systematic review of published studies evaluating PRVE programs, we found only one study where the author was a practitioner, which shows what place practitioners occupy in the evaluation process. Lastly, the program users or beneficiaries are among the actors who have the least power in the program evaluation process. They are often included in this process as objects for observation, or subjects for experimentation, or sources of information, but little is done to take their views into consideration in planning and conducting the evaluation or interpreting its results.

Box 5. When funders' and researchers' perspectives predominate

In our systematic review of published studies evaluating PRVE programs, we found that several of those studies that involved programs in Africa had been written in English, even though that was not the official language of the countries where these programs had been delivered. Many of these evaluations had been conducted by evaluators whose first language was English and who did not speak the language of the country in question. Most of these evaluations were of very high quality methodologically and fully satisfied well-established scientific criteria. But they obviously were not designed to meet the needs of local practitioners or to help improve PRVE practices. Most likely, these evaluations were conducted to meet the administrative needs of the development agencies that had funded the programs in question.

1. THREE SOURCES OF TENSION IN THE PROGRAM EVALUATION PROCESS

A successful program evaluation can be said to depend on successful negotiations among three different sets of actors with three different kinds of professional logic to reach an agreement on the overall framework, objectives and expectations for the evaluation. The obstacles, facilitating factors and ethical issues discussed in the preceding sections influence these negotiations while also reflecting the tensions between the perspectives of these three sets of actors. In the following pages, we analyze three particular sources of tension that seem to be central to the problems experienced by our interviewees. This analysis obviously entails some biases because of the backgrounds of our interviewees, most of whom were either researchers/evaluators or practitioners. The funders' viewpoint was represented mainly through these two types of actors. The views of program users will not, however, be explored in this analysis, not only because of the lack of information, but also because this analysis focuses on the three kinds of professional logic involved in this process.

1.1 Specificity versus harmonization

One hard-to-resolve tension in PRVE program evaluations arises between practitioners' tendency to see things from a unique, particular, local perspective and researchers/evaluators' (frequent but not invariable) tendency to try to analyze the local context with a model that is both harmonized and comparable. For example, for researchers and/or evaluators, the quality of the evaluation method is a critical aspect of the evaluation process. It ensures the validity and reliability of the results obtained, so that they can be compared with other results and thus enable valid conclusions to be drawn about the program's effectiveness and the factors that impeded or facilitated its implementation. This was, for example, the logic behind our systematic review (Madriaza *et al.*, 2022), in which the objective was to harmonize the approaches to make them comparable and generalizable—for example, by developing evaluation models that can be reproduced elsewhere. However, many authors have underscored the heterogeneity of programs that are designed and implemented in response to local particularities and specific intervention groups (Hirschi and Widmer, 2012; Lindekilde, 2012; Marret *et al.*, 2017; Mastroe and Szmania, 2016). There is a lack of consensus in the academic literature and among policy makers and practitioners on the definition of violent radicalization, and current counter-radicalization policy responses and procedures are informed by a weak and, at times, confused

---

4   According to Giddens's theory of structuration (1984), actions are understood as an interactive, intersubjective process and comprise three dimensions: the semantic dimension (shared interpretive schemas that carry shared meanings), the power and domination dimension (the system for allocating resources that exerts control over the organization's activities) and the legitimation dimension (the set of moral rules and values by which actions are justified). It is precisely these three dimensions that have been addressed in this part of this study.

understanding of the motivational and structural factors underpinning such a process. The result is a variety of interventions across the EU, signalling a lack of consensus on the purposes of counter-radicalisation. In addition, indicators of success of counter-radicalisation policies are often unclear or unspecified. One consequence of this is that assessments of the effectiveness of counter-radicalisation measures and policy responses are either lacking or often methodologically questionable, impairing our understanding of the impacts of counter-radicalisation interventions on targeted communities. The article investigates problems of assessing the impact of counter-radicalisation measures using Denmark as case study. It shows how the model of radicalisation underlying the Danish counter-radicalisation efforts translate into multilayered policy objectives and diversified policy solutions, and how the initial academic and official assessments of the impact of Danish counter-radicalisation policies on end target groups following is impaired and weakened by common methodological problems and challenges. The article concludes by suggesting some ways ahead for more systematic and valid assessments of the impact of counter-radicalisation policies in Denmark and elsewhere.

The results are thus highly dependent on the context in which the programs are implemented and, in some cases, are hard to compare with one another or transfer from one context to another. Indeed, one frequent recommendation is that PRVE interventions should be "tailor-made." For example, in its inventory of extremism prevention programs in Europe, the Radicalisation Awareness Network—a network of practitioners—says that its "DNA" requires interventions that are specific and adapted to local conditions:

> Each individual at risk is different, which calls for a case-by-case approach. It is important to understand the individuals' background, grievances, motivations, fears, frustrations, etc. to be able to develop a suitable intervention (RAN, 2019).

Practitioners' professional logic is grounded in the particularities of program users, which, in the context of program evaluation, conflicts with the professional logic of researchers/evaluators, who focus on modelling and harmonizing approaches. So here is the dilemma. Must an evaluation always meet the specific needs of the program being evaluated, even if that might prevent future comparisons and efficient use of its findings to make funding decisions for other programs? Or should the evaluation instead adopt a harmonized approach that can be applied in other contexts too, so that such comparisons can be made? The particularist logic of practitioners can also conflict with the logic of funders,

who seek to allocate resources more efficiently and make government actions more effective by applying program approaches that have received good evaluations in the past and avoiding those that have been deemed less effective.

1.2 Involvement versus independence

This tension between particularity and harmonization leads to another tension, related to the position of the researchers/evaluators in the program evaluation process compared with the positions of the program team and the program funder. As discussed in the sections of this study dealing with methodological dilemmas and ethics, the researchers/evaluators seem to find themselves in a precarious balance between their own need for independence in the evaluation process and the needs and pressures that they feel from the other actors. In program evaluation, the evaluators' independence is regarded as both axiomatic and a matter of professional ethics. Only evaluators who are sufficiently independent of and external to a program can claim to be able to view and assess it objectively, or at least from a healthy critical distance. The evaluators' independence from both the program team and the program funder protects them from certain potential biases and thus constitutes a red line that neither the team nor the funder should cross. This vision, of course, belongs to the logic of evaluative research, and it is particularly within this logic that it acquires a meaning and a legitimation.

Our interviewees fully accepted the principle of program evaluators' independence from program funders and saw any funder interference in evaluations as unacceptable and a major ethical concern. But regarding the involvement of the researcher/evaluator with the program team, a more nuanced view must be taken, in particular because of the tension described in the preceding section. As program practitioners see it, a completely external perspective on a program may fail to consider the complexities of their practice, in which they are always dealing with the contextual particularities of their programs. Thus, what the professional logic of evaluative research sees as a legitimate, independent view, the logic of practitioners would see as biased and illegitimate. In contrast, from the practitioners' standpoint, an internal view is sufficiently legitimate, because it is grounded in the experience of practice and an in-depth knowledge of the field and the issues that the program addresses. This tension partly explains practitioners' distrust of researchers and/or evaluators, the gap that they perceive between these professions are their own, and their perception that evaluation adds nothing relevant to their daily practices. In the most extreme cases, these two sets of actors behave as if they come from two opposing professional and ethical cultures, and their relations with

each other are marked by mutual mistrust. Practitioners thus perceive independent evaluation as something foreign to their practice and a low priority, something that adds yet another task in a context where time, money, and human resources are always limited.

1.3 Evaluation of programs versus evaluation of practices

The two points of tension just described lead to a third, which is expressed by two different conceptions of evaluation: evaluation of programs versus evaluation of practices. Evaluation of programs is the traditional model of evaluation, in which a program is assessed as a whole but the specific intervention practices used in it are not necessarily considered. This model makes sense to researchers, evaluators and funders. Evaluation of practices, on the other hand, is generally done formally or informally by the practitioners themselves, who focus on the particularities of the cases evaluated and the practices used, with the goal of improving their day-to-day work. The "evaluation of programs" model may thus conflict with the logic of particularity favoured by practitioners and hence not necessarily be of any interest to them.

In this sense, the design chosen for a program evaluation may also hinder the relationship with the practitioners; this is especially true of quantitative impact evaluations. Actuarial approaches, especially those that rely on turnkey tools and indicators and complex constructs that make no sense to practitioners, thus seem very distant from the realities of PRVE practice. The paradox of the control group, mentioned in the section on methodological dilemmas, is another example of how the "gold standard" that is central to the logic of traditional evaluative research is hard to apply in the field: why expose some at-risk individuals to the benefits of the intervention, but not others? To practitioners, program evaluation is thus more relevant when it shows its human side—when the evaluators treat program users not just as subjects for observation or sources of information or completed survey forms, but as people with whom they can establish relationships—in other words, when the research methods are similar to the methods that the practitioners use every day. That is why qualitative evaluation designs encounter less resistance. Instead of being an impediment, the evaluation design then serves as a way of connecting evaluators/researchers and practitioners with each other. Evaluation thus makes more sense when it comes closer to evaluation of practices, and instead of only targeting the program as an indivisible whole, also looks for ways to improve practitioners' day-to-day practices. At present, very few PRVE program evaluations incorporate this latter element.

2. THE COMPROMISE

The preceding analysis of these three sources of tension presents a somewhat exaggerated and only partial picture of the dynamics of conflict and co-operation among the three main sets of actors in PRVE program evaluations. But this analysis does illustrate the complexity of the evaluation process and partly explains the reasons that practitioners often feel less concerned by and involved in it.

First of all, these three sets of actors are not the only ones involved in the evaluation process. In addition, their professional logics and behaviours cannot be reduced to the positions that they hold or the organizations that they represent. Thus, just as these differences may, for example, be found and generate tensions within a single PRVE program, they may also generate compromises and potential solutions.

Second, to manage their resources efficiently, PRVE programs have to apply the same kind of administrative logic internally that funders apply from the outside, while many PRVE programs in Canada and elsewhere also have a very active research component; thus PRVE practitioners may become sensitized to the issues that the other kinds of professional logic address.

Third, funders are interested not only in efficient allocation of resources, but also in effective, relevant initiatives to solve social problems. Fourth, individuals from any given professional background can still be and often are highly sensitive to the issues dealt with by the professional logics of other actors. Moreover, as we have seen throughout this study, the various members of any given professional community do not necessarily all agree on how evaluation issues should be approached. The conflict is thus often not so much between individuals as between the professional logics and schemas for interpreting reality in which they are grounded and that inevitably colour their understanding and their fears regarding program evaluations.

CONCLUSION

In conclusion, this report highlights three elements to consider to improve evaluations of programs and practices for the prevention of radicalization and violent extremism. First, as we have discussed, the way in which evaluation is understood in any given professional field depends on the preconceptions (preconceived ideas or schemas for interpreting evaluations) that the various actors bring. Second, a program evaluation is not simply the application of a method to make an overall assessment of a program, but rather a complex process of interactions and co-ordination among various actors who must negotiate their interpretive schemas so that the evaluation is satisfactory for all parties concerned. Third, even if the preconceptions embodied in the evaluative collaboration process are sources of tension and conflict, they are also, and most importantly, the means of resolving them and achieving a successful evaluation—in other words, one that is satisfactory to all parties concerned. Evaluation as a task of collaboration and negotiation is, in fact, a process of aligning these preconceived ideas.[5]

When these ideas are aligned among the actors, an evaluation becomes a natural process that does not have to be forced. The extent of this collaboration depends on the extent to which the actors and organizations share similar interpretations of what the evaluation should be, what purpose it serves, and what kinds of findings it should provide. In order to reach this consensus, what is needed is not only a requirement that a program evaluation be done and a specific set of competencies and resources for doing it, but also actors who are sensitive and open to the issues that the evaluation imposes on each profession involved. PRVE program evaluations thus require "interface actors" who can translate these issues effectively and facilitate communication between professions. The role of researchers and/or evaluators is key to this process. The facilitating factors and obstacles identified in this report may help guide them in determining what paths to follow and what paths to avoid. In all cases, interpersonal relations remain a key factor in this process.

---

5   In an article inspired by Goffman's frame analysis, Snow, *et al.* (1986) proposed using interpretive frames and framing operations to understand how social movement organizations (SMOs) recruit new members and broaden the understanding of their struggles. The term "frames" designates "schemata of interpretation" that enable individuals to "locate, perceive, identify and label" occurrences within their life space and the world at large (Snow *et al.*, 1986, p. 364). Researchers in political science have used this model to understand how frames influence the implementation of policies: "How individuals and groups frame the problem opens up and legitimizes certain avenues of action and closes off and delegitimizes others" (Coburn, 2006, p. 344). Benford and Snow (2000) describe the strategies that organizations use to recruit, mobilize and acquire resources, which these authors refer to as "frame alignment processes," by which they mean "the linkage of individual and SMO interpretive orientations, such that some sets of individual interests, values and beliefs and SMO activities, goals, and ideology are congruent and complementary" (Snow *et al.*, 1986, p. 464).

RECOMMANDATIONS

On the basis of our findings in this international study, we have developed the recommendations presented below. We did so by analyzing the lessons that our interviewees had learned and the issues, facilitating factors and obstacles that they had encountered in conducting PRVE program evaluations, as well as the recommendations that they made to us directly on this subject. Although we have attempted to be faithful to what our interviewees actually told us, these recommendations are also based on our own analysis and interpretation of the information that we gathered from them.

07

Recommandations

We have divided these recommendations into eight categories representing eight different essential aspects of the program evaluation process. The **first four categories cover certain foundational elements** that must be taken into account: general recommendations, building capacity and developing evaluative thinking, funding, and the evaluation team and the collaboration process. The **second four categories deal with the successive phases of planning and conducting a PRVE program evaluation**: designing the evaluation protocol, building relationships with program practitioners and program users, conducting the evaluation, and preparing and publishing the evaluation findings.

For each recommendation, we use the following distinctive icons to identify the target audience or audiences to which it is addressed.

Distinctive icons to identify the target audiences.

Program managers

Funders

Government actors

Evaluators

Program designers

Practitioners

Program users

1. GENERAL RECOMMENDATIONS

a) Develop and encourage an evaluation culture within PRVE program organizations and, more broadly, the entire PRVE community.

b) Develop evaluation and best practice guides or guidelines for PRVE program evaluations so that organizations that do not have the resources and knowledge to conduct such evaluations can rely on high-quality information that is specific to this field. These guides or guidelines should be independent of the types of radicalization or extremism addressed and should be centred on factors related to the problem as a whole (for example, they should not be centred on Islamism or on any culture in particular).

c) When applying guidelines, always consider the local context.

2. BUILDING CAPACITY AND DEVELOPING EVALUATIVE THINKING

a) Form a qualified evaluation team and build its capacities.

b) In every organization that wants to conduct an evaluation of its own activities, train resource persons in how to use evaluation tools so that they can train new employees and thus ensure continuity in the evaluation function.

c) Provide all partners with regular training sessions on the use of evaluation tools, so that everyone uses these tools in the same way and can discuss any problems that arise and make adjustments as needed.

d) Provide training sessions on PRVE program evaluation to program staff, to build their capacities.

3. FUNDING

3.1 Relations between program evaluators and program funders

a) Funders should consider the field's needs regarding evaluation ahead of time, so that their expectations concerning evaluation will be realistic in terms of time, budget and results.

b) Clarify the funders' expectations concerning evaluation from the very start of the funding, to avoid misunderstandings and unrealistic expectations, and communicate with the funders regularly so that they can adjust their expectations in response to the realities in the field.

c) From the very start of the evaluation process, clarify that the findings may be negative and that in order to protect the integrity of the evaluation and the methods used and to adhere to the principles of ethics, these findings will not be altered under any circumstances.

d) Establish clear guidelines for the relations among evaluators, government actors and funders so as to ensure the evaluation's independence from the funders and avoid inappropriate pressures on evaluators.

e) Invite the funders to visit the field so that they can observe the realities that practitioners and evaluators deal with and better understand their situation. This will help to reframe their attitudes in terms of evaluation considering the budget allocated.

3.2 Budgets allocated for program evaluations

a) The funding for every PRVE program should include, from the start, a specific budget item for program evaluation (at least 10% of the total program budget).

b) This budget item must be large enough to ensure a high-quality evaluation, meaning that it is sufficient to hire personnel specialized in program evaluation and PRVE, to acquire the necessary materials for collecting and analyzing data, to cover travel expenses and to provide the time needed to do the evaluation properly. The time needed should be estimated according to the size and complexity of the program, the number of components that it has and the geographic areas that it covers.

c) Require program evaluations as a condition for program funding. But the renewal of a program's funding should not necessarily depend on the positive or negative results of its evaluation. Instead, it should depend on the quality of the evaluative research and on the program's ability to apply the evaluators' recommendations.

4. THE EVALUATION TEAM AND THE COLLABORATION PROCESS

4.1 Assembling the evaluation team

a) Involve the evaluators in the evaluation committee as early as possible, ideally while the program is being designed, but no later than the start of program implementation.

b) Establish an evaluation committee at an early stage of the project and train this committee so that it can be operational when the program begins its activities.

c) Assign one or more clear roles and specific tasks to every member of this evaluation committee.

d) Make sure that the evaluation team is diverse in terms of gender, religion and ethnic origin.

e) Ensure a good mix of professional backgrounds on the evaluation team. By including evaluators from a variety of sectors (including the one to which the program organization belongs), you enable the team to better understand the issues that the organization must deal with, build better relationships, and improve communication (for example, if the organization is in the law-enforcement sector, having at least one evaluator who is familiar with this sector will facilitate relations and understanding between the evaluators and the organization's staff).

f) Identify key actors and build a diverse network of local, national and international partners who come from multiple sectors, have experience in evaluating PRVE programs and will be able to help in the evaluation process.

g) When the budget allows, also try to include experts with experience in program evaluation on the evaluation committee, or to secure their support to guide this committee.

h) When the evaluations will be done in-house, invite a third party to join the evaluation team in order to provide an outside perspective, or, if the budget does not allow that, ask for a third party's opinion at individual key steps in the evaluation process (such as when the protocol is being designed).

i) When the evaluations will be done in-house, make sure to have access to resource people who have expertise in program evaluation and can provide helpful feedback about key tasks in the evaluation process (formulating theories of change, choosing indicators and metrics, choosing methodologies, deciding how to approach program users, etc.).

4.2 Collaboration and teamwork

a) Familiarize program staff with the concept of evaluation when they receive their initial training, to avoid mistrust and misunderstandings later on.

b) Encourage co-creation of the evaluation by involving program practitioners and program users in developing the evaluation protocol.

c) Align the visions and missions of the partners involved in the evaluation. Be familiar with all of the partners' strengths and weaknesses and their influence on the organization's work.

d) Develop a joint plan among the evaluators, the program organization and the partner organizations to better orient the evaluation effort.

5. DESIGNING THE EVALUATION PROTOCOL

5.1 Preliminary considerations

a) As much as possible, include provisions for administrative delays in the evaluation protocol (time needed to access records or obtain documentation, time needed to obtain responses from certain government agencies, etc.), and adjust this protocol as necessary.

b) Adopt flexible evaluation protocol and approach that let the team adjust to changes and to realities in the field as the evaluation proceeds.

c) Start developing the evaluation protocol as soon as you start designing the program, identifying the elements that will have to be evaluated and planning how they will be measured.

5.2 Theories of change

a) Ground the program and its evaluation in a clear theory of change and translate it into concrete, specific objectives that the evaluators can use to determine what change indicators they will evaluate.

b) Work together with the managers and practitioners of the program to be evaluated, in order to understand and formulate their theories of change, determine their objectives in practical terms, and choose the indicators that will be measured, all in a spirit of co-creation and within the framework of an iterative process.

5.3 Choosing indicators and methods

a) When the program is first established, select the variables to be measured, and then adjust them iteratively in order to adapt to the realities in the field.

b) Choose indicators that are clear, accurate, and suited both to the context of PRVE and to the methodology used for the evaluation.

c) Provide time in the evaluation schedule for team discussions about choices of indicators.

d) In evaluations of tertiary prevention programs (particularly in correctional settings) that use the recidivism rate as the success indicator in their impact evaluations, define what is regarded as recidivism in very concrete terms.

e) In evaluations of tertiary prevention programs (particularly in correctional settings), choose specific indicators and measurements to be taken after detainees have been released.

f) If possible, opt for sequential treatment (control) groups, so that all participants can benefit from the program.

g) For programs employing a hub or situation-table model, evaluate the effectiveness of the partnerships and the multisectoral approach, rather than try to isolate and evaluate the effect that each of the parties (for example, social workers and police officers) has had on the results observed.

6. BUILDING RELATIONSHIPS WITH PROGRAM PRACTITIONERS AND PROGRAM USERS

a) Some months before the program evaluation starts, take the time to build trust with the members of the program organization. To make this easier, introduce the evaluators to the program teams, explain the role of the evaluation and of the people who will carry it out, clarify the objectives of the evaluation and reassure the organization about its neutrality.

b) Before you start collecting data from program users, build trust with them. Some ways to do so include working through an intermediary whom these users already trust (such as a member of their own community), sharing some more personal information about yourself with them (if you feel comfortable doing so), establishing your independence from the police and the justice system, and guaranteeing the users' confidentiality and anonymity.

c) Build channels for regular communication with the program organization as well as with its partners, and have a presence in the field.

7. CONDUCTING THE EVALUATION

7.1 Evaluators' approach

a) Adopt a collaborative, non-judgmental approach, because the evaluation may make employees think that they are going to be judged, criticized and possibly reprimanded, which could make them feel anxious and mistrustful of the evaluators.

b) Pay close attention to how the program organization operates and adapt to it as much as you can. For example, find out whether the dress code is formal or informal, whether relationships among employees are fairly egalitarian or more hierarchical, and how the various teams communicate with one another. That way, you can avoid missteps and bad first impressions that would make it harder for you to secure people's co-operation from then on.

7.2 Implementing the evaluation protocol

a)   Once the evaluation has begun, do not hesitate to adjust your choice of indicators if necessary to make sure that they remain appropriate and measure what you want to.

b)   When working with program users who do not speak the languages spoken by the evaluation team, use translators and interpreters as necessary.

c)   Within your organization, adopt an open-door policy and make it easier for the evaluators to talk with your employees.

d)   Share the results of the evaluation with the program organization and its partners regularly to keep them up to date on its progress and obtain their feedback.

8. PREPARING AND PUBLISHING THE EVALUATION FINDINGS

a)   Since you know that any PRVE program evaluation is very unlikely to prove any causal links, qualify your findings by mentioning any other elements that may have influenced the changes that you observed.

b)   Keep the program organization and its funders informed about what will be written in the final evaluation report and, possibly, what will or will not be made public.

c)   Open the door to discussions and to correcting any errors. Be open to criticism. If necessary, send the stakeholders a preliminary report and discuss it with them.

d)   Formulate your findings in a non-accusatory way and accompany them with detailed context.

e)   Accompany any negative conclusions with concrete recommendations for improving practices.

f)   Publish your evaluation findings so that other PRVE actors can have access to a larger body of empirical data on PRVE program evaluation, because this is an emerging field in which such data is still scarce.

References

Baruch, B., Ling, T., Warnes, R., et Hofman, J. (2018). Evaluation in an emerging field: Developing a measurement framework for the field of counter-violent-extremism. *Evaluation*, *24*(4), 475-495.

Bellasio, J., Hofman, J., Ward, A., Nederveen, F., Knack, A., Meranto, A.S., et Hoorens, S. (2018). *Counterterrorism evaluation: Taking stock and looking ahead*. RAND. www.rand.org/t/RR2628.

Benford, R.D., et Snow, D.A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, 611-639.

Chowdhury Fink, N., Romaniuk, P., et Barakat, R. (2013). *Evaluating countering violent extremism programming: Practice and progress*. Center on Global Counterterrorism Cooperation.

Cicourel, A.V. (1995). *The social organization of juvenile justice* (Vol. 36). Transaction Publishers.

Clément, P.-A., Madriaza, P., et Morin, D. (2021). *Les intervenants et l'évaluation en prévention de l'extrémisme violent: Entre contraintes et opportunités*. Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (Chaire UNESCO-PREV), Université de Sherbrooke.

Coburn, C.E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation. *American Educational Research Journal*, *43*(3), 343-349.

Davey, J., Tuck, H., et Amarasingam, A. (2019). *An imprecise science: Assessing interventions for the prevention, disengagement and de-radicalisation of left and right-wing extremists*. Institute for Strategic Dialogue.

Feddes, A.R., et Gallucci, M. (2015). A literature review on methodology used in evaluating effects of preventive and de-radicalisation interventions. *Journal for Deradicalization*, *5*, 1-27.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.

Gielen, A.-J. (2017). Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How? *Terrorism and Political Violence*, *31*(6), 1149-1167. https://doi.org/10.1080/09546553.2017.1313736.

*Hassan, G., Brouillette-Alarie, S., Ousman, S., Kilinc, D., Varela, W., Lavoie, L., Fetiu, A., Harris-Hogan, S., Madriaza, P., Borokhovski, E., Pickup, D., Boivin, M., Srimathi Narayana, M., Rousseau, C., Gill, P., Thompson, S., McCoy, J., Venkatesh, V., et Morin, D. (2021a). *A Systematic Review on the Outcomes of Primary and Secondary Prevention Programs in the Field of Violent Radicalization*. Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV).

*Hassan, G., Brouillette-Alarie, S., Ousman, S., Kilinc, D., Varela, W., Lavoie, L., Fetiu, A., Harris-Hogan, S., Madriaza, P., Borokhovski, E., Pickup, D., Boivin, M., Srimathi Narayana, M., Rousseau, C., Gill, P., Thompson, S., McCoy, J., Venkatesh, V., et Morin, D. (2021b). *A Systematic Review on the Outcomes of Tertiary Prevention Programs in the Field of Violent Radicalization*. Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV).

Hassan, G., Ousman, S., Madriaza, P., Fetiu, A., Boily, L.-A., Levesque, F., Squalli, Z., Ajrouche, K., El-Tahry, N., Lampron-De Souza, S., Desmarais, L., Duong, E., et Moyano, R. (2020). *D'un océan à l'autre: Cartographie des initiatives de prévention secondaire et tertiaire oeuvrant dans un contexte de radicalisation et d'extrémisme violents au Canada*. Réseau des Praticiens Canadiens pour la Prévention de la Radicalisation et de l'Extrémisme Violents (RPC-PREV). https://cpnprev.ca/wp-content/uploads/2020/12/FR-Rapport-mapping-final-1.pdf.

Hirschi, C., et Widmer, T. (2012). Approaches and challenges in evaluating measures taken against right-wing extremism. *Evaluation and Program Planning*, *35*(1), 171-179.

Holmer, G., Bauman, P., et Aryaeinejad, K. (2018). *Measuring Up: Evaluating the Impact of P/CVE Programs*. United States Institute of Peace.

Horgan, J., et Braddock, K. (2010). Rehabilitating the Terrorists?: Challenges in Assessing the Effectiveness of De-radicalization Programs. *Terrorism and Political Violence*, *22*(2), 267-291. International Bibliography of the Social Sciences (IBSS). https://doi.org/10.1080/09546551003594748.

Levin, K. Study design III: Cross-sectional studies. *Evid Based Dent* **7,** 24-25 (2006). https://doi.org/10.1038/sj.ebd.6400375.

Lindekilde, L. (2012). Value for Money? Problems of Impact Assessment of Counter-Radicalisarion Policies on End Target Groups: The Case of Denmark. *European Journal on Criminal Policy and Research*, *18*(4), 385-402. ProQuest Central; ProQuest Sociology. https://doi.org/10.1007/s10610-012-9178-y.

Madriaza, P., et Ponsot, A.-S. (2015). *Preventing radicalization: A systematic review*. International Centre for the Prevention of Crime.

Madriaza, P., Ponsot, A.-S., Marion, D., Monnier, C., Ghanem, A., Zaim, B., Nait-Kac, W., Hassani, N., et Autixier, C. (2017). *La prévention de la radicalisation menant à la violence: Une étude internationale sur les enjeux de l'intervention et des intervenants* (p. 137). Centre International pour la prévention de la Criminalité. https://cipc-icpc.org/rapport/rapport-sur-la-prevention-de-la-radicalisation-menant-a-la-violence-une-etude-internationale-sur-les-enjeux-de-lintervention-et-des-intervenants/.

Madriaza, P., Morin, D., Hassan, G., Venkatesh, V., Plaude, M., Deli, C., Girard, M., Durocher-Corfa, L., Grijalva-Lavallée, R., et Poulin, K. (2022). *Les enjeux méthodologiques de l'évaluation de programmes de prévention de l'extrémisme violent : une revue systématique* [titre provisoire]. Montréal: Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (Chaire UNESCO-PREV).

Marret, J.-L., Bellasio, J., van den Berg, H., van Gorp, A., van Hemert, D., Leone, L., Meijer, R., Warnes, R., et Van Wonderen., R. (2017). *Final report providing background for using and further developing the validated toolkit Impact* (FP7 GA no. 312235). Impact Europe.

Mastroe, C., et Szmania, S. (2016). *Surveying CVE Metrics in Prevention, Disengagement and Deradicalization Programs*. Report to the Office of University Programs, Science and Technology Directorate, Department of Homeland Security.

Neveu, E. (2015). *Sociologie politique des problèmes publics*. Paris: Armand Colin.

Pistone, I., Eriksson, E., Beckman, U., Mattson, C., et Sager, M. (2019). A scoping review of interventions for preventing and countering violent extremism: Current status and implications for future research. *Journal for Deradicalization*, *19*, 1-84.

RAN. (2019). *Preventing radicalisation to terrorism and violent extremism: Approaches and practices*. Radicalisation Awareness Network. https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ran-best-practices/docs/ran_collection-approaches_and_practices_en.pdf

Ris, L., et Ernstorfer, A. (2017). *Borrowing a wheel: Applying existing design, monitoring and evaluation strategies to emerging programming approaches to prevent and counter violent extremism*. Peacebuilding evaluation consortium.

Romaniuk, P. (2015). *Does CVE work?: Lessons learned from the global effort to counter violent extremism*. Global Center on Cooperative Security.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed). SAGE.

Setia M.S. (2016). Methodology Series Module 3: Cross-sectional Studies. *Indian journal of dermatology*, *61*(3), 261-264. https://doi.org/10.4103/0019-5154.182410.

Snow, D.A., Rochford, E.B., Worden, S.K., et Benford, R.D. (1986). Frame Alignment Processes, Micromobilization, and Movement Participation. *American Sociological Review*, *51*(4), 464-481. https://doi.org/10.2307/2095581.

UNEG. (2016). *Normes et règles d'évaluation*. Groupe des Nations Unies pour l'évaluation.Vargas-Diaz, R. (2021). *Processus décisionnel autour du projet de vie de permanence pour de jeunes enfants placés en milieu substitut: L'acteur, l'interaction et le contexte*. Université de Montréal.

Williams, M.J., et Kleinman, S.M. (2014). A utilization-focused guide for conducting terrorism risk reduction program evaluations. *Behavioral Sciences of Terrorism and Political Aggression*, *6*(2), 102. ProQuest Sociology Collection.

IN PARTNERSHIP WITH

UNESCO
uni Twin

United Nations
Educational, Scientific and
Cultural Organization

"UNESCO Chair in Prevention of Radicalisation
and Violent Extremism", Université de Sherbrooke,
Concordia University, Université du Québec à Montréal

INTERNATIONAL
CENTRE
FOR THE
PREVENTION
OF CRIME