



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

IN PARTNERSHIP WITH



INTERNATIONAL
CENTRE
FOR THE
PREVENTION
OF CRIME

Constraints and opportunities in evaluating programs for prevention of violent extremism: how the practitioners see it

PIERRE-ALAIN CLÉMENT, PABLO MADRIAZA AND DAVID MORIN

JUNE 2021



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

HOW TO CITE THIS REPORT

Clément, P.-A., Madriaza, P., & Morin, D. (2021). *Constraints and opportunities in evaluating programs for prevention of violent extremism: how the practitioners see it*. UNESCO Chair in Prevention of Radicalisation and Violent Extremism (UNESCO-PREV Chair), Sherbrooke University.

FUNDED BY



Gouvernement
du Canada

Centre canadien d'engagement
communautaire et de prévention
de la violence

Government
of Canada

Canada Centre for
Community Engagement and
Prevention of Violence

THE AUTHORS



Pierre-Alain Clément

Doctor in political science from the University of Quebec in Montreal and is in charge of research on “terrorism and violent radicalization” at the Institute of High Studies of the Ministry of the Interior (IHEMI) in Paris.



Pablo Madriaza

Associate Professor at the University of Quebec in Montreal, 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

Associate 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).

Table of contents

Executive summary	5
Background	5
Highlights	6
Suggestions for action	7
Suggestions for policymakers	7
Suggestions for practitioners	7
Suggestions for designing evaluations	7
1. Introduction	8
1.1 Background of this study	8
1.2 Methodology of this study	9
1.3 Limitations of this study	10
1.4 Ambivalence around evaluation as both a constraint and an opportunity	10
2. Institutional constraints and practical challenges for PVE program evaluation	12
2.1 Influence of politics and the media	13
2.2 Institutional constraints on funding access	14
2.3 Lack of adequate definitions and indicators for evaluating PVE programs	16
2.4 Trial-and-error: formative self-evaluations as central to PVE practice	18
2.5 Restoring the specificity of the profession: evaluation as a professional challenge	19
3. Evaluations as opportunities for improving practices and autonomy	21
3.1 Evaluations promoting a coordinated definition of each actor's responsibilities	22
3.2 Evaluating practitioners' needs in terms of their professional careers	23
4. Adapting to local conditions as a basic principle	24
4.1 The right distance from the state: a crucial factor for PVE program success	25
4.2 Evaluating communities' needs and satisfaction	26
4.3 Evaluating service provided to the community as a means of defending a practice	27
5. Suggestions for making evaluations an effective, legitimate tool for improving PVE practices	29
5.1 Suggestions for policymakers	30
5.2 Suggestions for practitioners	30
5.3 Suggestions for designing evaluations	30
Thematic Glossary	31
Appendix A – Characteristics of the 57 PVE professionals who participated in the interviews and focus group	33
List of references	35

EXECUTIVE SUMMARY

This report was produced as part of the PREV-IMPACT project, an initiative of the UNESCO Chair on the Prevention of Radicalization and Violent Extremism (UNESCO-PREV Chair). It is based on semi-directed interviews conducted in an earlier study by the International Centre for the Prevention of Crime (ICPC) (Madriaza, Ponsot, & Marion, 2017) and a focus group conducted by the UNESCO-PREV Chair in Ottawa, Canada in March 2019.

The participants in these interviews and the focus group were 57 professionals involved in prevention of violent extremism (PVE), from six different regions of the world: North America, Europe, sub-Saharan Africa, North Africa and the Middle East, Asia, and Oceania. The purpose of the present report is to further examine what these professionals had to say about their experiences in delivering PVE programs in the field, and, in particular, about their experiences with the evaluation of such programs.



Highlights

The interviews and focus group revealed wide variations in the settings where the practitioners delivered their PVE programs, the needs of the communities that these programs served, the degree of satisfaction of the program participants and the outcomes of the program activities. Nevertheless, a certain common perspective on the evaluation of these programs emerged.

01

First, the practitioners **perceived program evaluations as an additional constraint for organizations** that are already operating with limited resources and under precarious conditions. The practitioners also perceived evaluations as, to some extent, both a tool for enforcing compliance and a form of rapid response to specific demands from politicians or the media. Lastly, the **practitioners generally saw PVE work as ill-suited to tightly scheduled evaluations using traditional performance indicators**. In short, a majority of the practitioners expressed a degree of skepticism, not to say a certain mistrust, about the way that PVE programs are currently evaluated, with a focus on summative evaluation (did this program work?) rather than on formative evaluation (how can PVE practices be improved?).

02

But the practitioners also acknowledged the importance of evaluating their programs. Evaluations give practitioners the **opportunity to establish professional standards, practices and qualifications**. Evaluations can also demonstrate the **need for autonomy and coordination in program** design and delivery, so as to divide tasks more fairly among the various stakeholders. Lastly, evaluations also let PVE workers **highlight needs associated with the professionalization of their jobs**: training, remuneration and career management.

03

But most importantly, the practitioners underscored the essential factors for the success of PVE programs. These factors can be summarized as a **focus on meeting the needs of the communities where these programs are carried out**. Such a focus requires not only maintaining an appropriate distance from State funding sources but also taking a genuine interest in the usefulness of PVE to these communities, so as to determine what services should be provided to meet which needs.



Suggestions for action

As just described, PVE program evaluations entail not only constraints but also opportunities. These evaluations can become useful tools for PVE practitioners on several of the listed conditions: all stakeholders must agree on the value of conducting formative evaluations as well as summative evaluations, including independent, outside evaluations; evaluations must recognize the qualitative, long-term nature of PVE interventions; the workload associated with evaluations must be divided more effectively among the various stakeholders (program sponsors, program delivery organizations, trusted third parties).

SUGGESTIONS FOR POLICYMAKERS

1. At the very outset, dedicate a portion of the program's funding to evaluations, without reducing the funding allotted to operations.
2. Conduct cross-evaluations: self-evaluations by practitioners to capture data from the field, internal evaluations within the organization to measure the program's effectiveness, and external evaluations to ensure that all organizations are treated equally.

SUGGESTIONS FOR PRACTITIONERS

3. Work with the evaluators to develop indicators that define the goals of all parties concerned.
4. Evaluate the organization's resources (training, talent retention, obstacles encountered).

SUGGESTIONS FOR DESIGNING EVALUATIONS

5. Evaluate the quality of the relationship developed with the community.
6. Develop qualitative indicators regarding practices.
7. Develop indicators to measure the organization's independence from the sponsor.

Introduction

1.1 BACKGROUND OF THIS STUDY

In recent years, as the traditional security response has failed to stem the rising tide of extremist and terrorist violence in various parts of the world, numerous programs have been deployed to attempt to counter such violence by other means. Unfortunately, these programs have not been grounded in any clear, rigorously defined conceptual and empirical foundations. Heydemann argues that this “blurring of boundaries reinforces perceptions of CVE [countering violent extremism] as a catch-all category.” (2014, p. 10) Given the urgent need to prevent violent extremism, the practitioners (see [glossary](#)) who do this work have had to adapt practices from their own and one another’s experiences in other fields, even though the effectiveness of many of these practices for prevention of violent extremism (PVE) has not yet been demonstrated. Thus, PVE practitioners have had to design their programs with relatively little evidence-based data in place to guide them, or promising, proven practices to attempt or rely on.

The literature on PVE programs¹, whether from official, scientific or “grey” sources, consistently identifies a lack of three things in this field: evidence-based studies, program evaluations, and data that capture the experiences of front-line practitioners.² As regards program evaluations, the UNESCO-PREV Chair has conducted a systematic review of evaluative studies of PVE programs (Madriaza,

et al., to appear in 2021) which found that since 2016, their number has increased substantially and their methodological quality has improved, although it remains limited on average. The increased number is due in particular to the contributions of centres of expertise and specialists outside the academic space. Nearly 60% of the studies reviewed came from the grey literature, and on average, their methodological quality was higher than that of the studies published in specialized journals. Although the shortage of empirical data has grown less severe over the past few years, when it comes to evaluating programs and capturing the experiences of front-line practitioners, shortcomings remain. Up to 2018, for example, only 16% of scientific publications used first-hand data, and half of these were evaluations (see [glossary](#)), while only 13 studies considered the views of practitioners (Ponsot, Autixier, & Madriaza, 2018, p. 2).

The question of how PVE programs should be evaluated has become all the more critical now, because of the risks that poorly designed programs entail. Some recent systematic reviews have shown that the number of programs evaluated is still limited and that problems with the methods used to evaluate them persist (Madriaza, et al., to appear in 2021; Hassan, et al., 2021a;

¹ Here we focus strictly on radicalization leading to violence (also known as violent radicalization or behavioural radicalization), which has been defined as the process by which an individual or a group adopts a violent form of action that is directly associated with an extremist ideology of political, religious, or social content which contests the established political, social or cultural order (Khosrokhavar cited by Madriaza, Ponsot, & Marion, 2017, p. 9). Violent radicalization is distinguished from simple or cognitive radicalization, which may be defined as embracing ideas which, on the basis of a rupture with the society to which one belongs, involve accepting, at least in theory, recourse to forms of political action that are unconventional and possibly illegal, if not violent (Sommier, 2012, p. 15). The connection between cognitive radicalization and behavioural radicalization is not clear-cut, but the former is often regarded, especially by practitioners, as an independent variable that is predictive of the latter as a dependent variable (Silber & Bhatt, 2007; Moghaddam, 2015). But there are many studies that treat both as dependent variables (Sageman, 2009) that may possibly co-vary (Skillicorn, Leuprecht, & Winn, 2012), which suggests that this dichotomy is not highly operational (Neumann, 2013, pp. 45-46).

² New data on these experiences have been gathered more recently, including at two events hosted over three days in Ottawa, Canada by the UNESCO-PREV Chair and the Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV). The first was a workshop entitled “The Prevention of Violent Radicalization: Evidence-Based Guidelines to Promote Efficient Interventions” (March 21 and 22, 2019). The second was a colloquium entitled “Prevention of violent radicalisation and extremism: practices, evaluation and cooperation: Dialogues between Africa, Europe and North America” (March 23, 2019).

Hassan, et al., 2021b). This is so for many reasons: the limited number of practitioners qualified to assess the methodological complexity of evaluations, limited funding, the practitioners' other competencies, and political constraints.

Despite these obstacles, there is some consensus among practitioners, researchers, and policymakers about the need to develop models for evaluating PVE programs. For practitioners, evaluations can provide ideas for improving their practices. For researchers, evaluations can provide a better understanding of the mechanisms and processes that make programs succeed or fail. For policymakers, evaluations can help to guide public-policy decisions and optimize the use of public funds.

According to Ponsot, Autixier, & Madriaza (2018, pp. 24-25), the current lack of evaluated practices in the field of radicalization prevention means that practitioners have very few best practices to draw on for their own interventions, and instead look to other fields or their own experiences to develop interventions or encourage innovations. The purpose of evaluating PVE programs is to correct this situation by identifying best practices that PVE practitioners can incorporate into their own programs in future.

Although the literature stresses the need for evaluations to provide useful information both to front-line practitioners and to policymakers, many evaluation studies are not really designed to facilitate knowledge transfer either vertically (between the front-line practitioners who deliver prevention programs at the operational level and the political/governmental policymakers in charge of them at the strategic level) or horizontally (among practitioners and policymakers from different programs, cities, and countries).

In this context, the present policy paper conveys the views that 57 PVE professionals expressed regarding PVE program-evaluation issues in interviews and a focus group conducted between 2017 and 2019 by the UNESCO-PREV Chair and the International Centre for the Prevention of Crime (ICPC). About one-third of these professionals were front-line practitioners working directly with participants in PVE programs. Another one-third were PVE program coordinators and one-quarter were PVE organization heads, most of whom were either working as front-line practitioners too or had done so in the past. The small remainder were PVE researchers, trainers, and independent contractors. These 57 professionals' observations yielded a number of preliminary lessons about PVE program evaluation that may help to inform future thinking about this subject, as well as about public policymaking and the actors in this field.

This paper is only a preliminary analysis. It is the first step in a broader project on evaluation of prevention practices, entitled PREV-IMPACT Canada and directed by the UNESCO-PREV Chair in partnership with the Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV) and the ICPC, with funding from Public Safety Canada. This project will take a systematic, structured approach to produce evidence-based data, in particular through a systematic review of PVE programs that have been evaluated in the past and an international comparison.

This paper is divided into five parts. This introduction (Part 1) presents the methods used to analyze the information that the 57 professionals provided about their experiences with PVE program evaluation. Part 1 also discusses the limitations of this study. Part 2 discusses the substantial constraints that the PVE practitioners see program evaluation as imposing, while Part 3 discusses the opportunities that they believe it affords. Part 4 discusses some factors that they see as crucial for successful PVE program evaluation. Lastly, in Part 5 of this study, the authors suggest a number of ways to make PVE program evaluation both simpler and more useful for practitioners.

1.2 METHODOLOGY OF THIS STUDY

This study analyzes the statements that the PVE practitioners made, in the interviews and focus group, about their experiences with evaluations of their PVE practices. This empirical approach was designed to capture and share the practitioners' experiences with practicing PVE in the field and to explore the approaches that they had developed for evaluating their own practices.

The interviews were conducted in an earlier study by the ICPC (Madriaza, Ponsot, & Marion, 2017), while the focus group was conducted by the UNESCO-PREV Chair in conjunction with a colloquium that it hosted in Ottawa, Canada in March 2019, entitled "Prevention of violent radicalization and extremism: practices, evaluation and cooperation: Dialogues between Africa, Europe and North America."

The objective of the 2017 ICPC study³ was to identify the main issues that front-line PVE practitioners face. More precisely, the goal was to gather specific, practical information about the implementation of PVE programs, in particular regarding the challenges and problems that such interventions involve and the methods of managing them. This study was divided into two phases. In the first, the exploratory phase, the researchers interviewed 27 experts from 14 countries. In the second phase, they interviewed 63 front-line practitioners from 23 countries

³ The complete methodology of this study is presented in appendices 1 and 2 of the original report (Madriaza, Ponsot, & Marion, 2017).

in North America, Europe, Africa, Asia and Oceania. Verbatim transcripts were produced for the interviews with most but not all of these informants: 24 of the experts and 54 of the practitioners (only 51 of these 54 interviews turned out to be usable). The researchers analyzed the remaining interviews by means of a grid that they filled out as they listened to the recordings.

In March 2019, the UNESCO-PREV Chair gathered similar data from six additional practitioners in the focus group referred to above.

The present study thus examines the content of interviews and focus-group discussions with a total of 57 PVE professionals from six different regions of the world and all types of organizations, involved in primary, secondary and tertiary PVE (for details, see Appendix A). To this content we applied an open coding process on various topics, including evaluation. The other topics included coordination among stakeholders, theoretical issues (definitions and approaches) and practices (relationship to religion and gender, security matters and relations with police, funding, etc.) and factors for success and failure.

For the present study, we concentrated on the participants' statements concerning evaluation: questions dealing with examples of successes and failures and the factors contributing to them, as well as their suggestions for improving programs. Other topics addressed in the interviews shed helpful light on the practitioners' evaluations of their own practices. These topics included issues of coordination, the role of terminological issues, practical procedures for conducting evaluations, evaluation of relations with police and intelligence services, and issues of funding.

1.3 LIMITATIONS OF THIS STUDY

Our interviews with PVE professionals, and especially with front-line PVE practitioners, represent a unique source of lessons on how to evaluate practices for preventing violent radicalization and extremism. This material does, however, have certain limitations.

First, the present study does not claim to provide an exhaustive analysis of the evaluation methods that the professionals use themselves. Evaluation of prevention programs was not the main subject of the interviews, and the professionals were actually asked very few specific questions about program evaluation as such. But they were asked several related questions, and these formed the basis for this study.

Second, only 18 out of the 90 interviews were transcribed verbatim; the contents of the remainder were summarized on interview grids.

Third, this study did not look for correlations concerning the evaluation or effectiveness of preventive measures. The fragmentary, qualitative nature of the interview content does not lend itself to statistical analysis (quantifying practices, quantifying effectiveness, looking for correlations, testing hypotheses, etc.); hence this study cannot be used to describe the effectiveness of practices and does not constitute an evaluation itself.

Fourth, this study examines practitioners' perceptions of their own practices, which is relevant in and of itself, but provides only partial, indirect answers concerning evaluation of prevention programs.

This study is also subject to certain epistemological limitations that are inherent in evaluating any anti-terrorism measures (Jackson, 2012). Theoretically, the way to evaluate the effectiveness of a program to prevent political violence would be to have two similar groups, one of which participates in the program while the other does not, and then compare their levels of political violence afterward. However, while such experiments may be possible, the ethical and physical obstacles to them are well known and obvious.⁴ Also, the statistical infrequency of violent events⁵ reduces the significance of each one and therefore limits the conclusions that can be drawn from any experiments of this kind.

To summarize, because of the limitations outlined above, caution should be exercised in drawing any conclusions from this study. As stated before, its primary objective was to give PVE practitioners a chance to make their voices heard, and to gather their impressions about the evaluation of prevention measures and, more broadly, of their practices. We hope that by doing so, we have shed some light on the development of PVE evaluation practices and suggested some avenues for pursuing it.

1.4 AMBIVALENCE AROUND EVALUATION AS BOTH A CONSTRAINT AND AN OPPORTUNITY

One of the first general findings that emerged from the interviews and the focus group was that the subject of evaluation seemed to be conspicuous mostly by its absence. Even when the participants were asked

4 For evidence-based studies, randomized control trials or, if those are not possible, quasi-experimental designs, are regarded as the most robust method of evaluating prevention measures (Hofman, J. and Sutherland, A. (eds.), 2018, pp. 127-137). But these two methods raise ethical problems of consent, confidentiality, access to the best intervention, and so on. They also pose some practical problems. If the program to be evaluated is voluntary, how does one control for self-selection of participants? If it is mandatory, how does one establish a control group?

5 "Global Islamist terrorists have managed to recruit fewer than 1 in 15,000 Muslims over the past quarter century and fewer than 1 in 100,000 Muslims since 9/11." (Kurzman cited by Schmid, 2013, p. 32).

direct questions about it, they either gave their personal evaluations of their own work or described some evaluation methods that had been put in place (essentially questionnaires). But in all cases, they provided very few details, did not really address program evaluation as such, and quickly went back to talking about their own practices. What should be made of this obvious omission? Is evaluation something that practitioners would prefer to avoid? Have the results been unsatisfactory? Do practitioners have objections to evaluations in principle, or simply to the way that they are done? In short, why evaluate?

A second finding, which strongly influenced the structure of this paper, was that when the practitioners did talk about evaluation, they described it in contradictory terms. On the one hand, they were aware of the need for it and its usefulness for improving their practices. But on the other, they also saw it as, by definition, a constraint on the way they did their jobs.

Reflecting this ambivalence, the practitioners were unanimous: they never described evaluation negatively in and of itself, as if being against it would be unthinkable. Moreover, they did not hesitate to judge the quality of their own work. Instead, they consistently painted a more nuanced picture. They explained that in their work, they focus on listening to program participants and encouraging tolerance and critical thinking, which are somewhat hazy, long-term goals. To achieve them, they need long-term financial and institutional support. The practitioners conduct frequent evaluations of program implementation and participant satisfaction. They underscored the inherent limitations of their interventions and readily acknowledged that they could do better in many respects. They cited the sometimes excessive expectations placed on them and their need for more resources to produce better results. Immediately, evaluation was seen as a practice space where various expectations are expressed between the strategic level and the operational level. It was also seen as a space for the expression of hierarchies and power relationships within the organization, between competing organizations and between the sponsor (see [glossary](#)) and the funded organization, even though this argument was mentioned very seldom.

In reality, although the practitioners did not necessarily address the subject of evaluation directly, they did delve into it at length in their interviews. As the practitioners themselves observed, many constraints interfere with the presentation of reliable, significant results. The practitioners very clearly identified the same epistemological challenges that researchers have—what to measure, when, and how, and with what degree of reliability—albeit in less scientific terms. Even researchers are not unanimous about the best way to

define the boundaries of evaluations. The dimensions that they examine—such as outputs (see [glossary](#)), outcomes (see [glossary](#)), satisfaction and recidivism—vary widely in their measurability and their relevance for public policy, and potential evaluation models vary accordingly.⁶ The large number of models reflects the lack of consensus on some fundamental elements: explicit criteria for success, independently verifiable data, and a systematic evaluation of one or more programs (Horgan & Braddock, 2010, pp. 285-286). According to Schmid (2013, p. 49), the choice of the term “good practices” rather than “best practices” in the title of the Rome Memorandum on Good Practices for Rehabilitation and Reintegration of Violent Extremists indicates that it is not yet possible to evaluate existing practices comparatively. But the absence of “best practices” that are valid for everyone is not necessarily a defect, because one of the few robust findings regarding PVE practices is how much they depend on cultural, social and political context.

Another, underlying consideration is that evaluations are used by funding sources (usually governments) as a tool for imposing standardization on sometimes widely differing practices and forcing them to conform to an overall strategy that will supposedly ensure consistency and legitimate use of funds in national or international programs. But practitioners are not mere passive recipients of new constraints. They too can use evaluation as a tool, to objectify certain findings, to provide empirical support for certain proven practices, and to reject certain unrealistic expectations. In this context, although practitioners may at first perceive evaluation as an additional burden or even a professional risk, they may also come to see it as a way to make policymakers realize their value and understand their needs.

Thus, as described in Part 2, PVE practitioners may initially see evaluation as an external constraint in a context of institutional dependency. However, as will be seen in Part 3, evaluation also represents an opportunity for practitioners to establish specific standards, practices and qualifications and to expand their toolkit so as to improve their practices. One tool endorsed by virtually all of the practitioners is the adaptation of PVE programs to local conditions (see part 4).

⁶ For example, Hofman and Sutherland (2018) propose no fewer than 24 different models.

Institutional constraints and practical challenges for PVE program evaluation

The first way that PVE practitioners see evaluation is as an additional constraint, or even as tool that sponsors—in most cases, governments—use to control them. Evaluation, of course, costs additional time and resources for organizations that are already struggling to find funding and attract talent. Evaluation is also often seen as an audit of professional practices, or even an attempt to impose conformity on them. On the other hand, national and international program staff do need to show policymakers and the public that the resources allocated to prevention are being put to optimal use and producing the desired results. Consequently, there is pressure for funded programs to obtain positive evaluations.

But what should these evaluations measure: decreases in violent behaviour? decreases in tolerant attitudes toward violence? satisfaction of program participants? program implementation? Depending on what indicator is used, the evaluation result will be different. Some indicators are quite simply impossible to create or give credence to—for example, indicators to measure deradicalization, which is the abandonment of a radical ideology. Evaluations are also highly time-dependent. For instance, a reduced appetite for ideology immediately after a program ends is meaningless unless it continues to be measured and is found to persist over the long term. PVE practitioners are well aware of such issues and are developing various evaluation methods to improve their practices incrementally, in a manner suited to the setting in which each program is delivered.

In this context, evaluation thus becomes an instrument of bureaucratization: at a minimum, it helps to codify the actors' professional practices, and at a maximum, it imposes uniformity on practices that previously varied widely.

Evaluation casts a spotlight on two issues that are always sensitive: the social role that PVE practitioners perform and the infrastructure that supports it. In their social role, practitioners face the difficult task of fitting into a vast social-engineering project and building ties with the other actors and the targeted communities (see [glossary](#)). But they must also maintain relationships with police and intelligence services; the nature of such relationships can be ambiguous, but they are necessary for coordinating measures in an integrated approach.

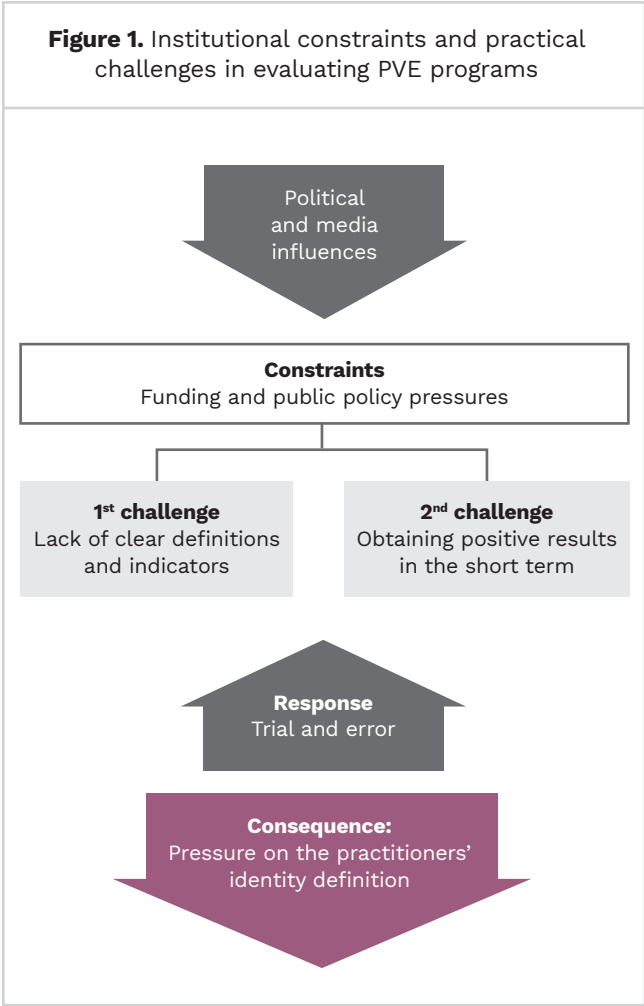
The infrastructure supporting practitioners is often insufficient. There are no consensus definitions of key concepts such as terrorism, radicalization and their derivatives, whose meaning is simply taken for granted. The PVE profession itself is weakly structured and relies heavily on practitioners' ability to draw on their knowledge of related areas (for instance, models for crime prevention⁷ and social cohesion).

Under these circumstances, evaluation represents a set of constraints that add complexity to a kind of work that is hard to evaluate. In principle, the work of PVE might even seem incompatible with evaluation, but as numerous

⁷ For example, a research project on increasing the evaluation capabilities of security services (ACESS, 2017-2018), conducted in France but including comparisons with Quebec and French-speaking Belgium, provides a framework for evaluating crime-prevention programs. Aspects of this framework might be relevant to and inspire the development of a framework for evaluating programs to prevent radicalization (Wuilleumier & Delpuech, 2019).

formative (self-)evaluations (see [glossary](#)) have shown, PVE practitioners are not opposed to the idea of improving their practices. The fact remains that evaluations impose a burden on an organization's operations and finances and are conducted over too short a time frame to provide reliable results. Lastly, even though PVE practitioners do their work in the field, they are well aware of the inevitable theoretical limitations of evaluations in the absence of a consensus on the definition and objectives of PVE and the indicators that should be applied to it.

To capture all of these issues, we have developed the model shown in Figure 1, which explains the constraints that practitioners face when an evaluation is to be done, the challenges that they have to meet, their responses and the consequences for their professional identity. In the following sections, we explain each of these elements in detail.



2.1 INFLUENCE OF POLITICS AND THE MEDIA

The issues involved in evaluating PVE programs are highly complex. For one thing, because PVE is a relatively young discipline, the lack of empirical data confirming its effectiveness tends to produce negative assessments that make practitioners mistrustful of the entire evaluation process (Madriaza, Ponsot, & Marion, 2017, p. 107). For another, PVE program evaluations are vulnerable to being diverted from their original purpose in order to serve political or bureaucratic ends.⁸ As one practitioner put it: “Evaluation is no longer a constraint, but a fulcrum” (Madriaza, Ponsot, & Marion, 2017, p. 94). The worst risk is that instead of evaluating an evidence-based policy, evaluators will go looking for policy-based evidence. Thus evaluation is not only a process for optimizing public policy, but also a public policy that demands scrutiny in and of itself. Practitioners are aware of this fact and readily point out that PVE evaluation takes place in a highly politicized, highly mediated context that raises various issues about evaluation as a process.

The only meaningful way to judge the effectiveness of PVE programs is in the long term, because the only meaningful indicator that an individual has renounced violence is an extended period without recidivism. But PVE programming is heavily influenced by shifting political priorities, which in turn are affected by shifts in media attention over the very short term. Evaluations conducted under such conditions may be distorted in a variety of ways.

In particular, short-term pressures may cause measurements of success to be inflated artificially, because they are taken before enough time has passed for recidivism to occur. The risk in this case is that too much trust will be placed prematurely in deradicalization entrepreneurs claiming overstated success rates for their “solutions”. Though the practitioners whom we interviewed did not refer to such businesses, they are a documented reality (Schmid, 2013, pp. 46-47).

The practitioners did, however, mention two types of problems that result from differences between their priorities and those of policymakers and the public and that make evaluations more complicated. The first of these problems is the way that politicians and the media make use of evaluations. Policymakers need to show that have met their political commitments to take action against political violence and can use evaluations as a tool for this purpose. As one practitioner from a police agency put it:

⁸ An example of Campbell's law (1979), which states: “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.” In other words, practitioners may perceive evaluations as a tool for monitoring or promotion, which gives them an incentive to overstate performance or justify increased resources. In addition, some prevention practices may persist whether or not they are actually effective, because of various biases associated with cognitive dissonance (persisting in certain beliefs despite contradicting evidence, overestimating results to justify the efforts made to achieve them, or believing a statement is true simply because it has been repeated so many times).

“ Sometimes it’s harder to deal with the [people at] the political level because [they are] not always guided by facts and knowledge. They tend to be guided by values, [which] are not always consistent with what’s needed to solve the problem. [...] But we’ve been blessed with political backing almost from the beginning, because it protects us from national criticism.” (EUR06)

In other words, if evaluations are used mainly for political messaging, they will produce false findings unless they are conducted as objectively as possible. On a highly sensitive subject such as political violence, evaluations must draw a not-always-obvious connection between public concerns and intervention practices:

“ [One of the challenges] is that our work is very much linked to political debates. Each attack in Paris, Brussels or right-wing attacks against mosques directly affects our work, since we want to show Islam as normal, part of [our country’s] culture. Surveys shows clearly that the rejection of Islam as a religion is very popular in [our country].” (EUR19)

Also, when support for a program fluctuates along with political priorities, that can have an impact on the program’s content. If an attack occurs after a program has been implemented, there can be the temptation to conclude that it was ineffective (in one such case, a practitioner reported their program was made “the scapegoat for the government and for terrorism.” (EUR10)) One can then expect to see a hardening both in the policy line and in the kind of programs that receive support:

“ At the new deradicalization centers, we give citizenship courses where we talk about things like republican values and laicity [secularism]. The assumption is that if a young man hasn’t fit into society, it’s because he hasn’t understood what the Republic means. But because of the change in the political climate, our contract didn’t get renewed.” (EUR10)

Obviously, no program can be expected to be totally effective, but it is still hard to defend a program to the public under such circumstances. On the other hand, if we keep switching from one model of PVE to another, what are we going to evaluate? How are we going to compare the results of different programs?

The second problem that the practitioners mentioned concerns the timing of evaluations. Program calendars raise evaluation expectations that are out of sync with the results that programs can reasonably be expected to achieve. Programs are often designed to last one year and

to meet very specific goals. But many practitioners regard such short periods and such definite goals as unrealistic:

“ It’s an ideal that is hard to achieve ethically, financially and in terms of timing as well. I think it would make more sense to examine young people’s trajectories over 10 to 20 years and see what programs they have participated in. Then you might conceivably have something that resembled a real impact evaluation.” (EUR28)

Obviously, for donors, such a long time horizon is unthinkable. As one practitioner put it, “The problem is that a lot of donors aren’t able to commit to long-term funding ... The nature of the work [we do] tends to require longer projects to measure impact, which they aren’t able to do in the way they want in a 12-month project.” (ASI04)

In highly destabilized countries, it is hard to see how a program can demonstrate its effectiveness when the level of violence prevents it from even being implemented:

“ Political volatility and the security situation are and will continue to be a huge challenge to ensure timely implementation of programs. [...] Money has already been spent for the activity so they have to answer to the donors because they have spent the money but didn’t do anything.” (ASI04)

To sum up, this fixation on the short term, both in political priorities and in funding schedules, is incompatible with the time needed to conduct valid assessments of PVE programs. The practitioners deplored this tendency to set tight timelines to meet messaging needs (“Our government has acted to counter radicalization.”) or to satisfy the bureaucracy (“Have the procedures been followed?”). In this regard, program evaluations would be more worthwhile if they did not have to deal with any considerations not directly related to prevention. One way would be to make sure that a single, external evaluation (see [glossary](#)) is not the only evaluation that a PVE program receives. It would also be ideal to evaluate the trajectories that program participants follow over the medium or long term.

2.2 INSTITUTIONAL CONSTRAINTS ON FUNDING ACCESS

PVE practitioners’ first reaction to political and media influences on program evaluation is to see it as an institutional constraint. In the most practical terms, this constraint affects organizations’ ability to access the funding they need to deliver PVE programs.

One practitioner reported that “money is definitely a challenge” (EUR18), while another said that “the government wants to give as little money as possible

and include as many things as possible in the budget.” (EUR06) The logic that guides funding also creates hierarchies among organizations; as one practitioner put it, “Grassroots initiatives get little money, but big institutionalized youth organizations get most of [it].” (EUR03) In this context, PVE organizations naturally perceive evaluations both as an additional cost and as a high-stakes game on which their survival may depend.

This precariousness results from the way that PVE organizations are funded. Like many other kinds of organizations, they have to secure funding on a project-by-project basis (EUR26), and projects live and die according to whether they get funding (SSA06). In an environment of constant pressure to rationalize public spending:

“...projects become the primary means by which institutions demonstrate their worthiness and respond and adapt to the uncertainties of changing economic and political conditions. Many organizations have become obsessed with designing projects as a way to justify their actions.” (Paturet, 2002, p. 65)

The project-based methodology divides human activities up into discrete packages that are bounded in space and time (generally one year), any one of which can be brought to a halt or replicated, depending on its results. Projects were designed to meet the needs of business, and some question their undesirable consequences, and whether it is really appropriate to apply them to public services where there is no commercial motive. The project approach has at least two effects that undermine the sincerity of evaluations: it forces programs with widely varying goals and practices to be evaluated according to standardized frameworks designed outside the organization, and it generates pressure to produce positive evaluations.

First of all, the project methodology affects the kinds of evaluations that get carried out. Their objective is to provide a “rational” basis for decisions to continue or end programs. They thus constitute summative evaluations (see [glossary](#)), quite differently from formative evaluations, which are still the kind more commonly used among PVE practitioners. Summative evaluations are designed to determine whether public funds are being put to good use, by verifying whether programs were implemented as planned (NA02). Such evaluations put pressure on practitioners to take practices that are ill suited to quantification and translate them into goals and success criteria. As one practitioner put it: “Any institution that grants funding also imposes some conditions. I try to impose my choices and hope that half of them will accepted.” (MENA02) But this does not always seem to be the case, and the pressure is not uniform. Practitioners MENA01 and MENA03, for example,

said that no one attempted to dictate the content of their programs or their practices in exchange for funding. When the pressure does exist, indicators and metrics are “... one of the things donors are very particular in looking at, because if that system is in place, it helps them to recognize that the project is well thought out.” (ASI04) In addition, the formality of the evaluation exercise imposes an administrative burden on organizations. One practitioner who coordinates the activities of several organizations recognizes this fact: “We have to stimulate the organizations but not overburden them.” (EUR23) Does such a development of procedures indicate an improvement in program evaluations? One practitioner clearly stated that the evaluations expected were unrealistic: “When we started [this program], the government wanted us to evaluate all of the young people individually. I told them it was impossible.” (EUR03)

This last quotation suggests that standardized evaluations are not something imposed by direct, top-down orders, but more a matter of general standards and objectives. It is up to the organization to interpret them and flesh out the details to satisfy the sponsor (SSA03, SSA04). The practitioners have leeway to negotiate a balance between satisfying evaluation criteria and carrying out the actions that they consider necessary:

“We want to make sure to do everything the government wants from A to Z, but at the same time we have no choice but to adjust to the young people’s needs. So for sure when I do my evaluation this year, I’m going to be much more focused on the young people in the field, on knowing what’s going on there and what they need.” (NA09)

Second, evaluations represent a matter of financial survival in an environment of limited resources. The participants in this study, in particular those in the focus group, dwelt on this subject at some length. One practitioner in particular was quite explicit about how evaluations serve the sometimes contradictory objectives of informing and persuading. She said that at one meeting whose purpose was to present a program’s results, compare them with its objectives, and draw relevant lessons:

“We tried to comment together and draw lessons, but in my opinion it wasn’t really enough to evaluate the work done. The reality was that we were at the end of the fiscal year and we had to justify the use of this money. I asked a lot of questions about the evaluation process, and in the end, I got the impression that I was annoying people. I understood the truth was that we had been given very little money and we had to say that we had achieved results anyway.” (SSA09)

Perhaps these unvarnished remarks reflect a situation that was more difficult than most. But other practitioners

whose programs were apparently subjected to more structured evaluations (of implementation, or outcomes or participant satisfaction) described a similar dynamic:

“The first problem is that evaluating overall impact isn’t easy. The second, of course, is pressures related to funding. No matter how good or bad a project may be, the biases related to funding strategies come into play; the operator needs funding and is going to evaluate the project in the most favourable possible light.” (EUR27)

The point here is not to suggest that systematic deception is going on, but rather to show that when various actors are competing for limited funding, they will be tempted to give the persuasive aspect of their evaluations precedence over the informative one. It is hard to see how evaluations, which impose additional costs, can be conducted rigorously and exhaustively if no additional resources are earmarked for them: “There is also the financial issue, where we try to ‘please’ to obtain as much funding as possible, because for community agencies, that’s difficult.” (NA09)

In conclusion, evaluations are used as a tool to impose a framework on PVE programs as a condition for funding them. Evaluations can thus jeopardize the very survival of organizations with shaky finances, tempting them to embellish their results while potentially giving large, institutionalized, professionalized organizations an advantage over them. In the absence of robust, independent criteria, evaluations may end up measuring an organization’s ability to meet existing criteria more than the actual effectiveness of its program. In order to distribute the costs and risks of evaluations more fairly, it may be advisable to divide the evaluation tasks among the various stakeholders. Also, to give small local organizations a chance, a portion of the funding could be reserved for evaluations.

In addition to these institutional constraints on access to funding, organizations face some practical challenges in evaluating PVE programs.

2.3 LACK OF ADEQUATE DEFINITIONS AND INDICATORS FOR EVALUATING PVE PROGRAMS

The PVE professionals who participated in the interviews and focus group were well aware of the continued lack of consensus on a definition of radicalization. As one put it, “We’re starting from a point that I’m not really clear on” (EUR25), while another explained that he had “a personal disagreement with the term radicalization.” (EUR24) How can anyone be expected to evaluate success in preventing radicalization if they cannot even define it?

Given this lack of clarity, concern for accuracy and

scientific consensus is superseded by various forms of pragmatism. Several of the practitioners believed that for all practical purposes, they did not need a definition. As one put it, “They have conducted different research on the concept, but they don’t have a fixed definition. ... With radicalization, you don’t need to have a definition.” (EUR01) A number of other practitioners preferred to adopt the definitions provided by governments (EUR02, EUR03, EUR14) or by police and intelligence services (EUR06, EUR07). This was to be expected among practitioners who worked for government organizations, but several practitioners working with community organizations made this same choice.

This uncertainty surrounding the definition of radicalization has profound consequences both for the programs implemented and for the expectations that evaluations should place on them. The fundamental question really comes down to whether programs should focus solely on treating behavioural radicalization, or whether they should address cognitive radicalization as well.

Most of the practitioners expressed uneasiness at the prospect of treating ideology. Several pointed out that some values now regarded as central to democracy, such as women’s right to vote, were once considered radical. Hence, as one practitioner put it, radicalization can be “something that makes society advance.” (NA01) The practitioners who believed that prevention programs should address cognitive radicalization were in the minority; most of them came from Germany, Switzerland or France. Only three of the practitioners (MENA02, MENA04, ASI06) judged radicalization in moral terms. The rest endorsed the definition that “radicalization was the gradual acceptance of violence or undemocratic means to further your agenda,” because “we are not the thought police; we are preventing them [from doing] something violent” (EUR06), from doing “harm [to themselves] or others.” (ASI01) They were aware of their own biases and so were reluctant to take a stance of ideological superiority, “First, because it’s hard to change other people’s ideas, and second, because we don’t feel we have the legitimacy to do that.” (NA01) Whether it is a matter of effectiveness or legitimacy, “They are not trying to say which ideology is important or not.” (EUR02) For them, it is therefore ethically and logically unthinkable to evaluate PVE programs according to their ability to change the participants’ beliefs.

The practitioners also underscored just how much they thought that “wanting to change their ideas and convince them to think differently is absurd” (MENA04) and that “talking only about the topic of radicalization will be boring for the participants.” (SSA07) More seriously, if practitioners even use the word “radicalization”, participants will trust them less and give them less credence. Instead, practitioners prefer an indirect approach: promoting tolerance and non-violence (ASI04, SSA05). In highly religious societies, prevention programs

can use this approach to defuse violent interpretations of sacred texts. Some practitioners involved in training programs for imams said that they were more open to the case for non-violence when it was supported by a traditional reading of the Koran (SSA05, SSA07, MENA03) or a demonstration that extremists were misconstruing it (ASI06).

We thus see that the most of the practitioners rejected the notion of deradicalization, either on philosophical grounds or because they considered it ineffective for working with PVE program participants. Instead, the practitioners prefer the concept of disengagement from violence, something that they believe it is feasible to evaluate, whereas they consider evaluating deradicalization very difficult if not impossible. Measuring effectiveness in changing participants' ideology would be a hazardous, ongoing, long-term undertaking. The contextualization of the intervention thus appears crucial for formulating arguments that the participants can hear. Once again, a localized evaluation tailored to the practitioners' actual scope for action is essential.

Beyond the matter of definitions, the practitioners operate in a climate of uncertainty about their ability to measure and prevent the risk of violence. Again, the lack of robust knowledge often deprives the practitioners of a list of factors that increase or decrease the risk of turning to violence, against which their interventions might be directed and measured. Many programs claimed as success stories by certain countries have turned out to be less impressive once they were subjected to external evaluations. When programs have been less successful, that has not always been made transparent. Lastly, in many cases, evaluation remains a luxury (Schmid, 2013, pp. 47, 52-53). In reality, "When it comes to de-radicalisation/disengagement and counter-radicalisation, [we conclude] that it is difficult to identify what works and what does not work in general, or what is even counter-productive." (Schmid, 2013, p. v)

Many programs present themselves as promoting tolerance and non-violence, which suggests that the practitioners accept the premise that grievances lead to violence. Causes of radicalization that the practitioners mentioned included economic and ideological factors, cultural stagnation and lack of information (MENA06), as well as social pressure on young men to be providers even when opportunities to work are scarce (SSA03). However, these widespread and seemingly plausible assumptions are not always supported by scientific sources and reliable data. As one practitioner put it, "We also assume that trauma and discrimination are the great trajectories, but [in fact] they are not, according to our research based on the data [that] we have." (NA04)

The practitioners did, however, express the desire to have reliable data to help make their interventions more effective: "If we understand the causes, we can target the roots of the problem." (SSA03) Evaluations can show when programs have been designed without enough input from researchers who might provide reliable hypotheses and data about prevention. As a corollary, evaluations can help to establish the need for more robust tools so as not to impose undue expectations on practitioners.

The practitioners are not asking for violence-predictor checklists: "CVE at the local level [has] to use evidence-based research to understand how and why the narratives of extremists work and how to counter them." (SSA07) Without such instruments, practitioners are forced to take generalized actions, with no clear direction: "preventing radicalization is not something that you can attempt at one point in time; it's a process of transmitting values that children internalize and grow up with." (MENA03) That picture is probably accurate as far as it goes, but it offers little in the way of operational guidance. When asked how their actions should be measured, the practitioners from police services joked about their top criterion: "Number one, we always joke that it's the fact that nothing has gone 'boom.'" (NA06) But joking aside, "If you're talking about primary prevention, where the program addresses a very broad population, the only outcome that you should expect is that nothing happens," which legitimizes the importance of "having an external perspective for a comprehensive evaluation of either the impact or the entire process." (EUR26)

Practitioners undeniably prefer to work on disengagement (changing behaviour), which is more readily achieved and measured, rather than on deradicalization, which involves changing ideas and is hence more problematic, at least in democratic societies. In the absence of clear, localized indicators, any evaluation expectations are just constraints that do not help to improve the quality of the service provided. Such indicators could, for example, be developed using a tool such as the Influencing Violent Extremist Organizations (I-VEO) Knowledge Matrix.⁹ This matrix used current knowledge to evaluate the effectiveness of 183 measures aimed at weakening violent extremist organizations. Out of the 32 measures most strongly supported by empirical findings (with scores of 7 to 9 out of 9), 12 suggest the usefulness of prevention programs in the broad sense, confirming the widespread intuition that democracy is effective, repression is counterproductive, and appeasement produces ambiguous results. Prevention programs can directly operationalize one of the most robust findings, namely that positive incentives are more effective than negative ones for achieving deradicalization or disengagement. More broadly, institutions that are pluralistic, legitimate,

9 Developed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland (<https://www.start.umd.edu/data-tools/iveo-knowledge-matrix>). The measures mentioned here are numbers 114, 69, 39, 95, 172, 136, 30, 7, 8, 150, 140, 181 and 152.

effective, and redistributive weaken violent extremist organizations, while brutal, indiscriminate repression does not. Conciliatory actions reduce the probability of attack, but negotiations stimulate sabotage. The UNESCO-PREV Chair is currently working on a web-based platform that will let users apply certain characteristics to search for the evaluated programs that this Chair has identified. This platform will include a section on indicators.

The PVE professionals in the current study made their position on evaluations quite clear: they are desirable in principle, but they have to be done right. They must take the qualitative dimension of PVE practice into account, which involves trials and experiments that are sometimes inconclusive. Evaluations may encourage deceptive practices that distort the results and may have little meaning unless there is continuity in the support for the prevention programs and in the prevention measures themselves. These problems are aggravated by the lack of clear, realistic goals and criteria. But evaluations can also be a useful tool that helps practitioners communicate their concerns to program sponsors while developing their standards, competencies and practices.

2.4 TRIAL-AND-ERROR: FORMATIVE SELF-EVALUATIONS AS CENTRAL TO PVE PRACTICE

In the situation just described, while awaiting external evaluations that are still infrequent, the practitioners conduct qualitative self-evaluations for the most part.¹⁰ These can be regarded as a kind of formative evaluation, providing continuous learning through feedback from the field. In the absence of well established protocols, the practitioners correct their practices whenever they identify a problem. Regarding success criteria, they state that learning through experience is important (EUR04), and that they are always learning from one another (ASI01). “We’re always learning by doing. So there’s no academic approach. We develop our approach by experimenting, and we can always draw connections between our own discoveries and more macro analyses.” (EUR15 and EUR17) This approach actually makes some sense scientifically: exploring through trial and error is a typical method of solving problems in the early phases of knowledge development. As one practitioner explained, “One of the indirect principles is that it is trial-and-error; therefore the government is aware that some methods will fail, so the advantage is that you can actually try.” (EUR19)

The PVE measures whose effectiveness has been best supported by scientific evidence are consistent with what general intuition would suggest: rewards

work better than punishments; targeting the general population or individuals at risk works better than targeting a community; and eliciting empathy for victims of violent radicalization works better than deconstructing its ideological foundations. The measures for which the evidence of effectiveness is most robust (scoring 9 out of 9 on the I-VEO Knowledge Matrix) suggest a balance between extensive use of rewards (soft incentives) and limited use of punishments (hard incentives). Measures for which there is moderate corroboration (scores of 4 or 5 out of 9) include active, voluntary enrolment of the organization’s leaders; exposure to other individuals and ideas; and long-term psychological support (especially because disengagement often occurs following a traumatic event). In primary and secondary prevention programs, practitioners have achieved positive results by stimulating empathy for victims of terrorism and thus building resistance to extremist ideas (Gielen, 2019, p. 12). However, stimulating cognitive and ideological changes remains a very uncertain approach and does not necessarily lay the groundwork for behavioural change (Pistone, Eriksson, Beckman, Mattson, & Sager, 2019, p. 20). Lastly, targeting Muslim communities as populations at risk is potentially counterproductive (Lindekilde, 2012, pp. 395-396; Gielen, 2019, p. 9).

Conceptually speaking, the constraints cause practitioners to give priority to two largely qualitative types of evaluations: implementation evaluations (see [glossary](#)) and impact evaluations (see [glossary](#)). Implementation evaluations are done to verify that funding has been properly used (EUR26) in accordance with the plan submitted to secure it. Impact evaluations measure the quantifiable outputs produced by programs, such as changes in school-enrolment rates (SSA05, SSA06) and number of participants in program activities (NA09). But the practitioners feel that whether such programs produce any particular outputs is less important than whether they produce the desired outcomes. Hence many practitioners try to measure, with the limited resources available to their organizations, how satisfied participants are with their programs and what cognitive impact these programs have on them. The major obstacle besides resource limitations is the limited availability of the participants. They are hard to reach once an activity is over (SSA09), both the activity and the outcome measurement take place at only one point in time (EUR27), and only the most motivated individuals participate in the activities (SSA10).

In one experimental protocol, a practitioner compared the results for her program’s participants with the average numbers that she had found in the literature for violent offenders in general (having found no separate figures for ideologically motivated offenders) and reported that

¹⁰ Some of the practitioners (OCE01, ASI03, ASI04, SSA01, SSA05, SSA06, EUR05, EUR20, and MENA03) referred to quantitative evaluations (including indicators and statistics) conducted by those who had the necessary time, funding and skills, but these were still in the minority.

her program's participants "were doing much better." (EUR20). But this protocol was specific to this one organization and this one program, which had qualitative and quantitative indicators and internal and external evaluations (see [glossary](#)) and so readily lent itself to a detailed, rigorous assessment.

One final specific aspect of PVE programs explains why rigorous evaluations of them are so scarce: the theoretical design of prevention success indicators is not very applicable in the field. As one practitioner explained, "Prevention is hard to measure. We have found very few evaluation criteria that we considered really applicable to our work, so we have done a lot of self-evaluation." (EUR16)

Self-evaluations of course raise concerns about bias, because practitioners are judging their own work, but other evaluation approaches have their own weaknesses. Internal and external evaluations are subject to bias too. One way to achieve some objectivity might be to compare multiple partial evaluations whose sources of bias are known. A "cross-evaluation" of this kind would have the benefit of being conducted by all of the stakeholders working together—front-line practitioners, program designers, donors and researchers.

One of the researchers interviewed is attempting to create a psychometric scale that could be used to measure radicalism. Practitioners would capture the necessary data by filling in questionnaires as they interviewed program participants. Practitioners could also use this tool to determine the degrees of engagement in and disengagement from radicalism by administering the questionnaire to participants at two points in time and measuring the change (EUR12). This particular research is being conducted in partnership with associations that deliver PVE programs and are in contact with participants. Such initiatives could help to address some of the challenges involved in evaluating practices in the PVE profession.

2.5 RESTORING THE SPECIFICITY OF THE PROFESSION: EVALUATION AS A PROFESSIONAL CHALLENGE

"We're still just speculating and groping for ideas." (SSA09) This statement shows how much PVE evaluation presents practitioners with the problem of defining their modalities of action and the outcomes expected from them. Are they trying to instil a spirit of pluralism in the population? Many practitioners see their role as very broad; they believe that they have to counter the radicalization process "as far upstream as possible." (EUR07) This work will involve strengthening democratic cultures and structures (EUR21), teaching young people democracy (EUR03), teaching children about empathy

and compassion "from the day they are born" (NA07) and "teaching conflict resolution from kindergarten." (EUR24) But some practitioners temper such ambitions with the voice of realism: "We don't really set it as a goal to create super-democrats. We just want to make them less prone to violence." (EUR06) PVE clearly does involve teaching non-violence starting in childhood, but this last comment underscores the limitations of such efforts. One cannot judge a program from its ability to produce good citizens: that may be a worthy goal, but as an outcome, it is hard to measure, and the burden of achieving it cannot rightly be placed on PVE practitioners alone.

In other words, evaluating PVE presents the challenge of quantifying practices that are essentially qualitative, custom-tailored, and hard to replicate. Although PVE practitioners' mission consists, to some extent, in reducing individual behaviours that are deemed antisocial, it also frequently involves building social ties and reducing polarization within society, which makes this mission quite complex. The practitioners themselves question their own grasp on the problem and especially their ability to solve it: "We are dealing with a phenomenon that cuts across all of society, and our approach is not going to solve the problem. Socio-economic or religious approaches are not going to solve the problem." (EUR15 and EUR17)

The mission just described demands sincerity and spontaneity, two qualities that are hard to capture in an evaluation and that can even be said to contradict the very principle of evaluation. PVE depends on interpersonal skills that do not fit into neat checklists. The practitioners stressed how vital they consider it to build open, empathetic relationships with program participants. Naturally, they resist breaking such relationships down into processes: "How do you measure a qualitative process? You're building relationships, you're building trust, how do you measure that?" (NA06)

Many practitioners underscored the importance of meeting young people's need to express themselves. As one practitioner put it in his own evaluation, "90% [of my participants] were thankful for the open ear." (EUR22) But participant-satisfaction surveys (which are not very common) can provide only a fragmented picture of a program's effectiveness, and the problem is far broader. When asked to describe her greatest challenge, one practitioner said that it was getting all of the stakeholders to talk with one another:

“Getting the government, the religious leaders, the parents, the young people, the representatives of civil society and [the practitioners] to all sit down around a table and state the problem of radicalization of youth in a clear way. That's the challenge. Everybody knows it, but nobody says it out loud.” (SSA05)

Even if one imagines a more modest study comparing propensity to engage in violence among PVE program participants and control groups, such programs use methods that would be hard to quantify: “How do you measure when you prevented somebody from engaging in violence? I’m not aware of something that we’ve developed that can show us that.” (NA08)

Another issue is that PVE programs are still too new to have resolved the many remaining questions about PVE practice as a whole. How do you determine what actions work when “you need to have more than 10 years to tell whether it works or not” (NA01)? It can be hoped that as

PVE programs gain experience and maturity, indicators can be defined, but that remains uncertain. In this regard, practitioners are not opposed to evaluations, but they are encouraging or demanding external evaluations (EUR19, EUR06, EUR08, EUR26), both to relieve their organizations of this burden and to differentiate the tasks of program evaluation and program delivery (being a good practitioner does not necessarily mean that you know how to evaluate your own practices properly). One practitioner described the situation as follows: “We can’t afford to do real evaluations, and it’s difficult to do academic evaluations. We’d want external evaluations, but we don’t have the budget and [our work force] is limited.” (EUR06)



Evaluations as opportunities for improving PVE practices and autonomy

Evaluations unquestionably place cost and other pressures on organizations that deliver PVE programs. However, evaluations also represent tools that PVE practitioners can appropriate to improve their practices and even enhance their own autonomy. Governments lack the credibility to address persuasive arguments against radicalization to the audiences that might need to hear them. The state therefore needs volunteer and community groups to carry out action plans that would otherwise be seen as imposed from above in service of some official ideology. The PVE professionals in this study showed that they understood this social function: “Young people are usually considered in our societies [as] part of the problem or [as] troublemakers. They have never been considered as a part of the solution and as equal partners.” (ASI03)

In this sense, evaluations give practitioners the opportunity to show sponsors the factors that are the keys to success and to demonstrate the ongoing need to provide greater autonomy and coordination and increase their resources and capacities. These capacities may be tangible as well as intangible. For one thing, good evaluations can demonstrate an organization’s relevance and hence the need to renew its funding. For another, evaluations can present the practitioners’ ideas and values persuasively, demonstrating the specific attributes of the organization and its practices, the value that they add, and hence the central role that they can play in efforts to prevent violent extremism.

Thus, the evaluations conducted by the practitioners highlight these specific attributes and needs. Most of these evaluations are formative, which can justify the

continuation and improvement of a practice. They avoid concluding that a program, as a whole, needs to be terminated, as a summative evaluation might. The two most common types of evaluations—impact evaluations and implementation evaluations—can model the type of information that comes in from the field. More specifically, impact evaluations can highlight the activities carried out in the field. This community focus lets practitioners show policymakers that top-down approaches are doomed to failure; it also lets them demand a certain autonomy in their work. Similarly, implementation evaluations can be used to promote a certain conception of the intervention that argues for sharing responsibility among all stakeholders. This includes sharing responsibility not only for potentially negative assessments but also for criticizing state violence in conflict settings. Satisfaction evaluations (see [glossary](#)) and needs evaluations (see [glossary](#)) afford opportunities to argue for localizing PVE programs—not only adapting them to local needs and conditions, but also giving organizations the autonomy to deal with them.

Thus evaluation is not merely a tool for control and standardization, and the choice of what type of evaluation to conduct is not driven solely by constraints in terms of resources, indicators or volunteer participants. This choice also reflects the aspects of the profession that are valued and that are directly useful to its practitioners. The issue for them is to strengthen their professional identity, which is still a delicate work in progress.

This issue is manifested in at least two ways. First, evaluations provide a means of underscoring just how much the problem of violent extremism surpasses

the abilities of practitioners to deal with on their own and requires a coordinated definition of each actor's responsibilities (section 3.1). Second, evaluation provides an opportunity to address the issues of training and retention of practitioners who must be able to see themselves making careers in the PVE profession (section 3.2).

3.1 EVALUATIONS PROMOTING A COORDINATED DEFINITION OF EACH ACTOR'S RESPONSIBILITIES

As the practitioners in this study confronted the immense complexity of the phenomenon of radicalization, they were well aware of the difficulty of obtaining and measuring their results. They were also aware of the need for multiple actors to work with program participants and with communities. (Formative implementation evaluations are an effective tool for this purpose.) The practitioners saw government agencies as necessary partners whose role must be well defined: "We engage first with the community to gain trust and support, then [with] the police, and we also work with other organizations. We need to work with the government." (SSA02) Under these conditions, the practitioners see the government's coordinating role as a factor for success: "This success was the result of our having taken an 'all-of-government' approach." (EUR06)

Basically, the practitioners think of PVE as a combination of doing social work and promoting democracy. Aware of the size of the task, they do not see themselves as the solution, but rather as one link in a longer chain. Evaluation is also a way of sharing responsibility among all of the stakeholders. By evaluating how a PVE program is implemented, and in particular how the roles of the various actors are coordinated, the commitments and responsibilities of each of them can be better defined. It would be unfair to make practitioners bear all of the professional or moral responsibility for the failures that programs will inevitably experience, as was shown by the attack in Vienna on November 2, 2020.

The first place where such coordination is encouraged and occurs is between organizations with similar missions. For the practitioners, being evaluated by a fellow professional who is familiar with the realities of their work lends legitimacy to the process: "We have monitoring evaluations, as well as with peer organizations. We also discuss with their directors, we review about the strategies used." (SSA02) Evaluating the implementation of a program also entails defining the setting in which it is delivered, and in particular the other actors with whom the program works: "We have to do deep research on [the] needs and demands and interests of each religious group and the positions of each Muslim community." (ASI05) Evaluation of coordination, however, is still in its infancy, and the right balance has yet to be found. If too

few actors are involved, each of them ends up working alone. But if too many are involved, then coordination dilutes responsibility and impedes the intervention:

“ We could be better at involving the community to help them build these relationships. A lot of things that could be done by the community in terms of social welfare are done by the government. Basically, so much help from the government that people don't help each other in the community. We would like to mobilize that, but we haven't been good at doing it. We are reactive; we would like to be proactive.” (EUR06)

The sphere in which each of the actors operates therefore needs to be defined. Implementation evaluations can provide data about which coordination mechanisms work, with practitioners compiling information not only about how many actors are involved in a program but also about how well their efforts in this program have been coordinated.

Coordination should not be limited to PVE organizations, but should extend to all institutions that contribute to socialization, starting with families and schools:

“ We should invest more in schools and educational institutions. Many parents have problems to raise their children. Therefore, it is important to provide support at the early stage of parenthood, reaching the mother already when she is pregnant. All our cases with young radicals, they all came from a large family. It is difficult to raise a large family, since there is the risk of child neglect.” (EUR23)

Evaluation of such conditions, which have an indirect impact on the performance of PVE programs, is an integral part of implementation evaluations. It provides arguments for government investments that extend beyond PVE as strictly defined and address priorities for education, families, cities, and so on.

The practitioners identified support from government as an important part of the context for implementation evaluations. This support reflects media and political pressures, a constraint mentioned earlier. But this constraint becomes an opportunity when evaluations are used to justify increased attention from government. The practitioners expect the government to provide support in the form of coordination: "We had to learn to work together, to share necessary information without violating professional secrecy, and to trust one another." (EUR14) This coordination also involves civil society: "We also should reinforce dialogue between law enforcement and religious leaders." (ASI05) In tertiary prevention programs, coordination with correctional institutions is indispensable: from a security standpoint, it might

seem logical to select the most radicalized individuals to participate in such programs, but experience has shown that it is more effective to choose the individuals who will have the most influence on those around them:

“ The main challenge is to make sure that the right people are selected. We would think that they would choose the most radicalized to go on these courses, but in the course, what they are trying to achieve is a cascade effect. In other words, they want a good proportion of people attending these courses to go spread the message, to communicate their newfound skills. That’s where you get the best results.” (EUR02)

This kind of counter-intuitive result should appear in an implementation evaluation under the heading of the importance of having highly trusting relationships between security institutions and PVE practitioners.

Lastly, implementation evaluations help to communicate the complexity of PVE, and in particular the need for government to define and coordinate the responsibilities of the various actors. For this to happen, everyone’s roles and expectations must be spelled out and followed, as this practitioner suggested:

You need to set specific objectives for a process and a deadline for submitting the evaluation. I would add the co-construction aspect, which can be useful and productive. In other words, the objectives should be set by the government authorities and the donors, but at certain times there should be coordination to define the modalities. (EUR27)

The co-construction of clear, specific indicators was thus something that the practitioners asked for repeatedly. Evaluations should absolutely include it.

3.2 EVALUATING PRACTITIONERS’ NEEDS IN TERMS OF THEIR PROFESSIONAL CAREERS

PVE practitioners’ informal evaluations of their own practices serve yet another need: helping them to make PVE a credible professional career. Evaluations also let PVE organization managers demonstrate the need to provide stable funding for their programs and to provide their practitioners with the tools required to do their jobs.

Most of the practitioners in this study believed that they still needed better training in PVE, and in particular regarding “radicalization, jihadism, Middle East conflicts, migration and integration.” (EUR23, EUR06, EUR09, EUR21, EUR25) It might seem surprising that they reported training gaps on such crucial subjects; but the reality is that the demand for PVE programs, and hence for PVE training, had shot up dramatically, and the supply had not yet caught up. Therefore, this demand was being met by organizations that had not originally had any expertise in PVE. If the interests of PVE program participants, practitioners and society at large are to be met, it is essential to give practitioners the means of acquiring appropriate training and of evaluating their own competencies in PVE.

In general, PVE programs are delivered by not-for-profit organizations. The smaller the organization, the less funding it receives, and hence the more difficulty it has in attracting people with the required skills (EUR27, EUR21), in particular for a kind of work that does not have much status and is not very well paid: “It’s hard to find anyone who’s willing to do this work.” (MENA03) Retaining employees who have credibility (such as reformed extremists and people who are respected locally) is a tough challenge. (NA07) The difficulty of assessing practitioners’ qualitative achievements objectively and maintaining some psychological perspective on their work also weighs heavily on them:

We don’t chase numbers, “how many beneficiaries have we seen today”. We do try to be as careful as we can with our work, on quality. We are very aware of our strengths, our strong points, our abilities but also our limitations. One mistake that many people make is to believe they can fix everything with everyone, which is never the case. We have this messiah syndrome, “I’m going to save this and this person.” We are not saving anyone, anyway. (EUR25)

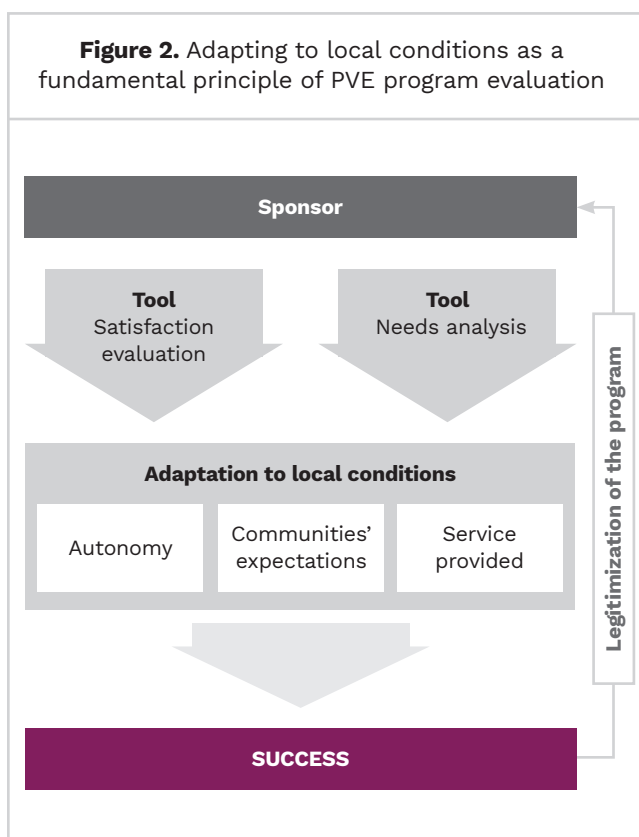
This cynical observation highlights how much importance should be given to selecting and evaluating the professional qualifications expected of PVE practitioners. In order to ensure the quality of PVE programs, it is essential that the scope of the work to be accomplished, its attractiveness and its constraints be clearly defined from the outset. Hence the criteria that will be used to determine whether the work has been performed satisfactorily must be defined realistically.

Adapting to local conditions as a basic principle

In their interviews, the practitioners identified certain action principles that they believe can ensure that PVE programs receive positive evaluations or, at the very least, that they are evaluated effectively. The key idea is to avoid approaches that are too centred on government. Instead, it is critical to get as close as possible to the local communities and local settings. According to the practitioners, for this approach to work, they must be given the necessary autonomy to carry it out. Adaptation to local conditions then becomes a key part of the process and an indicator to evaluate.

Figure 2 presents this concept schematically.

By adapting their PVE programs to local conditions, practitioners can legitimize them in the eyes of the communities that they are trying to serve and whose needs they are trying to meet. In this way, a social approach in the service of the community becomes the central focus of PVE practice, rather than a security approach in the service of the state. At a more symbolic level, evaluation based on professional experience enables practitioners to justify receiving a degree of autonomy from state institutions that program participants tend to perceive negatively.



4.1 THE RIGHT DISTANCE FROM THE STATE: A CRUCIAL FACTOR FOR PVE PROGRAM SUCCESS

The PVE practitioners believed that for their efforts to be effective, they must walk a fine line between carrying out PVE programs sponsored by the state and demonstrating that they are not agents of the state. Simply receiving government funding may make a program suspect, but in addition, for practitioners, the government can be an awkward partner. In the eyes of program participants, a practitioner who is too visibly associated with the state loses credibility and is not worth listening to. To reduce this risk, some PVE organizations prefer not to rely on one donor but instead work with several (such as the UN, UNICEF and USAID). These PVE organizations believe that relying on international humanitarian funding helps them to preserve their transparency and credibility, which might otherwise be a challenge (MENA05). In all cases, the source of practitioners' funding may undermine their credibility. But if an organization's funding arrangements mean a closer relationship with the state, evaluations can be a way of maintaining some distance. When suitably designed, they can help to objectively establish and justify a need for autonomy.

In theory, implementation evaluations are the kind that sponsors can most readily use to exercise control over organizations. But implementation evaluations can also provide data on implementation strategies, on the practitioners and partners involved in implementing programs, and on the contexts in which programs are implemented. When used in this way, implementation evaluations can demonstrate the need for organizations to retain some independence in selecting practitioners, selecting prevention practices, and taking local conditions into account. Implementation evaluations can also vividly demonstrate the need for PVE organizations as trusted intermediaries between program participants and the state: "One of the difficulties is that the [same] attorney who lead[s] the program [does] the prosecution. [...] Also, the government cannot deal with uncertainties." (NA04)

This is especially true in countries where repression is the preferred response to radicalization. As two practitioners described it, "We didn't want government funding because of all of the stigma" (NA04), knowing that, despite everything, the government's concerns "are still skewed heavily toward jihadism" and that "the government and the media are reluctant to [use the term] white extremism." (NA07)

Satisfaction evaluations can be similarly used to demonstrate the need for PVE organizations to have some autonomy. One European practitioner believed that after the November 2015 Paris attacks, "parents started complaining about being treated as if they were suspects or guilty, whereas before they were treated as victims" (EUR10), in a country where "there was nearly no talk

about prevention, and repression was the only response." (EUR11) In countries where conflict is imminent or has already broken out, this effect is even more pronounced: "[The] interventions that have the least success are those where you have the national government as a partner. It becomes difficult because the national government perceives civil society as an opponent." (SSA01) In such situations, the findings of both implementation and satisfaction evaluations will generally point to the need for practitioners to have a degree of autonomy.

Data from evaluations can legitimize the need for PVE organizations to have autonomy as well. Some of the direct indicators used in evaluations—such as propensity to talk to the police, number of participants in the program, reporting of human-rights violations, and abandonment of corporal punishment—imply the ability for organizations to criticize the state's actions in repressing radicalization and, more generally, human rights. The same is true of some of the indirect indicators, such as acceptance of violence, acceptance of different ideas, and ability to think critically. These success criteria are unattainable if the practitioners are seen as mouthpieces for the government:

“ They prefer to have a distance with the government to maintain their independence. They are happy to support efforts, and also critique efforts. They obviously appreciate what the government is trying to achieve in this area. However, governments aren't necessarily best placed [to engage in] counter-narrative work. It has to be fulfilled by civil societies and different bodies.” (EUR01)

Needs evaluations can also justify the need for autonomy. The state both employs and prevents violence. It has recourse to both repression and prevention. This ambiguity makes it hard for practitioners to succeed if they are not explicitly independent. By identifying the needs of communities on the ground, the practitioners can objectively establish their need to be seen as independent actors. The source of practitioners' funding is, in fact, a criterion that participants and communities use to evaluate them:

“ The main concern is about Western funding. We try to inform the people, to reassure them, and to be transparent. They see the forces of order as the enemy, and we avoid being lumped together with them. We stress that we are here to listen to people, not to call in the police.” (SSA05)

In a country in conflict, it can be hard to “fully understand who you are dealing with, [because of] the omnipresence of radicalized members everywhere in the community” (SSA03) or to regain communities' trust after the government has engaged in extrajudicial killings or forced

disappearances. (SSA01) In all of these cases, impact evaluations and needs evaluations show the importance for organizations to have a transparent, independent positioning.

By underscoring the public's mistrust of interventions that simply echo government rhetoric, the practitioners are tacitly arguing for increased autonomy for their profession. Also, the state, as a necessary source of funding for PVE organizations but a focus of suspicion for the target public, should never be the sole evaluator of PVE programs. A more appropriate alternative would be an independent evaluation that neither caters to nor is constrained by the state but instead focuses on the issues that concern communities and PVE practitioners.

4.2 EVALUATING COMMUNITIES' NEEDS AND SATISFACTION

In keeping with their vision of PVE as a service provided to communities, the practitioners underscored the need for their practices to be locally grounded to meet local needs. This means that the results achieved by PVE programs must be evaluated locally as well. It also means that participant needs and satisfaction must be evaluated (although the practitioners mentioned these two types of evaluations far less often).

The practitioners' comments indicate that current practices with regard to satisfaction evaluations need to be improved. One said, "We should design a satisfaction questionnaire with some open questions at the end. Right now, all we get are spontaneous, unstructured comments that don't really tell us very much." (EUR08) Other practitioners envisioned a more comprehensive impact evaluation that looks at the service provided to the community and the coordination of the actors:

“ We conduct a satisfaction survey, in which we question teachers and social workers and ask them to question the young people they work with and pass their responses along to us. That's really essential for their own evaluations, but none of that adds up to an impact evaluation. I'm not even talking about scientific accuracy, but rather about the medium and long-term impact of programs of this kind, which we know are only part of far broader educational efforts over a far longer time.” (EUR27)

Satisfaction evaluations provide the opportunity to defend a local approach to PVE against the top-down vision that is embraced not only by some sponsors, but also by some practitioners as well:

“ Too many practitioners think they know everything and want to immediately go to communities and tell them what to do. In my experience, what they know 'up there' is absolutely different from what communities feel and experience on a daily basis.” (SSA01)

This does not mean that programs cannot also benefit from the clarity and direction that a national focus can provide. As this police practitioner explains:

“ You have failed as the federal government when you have neglected the need for a local actor, local supportive, local directive approach. You want a national opinion, a national focus, I understand, but you gotta hit a structured flexibility.” (NA05)

With a properly designed needs evaluation or satisfaction evaluation, practitioners can document such flexibility.

Needs evaluations are, in fact, the other type that the practitioners have begun to attempt, though not in the most sophisticated possible way. One practitioner described an initiative in which communities were asked to choose among options, which is one way of assessing their needs:

“ We gave the communities a menu to choose what they really want. The CVE program comes from the government, so that cities had little choices what to do. [Our] framework was created to be diverse, inclusive and to allow communities to design their own path, programs that fit their needs.” (NA04)

But if communities are to be offered a diverse range of services, evaluations must capture the diversity of their needs. Communities must be able to evaluate each option on its own merits. Like satisfaction evaluations, needs evaluations are being done in an indirect, partial way, but remain a recurring concern: “When I do my evaluation this year, what I really plan to focus on is the young people in the field—I want to find out what is going on and what they need.” (NA09)

In some cases, practitioners did not originally perceive the need to focus on local conditions, but were forced to do so by circumstances. For example, several practitioners mentioned having attempted PVE initiatives based on values that they considered fundamental—such as secularism, gender equality, and children's rights—but that turned out to have little resonance in the community, because they were disconnected from its practical concerns. It would therefore be unfair to judge the results of a program against these criteria. The reason that practitioners evaluate participants' satisfaction with

the results of their programs is to demonstrate their detailed knowledge of the field and their concern for the communities' interests.

The treatment of religion in PVE programs typically requires an expert understanding of the field. Policymakers tend to perceive religion as the gateway to radicalization. But a focus on that possibility may prevent PVE programs from building trust with highly religious participants and communities. For example, talk about the Muslim religion from non-Muslim practitioners does not go down very well: "If you tackle the issue head-on and say things like 'That is not what the Muslim religion really says, here is the truth, you're being tricked, you're being lured into conspiracy theories, and so on,' then that is very obviously not going to work." (EUR14) And in majority-Muslim countries, the subject is no easier to address:

“ To be able to do radicalization-prevention work in Niger, you have to separate radicalization from religion, because in some contexts, talking about radicalization is seen as stigmatizing Islam, and usually people become radicalized because they feel stigmatized.” (SSA05)

In reality, religion is not an obstacle to PVE and, to some practitioners' surprise, can even facilitate it: "Interesting findings came out, especially in the case of madrassas in Pakistan. Islamic seminaries can also play a role in peace education that really works. Because for a long time, the predominant rhetoric has been to look at madrassas as breeding grounds of terrorism." (OCE02) Such constraints and results would have been hard to foresee before the practitioners had pointed them out. It is therefore crucial for evaluations to take practitioners' experience into account, both to avoid approaches that could be harmful and to take advantage of unexpected opportunities. It is to incorporate these essential lessons that the practitioners underscore the need to pay attention to participants' satisfaction and participants' needs.

One current theoretical approach to PVE involves finding universal mechanisms by which people adopt radical beliefs and behaviours. But the practitioners in this study stressed that, contrary to their first intuitions, this approach is often counterproductive. They develop metrics designed to gauge participants' needs and satisfaction instead, so as to avoid counterproductive approaches of this kind. Paradoxically, the criterion of localization appears to be one of the few "best practices" that all practitioners can apply. A sincere evaluation of PVE should therefore strive to strengthen these efforts to measure needs and satisfaction.

4.3 EVALUATING SERVICE PROVIDED TO THE COMMUNITY AS A MEANS OF DEFENDING A PRACTICE

To be effective, any evaluation of a PVE program must address the specifics of the community where it is delivered: "The interventions that have the highest impact are community-based interventions." (SSA01) Impact evaluations are therefore a natural way to demonstrate the benefits of bottom-up PVE practices that focus on providing a service to the target community. Impact evaluations must pay special attention to the community's perceptions of the intervention. An unfortunate choice of words or an unspoken attitude can hurt a program or derail it entirely: "I work in a neighbourhood where 80% of the people are from North Africa, so when you say the word 'radicalization', it really gets to them. It's going to frighten them, but it's also going to annoy them, because they're a little tired of always hearing the same speeches." (NA09) Another practitioner reported seeing the group of women in her program become withdrawn after hearing what they considered hostile comments about their veils. (EUR05)

These realities show that practitioners must not only display great cultural sensitivity, but also clearly understand their own roles. To be effective, they cannot show up acting like know-it-alls on the subject of extremism. They must ask themselves questions such as "How do you engage communities? How do you work with communities? How do you respect communities? We cannot offer solutions unless we have a common understanding of the problem." (NA04)

More specifically, instead of trying to convince program participants not to engage in extremism, practitioners should be trying to meet their practical needs and build trusting relationships with them. The ultimate goal is to reduce the potential attraction to violence. "Failure is ignoring what people really want. The factor of feeling I'm going to tell [them] what democracy is and should be and how young people [should] behave and live." (EUR03) The direct indicators used in evaluations, and especially in impact evaluations, reflect this commitment to providing a useful service. These indicators include the content of written materials handed out to program participants, the propensity to talk to the police, the number of participants in the program, reporting of human-rights violations, elimination of corporal punishment, school-enrolment rates, and literacy rates.

This judgment reflects the gap between government expectations and realities in the field. On the one hand, the reason that sponsors commission PVE programs is to prevent political violence. But on the other, practitioners stress the need to be a source of help for the participants with whom they work: "We work with law enforcement,

but we are negotiating that we won't be a security-oriented program." (NA04) This perspective is shared by practitioners who work for government, including this police officer: "You can't go after [young people] all the time, because if they haven't done anything wrong for a couple of weeks, but we, the police, come and check [on them] all the time, it becomes easy [for them] to see the police as the enemy." (EUR07) Instead of seeing young people as potential threats to which practitioners apply solutions developed in some laboratory, it is essential to successfully "give full ownership of the program to the young people themselves, because then they feel that they are being given acknowledgments, [that] they are being recognized." (ASI03) In the end, the outcomes of PVE programs are reconstructed from indirect indicators measured among the participants: acceptance of non-violence, well-being, acceptance of different ideas, ability to think critically, etc.

As opposed to a top-down vision of communities as dangerous, at-risk places from which disruptive elements must be purged, impact evaluations see them as groups of people whose needs the government must try to meet. By applying indicators that do more than simply try to anticipate primary violence or recidivism into violence, impact evaluations attempt to operationalize this criterion of trust as the best factor for success. In this way, they send the message that it is crucial for the field practitioners to be involved in designing and conducting the evaluations. As one practitioner said, "The government has to understand this issue." (SSA07) Moreover, this is something that policymakers have begun to demand—"They want to know what is happening here, so [the practitioners] document what they do, to show them" (EUR04), but in terms compatible with their roles as front-line practitioners. It would therefore be incongruous to design an evaluation that did not incorporate this dimension. For this reason, it is necessary to develop dedicated indicators for measuring it.



Suggestions for making evaluations an effective, legitimate tool for improving PVE practices

The preceding analysis of PVE professionals' comments about the evaluation of PVE practices has revealed a wide variety of situations. But out of this diversity, a picture does emerge of how evaluations can be made a relevant, useful tool for PVE practitioners.

For PVE practitioners, evaluations can present both constraints and opportunities. To develop a PVE evaluation culture that makes the most of the opportunities, two conditions must be met. First, all stakeholders must agree that formative evaluations will be just as helpful as summative evaluations. Second, evaluations must be designed so as to reflect the conditions under which PVE work is actually done, i.e. its qualitative and long-term dimensions.

The preceding sections highlight the kinds of improvements that can be achieved if these two conditions are met. In general, the types of evaluations that are already being performed fairly often (such as implementation evaluations) should be continued and enhanced, while those being performed less often (such as need evaluations) need to be developed. But these tasks should not fall exclusively on the organizations that deliver PVE programs. The work of evaluation should be shared among multiple stakeholders (program sponsors, program delivery organizations, trusted third parties, and possibly communities) in accordance with the resources available to each of them.

Regarding constraints, while PVE program evaluations do need to be redesigned to better reflect the specific features of PVE practice, PVE organizations should not have to provide all of the financial and human resources

for this task. The development of PVE by trial and error suggests that formative self-evaluation indicators should be defined by practitioners and evaluators working together. Because of the impact that evaluations can have on funding, they are vulnerable to manipulation in ways that compromise their integrity. It is therefore important to develop independent external evaluations. At the same time, funds must be dedicated specifically to the evaluation process. The short time frames of programs and the volatility of political and media attention require evaluations from multiple sources (PVE program organizations, sponsors, and trusted third parties). In addition, evaluations over the medium term would provide more robust results and compensate for shorter-term variations. Lastly, the development of qualitative indicators is absolutely essential, which will require involving researchers in external evaluations.

Regarding opportunities, in order to assess the services that PVE programs provide, impact evaluations must be improved. They should focus especially on the quality of the relationship built with the community, and practitioners should be involved in creating and refining the indicators used to assess this quality. The need for a coordinated definition of each stakeholder's action perimeters also calls for suitable indicators to be developed together with practitioners. This comes down to making implementation evaluations more robust, in particular through evaluation of the intervention context. Evaluation of communities' needs and satisfaction should be more systematic and should include the specific local features of every intervention terrain. The delicate balance that practitioners must maintain between the state and the communities argues in favour of evaluations

done by trusted third parties who are independent both of the practitioners and of the sponsors. These external evaluations should include indicators of the practitioners' autonomy. The strain that organizations' limited funding places on their ability to perform such vital tasks as attracting talented practitioners and providing them with appropriate training demonstrates the need to incorporate these aspects into evaluations as well.

These suggested improvements in PVE program evaluation can be classified under three headings: suggestions for policymakers, suggestions for practitioners, and suggestions for designing evaluations.

5.1 SUGGESTIONS FOR POLICYMAKERS

When evaluations are added onto existing programs after the fact, practitioners tend to perceive them as an attempt to impose control and an additional burden. But when evaluations are instead incorporated into programs when they are first designed, to enable their continuous improvement, practitioners tend to perceive them as relevant. For the findings to be as objective as possible, multiple situated evaluations should be conducted, reflecting the concerns of each of the actors, and they should then be compiled into the final evaluation.

To that end, we offer the following suggestions:

1. At the very outset, dedicate a portion of the program's funding to evaluations, without reducing the funding allotted to operations.
2. Conduct cross-evaluations: self-evaluations by practitioners to capture data from the field, internal evaluations within the organization to measure the program's effectiveness, and external evaluations to ensure that all organizations are treated equally.

5.2 SUGGESTIONS FOR PRACTITIONERS

3. Work together with evaluators and researchers to construct indicators that define the goals of all parties concerned.
4. Evaluate the organization's resources (training, talent retention, obstacles encountered).

5.3 SUGGESTIONS FOR DESIGNING EVALUATIONS

All of the practitioners questioned for this study stressed the qualitative dimension of their profession, and in particular the need to build trusting relationships with target communities and program participants. The only way that practitioners can build such trust is by approaching prevention in terms of services to be provided to communities. For this purpose, it is essential to identify community and participant needs before a program is designed and to evaluate community and participant satisfaction after it is delivered. The current scarcity of definitions and indicators for evaluating PVE programs makes this qualitative dimension hard to measure.

To address this issue, we offer the following suggestions:

5. Evaluate the quality of the relationship developed with the community.
6. Develop qualitative indicators regarding practices.
7. Develop indicators to measure the organization's independence from the sponsor.





Thematic Glossary

01

GENERAL

Sponsor: agency or institution that engages an organization to deliver a PVE program and provides it with the funding to do so

Community: group, within a society, that is targeted by a particular PVE program, located in a specified geographic area, and defined by shared traits (for example, the residents of a neighbourhood, the members of a religious group, the students at a school, or the members of an association)

Evaluation: the systematic assessment of the design, implementation or results of an initiative for the purposes of learning or decision-making (Canadian Evaluation Society, cited in Leblanc (ed.), 2017, p. 5)

Outputs: operational results that arise directly from a program's activities and take the form of goods and services delivered to its clients (for example, number of hours of services delivered, number of group activities conducted) (IRDPQ, 2015, p. 19)

Outcomes: results directly related to the program's objectives and to the changes that the program's activities are intended to achieve. Outcomes may be foreseen or unforeseen. (IRDPQ, 2015, pp. 19-20)

Practitioner: a PVE professional who works directly with target communities and program participants

Organization: an institution for which a PVE practitioner works

Participant: recipient of a PVE program delivered by an organization

02

GOALS

Formative evaluation: evaluation for the purpose of continuously improving services (IRDPQ, 2015, p. 17)

Summative evaluation: evaluation for the purpose of deciding whether a program should be continued or terminated (IRDPQ, 2015, p. 17)

03

ACTORS

External evaluation: the evaluation of a development intervention conducted by entities and/or individuals outside the donor and implementing organizations (OECD, 2002, p. 23).

Internal evaluation: evaluation of a development intervention conducted by a unit and/or individuals reporting to the management of the donor, partner, or implementing organization (OCDE, 2002, p. 26)

04

TYPES

Needs evaluation: an evaluation conducted to determine the needs of a population before an intervention is carried out (Meunier and Michaud, 2018, p. 26). A needs evaluation determines whether the program in place is sufficient to meet the actors' needs (IRDPQ, 2015, p. 18).

Satisfaction evaluation: an evaluation that considers the opinions and viewpoints of the actors concerned by the program [...], including how the actors perceive the quality of such things as the way the program's services are organized, the way it operates, the quality of its practices, the services that it provides, and the outcomes that it achieves. Satisfaction is measured by the difference between the quality that the client expected before the program was delivered and the quality perceived while it is being delivered or after it has been delivered. Satisfaction evaluations can contribute to needs, implementation and impact evaluations (IRDPQ, 2015, p. 20).

Impact evaluation (sometimes called outcomes or effectiveness evaluation): an evaluation conducted to determine whether a program that has already been in operation for some time is meeting its stated objectives. An impact evaluation assesses the changes actually produced by the program's activities and determines how closely these changes match the program's desired outcomes (IRDPQ, 2015, p. 19).

Implementation evaluation (sometimes called process evaluation): an evaluation of the internal workings of a program. An implementation evaluation examines the internal and external factors influencing the program's delivery and determines whether the program is being implemented in accordance with what was foreseen at the time that it was designed (Secrétariat du conseil du trésor, 2013, p. 9).

Economic evaluation: a comparative analysis of possible options on the basis of their costs and their consequences (Drummond, cited in Lance, 2004).

LEGEND

Characteristics of the 57 PVE professionals who participated in the interviews and focus group

World Region: the part of the world where the practitioner's organization operates (To protect the professionals' anonymity, they have been geographically identified only by world region and not by country.)

SSA: Sub-Saharan Africa

NA: North America

ASI: Asia

EUR: Europe

MENA: Middle East and North Africa

OCE: Oceania

Sector: Whether the practitioner's organization is in the public sector or the private sector.

Organization: The type of organization that the practitioner works for. NGO (non-governmental organization) is any private, not-for-profit entity (association, research centre, network, think tank, etc.) Sub-national government is any government entity at a lower level than the national government (province, region, département, municipality, etc.) Out of the 57 professionals questioned, about one-third were PVE practitioners working in direct contact with PVE program participants. Another one-third were PVE program coordinators or directors, one-quarter were heads of organizations involved in PVE, and the remainder were PVE researchers, trainers, and independent contractors.

Prevention Level: The level of prevention that the professional's program or organization provides (when the professionals were not specific about the level of prevention on which they worked, it was inferred from their statements, thus entailing a degree of uncertainty).

P: primary (general prevention aimed at broad segments of the population without distinction)

S: secondary (prevention specifically aimed at segments of the population identified as vulnerable to becoming radicalized)

T: tertiary (prevention of recidivism among individuals who have been convicted of crimes related to radicalization or terrorism).

World Region	Sector	Type of Organization	Prevention Level
SSA01	private	NGO	PS
SSA02	private	NGO	P
SSA03	private	NGO	P
SSA04	private	NGO	P
SSA05	private	NGO	PS
SSA06	private	NGO	P
SSA07	private	NGO	P
SSA09	public	National government	P
SSA10	public	National government	PS
NA01	private	NGO	PS
NA02	public	Educational institution	P
NA04	private	Medical institution	PS
NA05	public	Police	PST
NA06	public	Police	PST

NA07	private	NGO	T
NA08	public	National government	PST
NA09	public	Sub-national government	P
ASI01	private	NGO	PST
ASI03	private	NGO	PS
ASI04	private	NGO	PS
ASI06	private	NGO	PST
EUR01	private	NGO	PS
EUR02	private	NGO	PS
EUR03	private	NGO	PS
EUR04	private	NGO	PS
EUR05	private	NGO	PS
EUR06	public	Police	PST
EUR07	public	Police	PST
EUR08	private	NGO	PST
EUR09	private	NGO	PST
EUR10	private	NGO	PST
EUR11	public	Educational institution	P
EUR12	public	National government	PS
EUR14	public	Sub-national government	PS
EUR15	public	National government	PS
EUR16	public	National government	PST
EUR17	public	National government	PS
EUR18	private	NGO	PS
EUR19	private	NGO	P
EUR20	private	NGO	PST
EUR21	private	NGO	PS
EUR22	public	Police	PST
EUR23	public	National government	PST
EUR24	private	NGO	P
EUR25	private	NGO	PST
EUR26	public	Sub-national government	PST
EUR27	public	Sub-national government	P
EUR28	public	Sub-national government	PS
MENA01	public	National government	PST
MENA02	public	Sub-national government	P
MENA03	private	NGO	P
MENA04	private	NGO	P
MENA05	private	NGO	PS
MENA06	private	NGO	PST
OCE01	private	NGO	PS
OCE02	public	Educational institution	P
OCE02	public	Educational institution	P

List of references

- Campbell, D.** (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67-90.
- Gielen, A.-J.** (2019). Countering violent extremism : a realist review. *Terrorism and Political Violence*, 31(6), 1149-1167.
- Hassan, G., Brouillette-Alarie, S., Ousman, S., Kilinc, D., Varela, W., Lavoie, L., Venkatesh, V.** (2021a). *A Systematic Review on the Outcomes of Primary and Secondary Prevention Programs in the Field of Violent Radicalization*. Montréal: Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV). Récupéré sur <https://cpnprev.ca/wp-content/uploads/2021/03/CPN-PREV-2nd-Systematic-Review-2.pdf>
- Hassan, G., Brouillette-Alarie, S., Ousman, S., Kilinc, D., Varela, W., Lavoie, L., Venkatesh, V.** (2021b). *A Systematic Review on the Outcomes of Tertiary Prevention Programs in the Field of Violent Radicalization*. Montréal: Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV).
- Heydemann, S.** (2014). Countering violent extremism as a field of practice. *United States Institute of Peace Insights*(1).
- Hofman, J. et Sutherland, A. (dir).** (2018). *Evaluating interventions that prevent or counter violent extremism*. Santa Monica et Cambridge: RAND Europe. Consulté le 03 17, 2021, sur https://www.rand.org/content/dam/rand/pubs/research_reports/RR2000/RR2094/RAND_RR2094.pdf
- Horgan, J., & Braddock, K.** (2010). Rehabilitating the terrorists : challenges in assessing the effectiveness of de-radicalization programs. *Terrorism and Political Violence*, 22(2), 267-291.
- IRDPO.** (2015). *Cadre de référence en élaboration de programmation et en évaluation de programme*. Québec: Institut de réadaptation en déficience physique de Québec. Récupéré sur https://www.ciusss-capitalnationale.gouv.qc.ca/sites/default/files/cadreref_elaborationevprogramme2015.pdf
- Jackson, R.** (2012, 02 06). *Does counter-terrorism work ? Or, counter-terrorism as divination...* Consulté le 03 17, 2021, sur richardjacksonterrorismblog: <https://richardjacksonterrorismblog.wordpress.com/2012/02/06/does-counter-terrorism-work-or-counter-terrorism-as-divination/>
- Lance, J.-M. R.** (2004). Les approches d'évaluation économique utiles au champ de la santé publique. *Journées annuelles de santé publique 2004* (p. 11). Montréal: Institut national de santé publique du Québec. Récupéré sur <https://www.inspq.qc.ca/sites/default/files/jasp/rchives/2004/20041129/03evaluationeconomique/jasp2004-lance-approches.pdf>
- Leblanc (dir.), J.** (2017). Guide pour soutenir la mise en place d'une démarche évaluative. Avenirs d'enfants. Récupéré sur https://agirtot.org/media/488447/ae-guide_en_evaluation.pdf
- Lindekilde, L.** (2012). Value for money ? Problems of impact assessment of counter-radicalisation policies on end target groups : the case of Denmark. *European Journal on criminal policy and research*, 18(4), 385-402.
- Madriaza, P., Morin, D., Deli, C., Girard, M., Durocher, L., Grijalva-Lavallée, R., Preteroti, J.** (à paraître en 2021). *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).
- Madriaza, P., Ponsot, A.-S., & Marion, D.** (2017). *Prévention de la radicalisation menant à la violence*. Montréal: Centre international pour la prévention de la criminalité. Consulté le 03 17, 2021, sur https://cipc-icpc.org/wp-content/uploads/2019/08/FR_Rapport_Radicalisation_Final_Aout2017.pdf

Moghaddam, F. (2015). The staircase to terrorism : a psychological explanation ». Dans P. Neumann (dir.), *Radicalization* (pp. 161-169). Londres et New York: Routledge.

Neumann, P. (2013). The Trouble With Radicalisation. *International Affairs*, 29(4), 873-893.

OCDE. (2002). Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats. Paris: OECD Publication. Récupéré sur <http://www.oecd.org/dac/evaluation/2754804.pdf>

Paturet, J.-B. (2002). Le projet comme « fiction commune ». *Empan*, 45(1), 63-68.

Pistone, I., Eriksson, E., Beckman, U., Mattson, C., & 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.

Ponsot, A.-S., Autixier, C., & Madriaza, P. (2018). Factors facilitating the successful implementation of a prevention of violent radicalization intervention as identified by front-line practitioners. *Journal for deradicalization*(16), 1-33.

Sageman, M. (2009). Hofstad case study and the blob theory. Dans ARTIS, *Theoretical frames on pathways to violent radicalization* (p. 117). Theoretical Frames on Pathways to Violent Radicalization. Consulté le 03 17, 2021, sur https://sites.lsa.umich.edu/satran/wp-content/uploads/sites/330/2015/10/ARTIS_Theoretical_Frames_August_2009.pdf

Schmid, A. (2013). *Radicalisation, de-radicalisation, counter-radicalisation : a conceptual discussion and literature review*. La Haye: International centre for counter-terrorism. Consulté le 03 17, 2021, sur <http://www.icct.nl/app/uploads/download/file/ICCT-Schmid-Radicalisation-De-Radicalisation-Counter-Radicalisation-March-2013.pdf>

Secrétariat du conseil du trésor. (2013). *Glossaire des termes usuels en mesure de performance et en évaluation. Pour une gestion saine et performante en évaluation*. Québec: Gouvernement du Québec. Récupéré sur https://www.tresor.gouv.qc.ca/fileadmin/PDF/publications/glossaire_termes_usuels.pdf

Silber, M. D., & Bhatt, A. (2007). *Radicalization in the West: The Homegrown Threat*. New York: New York Police Department, NYPD Intelligence Division.

Skillicorn, D., Leuprecht, C., & Winn, C. (2012). Homegrown Islamist radicalization in Canada : process insights from an attitudinal survey. *Revue canadienne de science politique*, 45(4), 929-956.

Sommier, I. (2012). Engagement radical, désengagement et déradicalisation – Continuum et lignes de fracture. *Lien social et Politiques*(68), 15-35.

Wuilleumier, A., & Delpeuch, T. (2019). *Évaluer l'action de sécurité publique*. Paris: Institut national des hautes études de la sécurité et de la justice. Récupéré sur https://www.vie-publique.fr/sites/default/files/collection_number/portrait/photo/9782111459700.pdf.



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

IN PARTNERSHIP WITH



INTERNATIONAL
CENTRE
FOR THE
PREVENTION
OF CRIME

FUNDED BY



Gouvernement
du Canada

Centre canadien d'engagement
communautaire et de prévention
de la violence

Government
of Canada

Canada Centre for
Community Engagement and
Prevention of Violence



UNIVERSITÉ DE
SHERBROOKE

2500, boul. de l'Université
Sherbrooke (Québec)
J1K 2R1
CANADA



UNIVERSITÉ
Concordia
UNIVERSITY

1455, boul. De Maisonneuve O.,
Montréal (Québec)
H3G 1M8
CANADA

UQÀM

1430, rue Saint-Denis
C.P. 8888, succ. Centre-ville
Montréal (Québec)
H3C 3P8
CANADA