
Evaluating Terrorist and Extremist Reintegration Programming: A Systematic Literature Review

Megan K. McBride^{a1}, Marley Carroll^b, Jessa Mellea^b, Diána Hughes^c, Elena Savoia^d

^aFellow, Harvard T.H. Chan School of Public Health, ^bResearch Assistant, Harvard T.H. Chan School of Public Health, ^cSenior Graduate Research Assistant, Harvard T.H. Chan School of Public Health, ^dPrincipal Scientist, Harvard T.H. Chan School of Public Health

Abstract

This systematic literature review focuses on the evaluation of programs or interventions designed to deradicalize, disengage, rehabilitate, and/or reintegrate terrorists and/or extremists. Though a robust literature describing such interventions exists, it has long been recognized that more data are needed on the effectiveness of programs designed to facilitate reintegration. Thus, the objective of this review is to present an overview of academic and grey literature on the evaluation of programming designed to facilitate the deradicalization, disengagement, rehabilitation, and/or reintegration of terrorists and/or extremists. Our initial queries yielded 271 seemingly relevant peer-reviewed and grey literature articles, but after a more robust screening we determined that only 37 of those articles directly related to the evaluation of interventions. These articles are presented by evaluation design (e.g., quantitative, qualitative); we also review the evaluation mechanisms (e.g., survey, interviews, note review), study design (e.g., experimental, quasi-experimental), data types (i.e., quantitative or qualitative), and findings. Finally, we include ten articles discussing theory of program evaluation. Though the review is limited to available data (e.g., not including unpublished evaluations or government evaluations), we conclude by discussing the state of program evaluation relating to interventions designed to deradicalize, disengage, rehabilitate, and/or reintegrate terrorists and/or extremists, and offering several recommendations for how to improve evaluation methods and overcome barriers to evaluation.

Article History

Received May 24, 2022

Accepted August 16, 2022

Published Sept 30, 2022

Keywords: Systematic Literature Review, Terrorism, Extremism, Deradicalization, Reintegration, Tertiary Prevention, State of the Art

1. Introduction

In the two decades since 9/11, the U.S. approach to stopping terrorist attacks has evolved considerably. Early post-9/11 efforts were framed in the language, and pursued with the actions, of conflict and war. As part of this counterterrorism effort, suspected and confirmed terrorist actors were surveilled, detained, and even killed within a military framework that

¹ Corresponding Author Contact: Megan K. McBride, Email: mmcbride@hsph.harvard.edu, 90 Smith Street, Boston, MA 02115.

authorized responses along a spectrum, ranging from the nonviolent to the lethal. As the threat shifted toward the homeland, and the risk of domestic radicalization became a national preoccupation, a new set of tools became necessary because some earlier efforts (e.g., mosque surveillance programs) had come under serious and legitimate critique while others (e.g., drone attacks) were simply impermissible in a domestic context. The tools of addressing this domestic challenge—those of countering violent extremism (CVE)—were largely articulated within law enforcement frameworks. As such, the priority was that of collecting sufficient evidence to build cases against alleged perpetrators. Unfortunately, this approach, though more appropriate to a domestic environment, has proven to have limits (e.g., many parents are reluctant to report their potentially radicalizing children to law enforcement agencies). Thus, in recent years, a move has been made to situate interventions—both existing CVE programming and programming designed to prevent violent extremism (PVE)—within a public health approach. In particular, practitioners have begun to apply the public health matrix of primary, secondary, and tertiary prevention to violent extremism; this framework includes a progression from blocking the development of violent extremism in societies previously untouched by such violence (primary) to harm reduction practices aimed at radicalizing populations (secondary) to interventions seeking to deter recidivism or other recurrence of violent extremism (tertiary; Koehler, 2020b, p. 19; De Jong, 2010). The shift to this approach was likely precipitated by (a) recognition of the limits of previous efforts, (b) critiques rightly arguing that some CVE programming problematically profiled and targeted already vulnerable populations, and (c) the reality that white supremacist violence has overtaken jihadi violence as the most pressing domestic threat (effectively increasing domestic sympathy for those we now describe as “vulnerable” to the risk of radicalization). At the current moment, elements from all three approaches are active, and the United States engages in the work of countering and preventing terrorism and violent extremism within military, law enforcement, and public health frameworks.

Critically, the U.S. experience in the first decade after the 9/11 era was not a universal experience. Countries in the Middle East, for example, experienced the decade as one characterized by a domestic terrorist threat. As such, they pursued law enforcement and public health approaches far earlier than we did in the United States. As a result, for nearly two

decades now, local- and national-level governments have been working to develop and implement programs designed to prevent, counter, and undo the radicalization of would-be terrorists. And yet despite this increasingly robust effort to address the problem, and the universal recognition of the issue's importance, there is remarkably little known about which of these programs work.

Understanding which programs work is critically important. A number of scholars have produced literature reviews of the academic work on deradicalization, disengagement, rehabilitation, and reintegration of terrorists and extremists (Windisch et al., 2017; Grip et al., 2019; Doosje et al., 2016; Suarda, 2016; Feddes & Gallucci, 2015). However, it has long been recognized that more data are needed on the effectiveness of tertiary programs designed to facilitate reintegration. A 2011 report by Disley et al. noted that “while there are many descriptions of de-radicalisation interventions, [this rapid evidence assessment] identified no published robust evaluations of their effectiveness” (Disley et al., 2011). A few years later, a 2013 ICCT paper noted:

The conclusion is unavoidable that knowledge about de-radicalisation programmes is at this stage still very fragmented and uneven, which makes the comparative evaluation of various local initiatives an almost impossible task. External, in-depth evaluations of individual country programmes or detailed comparative evaluations of a rigorous sort do not yet seem to exist in the available literature (and it is very doubtful that they exist at all) (Schmid, 2013).

The paper went on to cite a 2010 article by John Horgan and Kurt Braddock that identified three major barriers to successful evaluation: the reality that few programs articulate clear criteria for success, the absence of data that can be validated independently, and the lack of a systematic effort to evaluate the programs alone or as a set (Horgan & Braddock, 2010).

Almost a decade later, the problem persists. A comprehensive 2021 systematic review of the literature on disengagement and deradicalization noted that evidence regarding program assessment and efficacy is “still of patchy quality,” and that “efforts are needed to improve data in this area” (Morrison et al., 2021, p. 43). The review itself offered an excellent

summary of the “major factors involved in these processes” with the goal of “assess[ing] the extent to which knowledge and understanding is progressing in this critical field” (Morrison et al., 2021, p. 43). As part of this work, however, it explicitly identified information on program evaluation as a knowledge gap. It pointed out that “the available evidence so far suggests that programmes, in general, do have positive benefits” and that “most participants report or show some positive impacts for most programmes” (Morrison et al., 2021, p. 43). It also said, though, that this type of broad data is not adequate: “The current state of knowledge is poor at identifying what elements of the different programmes have the most impact. As most interventions comprise multiple elements, this creates uncertainty over what works best” (Morrison et al., 2021, p. 43). In short, Morrison et al. point out that it is critical to understand not only which programs work, but which components of which programs work, in order to assess the effectiveness of these initiatives. This literature review focuses on that gap, speaking not to the variables involved in the process of deradicalization and disengagement (work already done by Morrison et al.), but to the effectiveness of programs seeking to promote deradicalization and disengagement. This work is especially important given that, as Daniel Koehler has warned, ill-designed and ineffective deradicalization or disengagement programs can *increase* the risk of terrorism by failing to detect high-risk cases (Koehler, 2017a).

Unfortunately, designing effective programming requires data that do not yet exist. And collecting those data—that is, evaluating the effectiveness of existing programming so as to determine what components do and do not work—requires developing, designing, and implementing *evaluations* of such programming. Recognizing the need for further scholarly literature on program evaluation and assessment, this review is focused on literature that deals directly with the *evaluation of tertiary prevention programs or interventions*, i.e., those designed to deradicalize, disengage, rehabilitate, and/or reintegrate terrorists and/or extremists. Thus, the primary objective of this literature review is to offer an overview of academic and grey literature on the *evaluation* of tertiary prevention programming, rather than focusing on program components and effectiveness. As such, it aspires to collect what is known about the evaluation methods themselves: how they have been implemented, what limitations they bring, and what paths forward are most viable and promising.

2. Methodology

This literature review was conducted following the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) statement whenever possible given the nature of the articles included (Page et al., 2021). A review protocol was not prepared or registered.

2.1 Definitions

The phrase “deradicalization, disengagement, rehabilitation, and/or reintegration of terrorists and/or extremists” is conceptually messy. The terms are entangled in even the most theoretically rigorous of academic papers, and the consistency with which they are defined deteriorates upon application in the real world. Program managers and policymakers, as one example, are unlikely to expend resources arguing over whether they have chosen the correct term when describing their program (and may, in fact, adjust language to meet the requirements of different funding mechanisms). In scoping this project, however, it was clear that our primary object of analysis was tertiary prevention programs that facilitate the reintegration of extremists (with deradicalization, disengagement, and rehabilitation assessed to be not necessarily essential to this ultimate objective). Deradicalization is an important goal, but is neither prerequisite nor equivalent to successful reintegration. Disengagement is perhaps more appropriate as it is a prerequisite to reintegration; however, it is also possible to be disengaged from violent extremism without being successfully reintegrated into society. Rehabilitation simply is not as commonly used within the literature. And so, for the remainder of this paper, we will use “reintegration” to refer to all four processes: deradicalization, disengagement, rehabilitation, and reintegration.

Similarly problematic, of course, is the language of terrorist and extremist. We understand that these are not overlapping categories and that no single term would be perfect. The word “terrorist” was particularly complicated given that it was not always clear what criteria the programs had used in admitting participants; not everyone involved in the evaluated programs met the legal definition of terrorist; and the programs occurred in different countries, meaning that the legal definition of terrorism varied. As a result, we use

the phrase “terrorists or extremists” to refer to the population being discussed throughout the remainder of this paper.

2.2 Information sources and search strategy

We conducted searches in the following sources: Web of Science, Social Science Premium, Policy File, PsycINFO, International Bibliography of the Social Sciences, Criminal Justice Abstracts with Full Text, and the Journal of American Medical Association using the following search string:

su(Extremis OR Terroris* OR Radical* OR Violen*) AND (De-radicali* OR Deradicali* OR Reintegrat* OR Re-integrat* OR Disengage* OR Recidivis* OR Rehabilit*) OR ti(Extremis* OR Terroris* OR Radical* OR Violen*) AND (De-radicali* OR Deradicali* OR Reintegrat* OR Re-integrat* OR Disengage* OR Recidivis* OR Rehabilit*).*

In addition to collecting peer-reviewed articles, our team conducted a comprehensive collection of the grey literature on this topic. We first made a list of approximately 50 websites as potentially relevant data sources (see Appendix 1). This list was compiled using Harvard Library’s Think Tank Search as a preliminary set of sources. We then excluded any sites with a focus on topics irrelevant to our research, and added sites identified as relevant through both previous research and consultation with experts. Though many excellent pieces have been written on tertiary programming itself, our object of analysis was *evaluations* of tertiary programming. Thus, when we searched the grey literature, we focused on formal reports and excluded blog entries and articles, as we deemed these unlikely to include the kind of literature we were hoping to find. With the list in place, we searched these sites, including articles published between January 2000 and July 2021, using the following general string in Google with some adaptations for the specific sites:

site:[INSERT] AND filetype:pdf AND (Extremis OR Terroris* OR Radical* OR Violen*) AND (De-radicali* OR Deradicali* OR Reintegrat* OR Re-integrat* OR Disengage* OR Recidivis* OR Rehabilit*)*

This methodology leverages Google’s search function to search at the URL level for results that meet the user’s criteria. As an example, the search string “site:brookings.com filetype:PDF deradicalization” will return every PDF on the Brookings website that includes the word deradicalization. This process is imperfect, as are all search strategies, but it allowed us to search the grey literature more effectively than a manual search. Aware of the difficulty in reaching high sensitivity with any search string in this topic area, we consulted with scholars working in this field to determine the comprehensiveness of the search strategy and ended up including ten additional articles based on their suggestions.

2.3 Eligibility (inclusion and exclusion) criteria

Our query was designed to return articles that had (i) been published after January 1, 2000, (ii) been published in a peer-reviewed journal, and (iii) been written in English. Because our search was completed in July 2021, this strategy returned approximately 21 years of peer-reviewed, English-language articles on the reintegration of terrorists or extremists. The articles that had no direct relevance to deradicalization, disengagement, rehabilitation, and reintegration programs for terrorists or extremists were removed from the dataset and marked as irrelevant. We also excluded any evaluations of programs primarily focused on primary or secondary prevention; that is, any programming where participants were not already fully radicalized. This included evaluations of programs aimed at a broad population, such as the residents of a specific region. As the team began a closer analysis, it became clear that a number of these results were books, book chapters, and book reviews, which were excluded from the analysis due to access issues arising from the COVID-19 pandemic.

2.3.1 Selection and data collection process

The results of the search queries were exported as RIS files from these databases and added to a Covidence project. Following this, we conducted a title and abstract screening.

Double screening was used to assess all articles, with two separate researchers independently coding each article. The reviewers worked independently to classify the selected articles. If consensus could not be reached, then a third member was consulted. In cases of disagreement, the team met to discuss both how to categorize challenging studies and how to describe the results.

2.3.2 Directly related and indirectly related articles

Once the initial text screening was complete, we reviewed in their entirety each of the articles in the dataset and marked them as either “directly related” or “indirectly related,” depending on their relationship to our core interest in the *evaluation of* reintegration programming. Our criteria for marking a piece of literature as directly related were as follows: 1) the article *mentions a specific (i.e., named) tertiary program or intervention* related to the deradicalization, disengagement, rehabilitation, or reintegration of terrorists or extremists and 2) the article *includes an evaluation of that program or intervention*. If the article described an intervention but did not include an evaluation of the intervention, we marked it as indirectly related. We made a few exceptions for literature that did not name a specific program but contained information especially relevant to our literature review.

In particular, we marked as directly related five articles that did not meet the criteria outlined above, but that explicitly discussed how to evaluate deradicalization programs effectively (Gielen, 2018; Mitchell, 2016; Williams & Kleinman, 2014; Koehler, 2017b; Koehler 2020; Horgan & Braddock, 2010; Marsden, 2015), as well as one article describing a coding mechanism to determine an individual’s level of disengagement from extremist ideologies (da Silva et al., 2019; Kutner, 2016; Bellasio et al., 2018). We chose to mark these articles as directly related because they had direct bearing on our primary object of analysis: literature on the evaluation of reintegration programming.

2.4 Data items and assessment of bias

After each article was categorized as directly or indirectly related, the full text of the article was used to extract and code (when available): (1) the program or intervention name; (2) the country where the program or intervention occurred; (3) whether the program or intervention

was voluntary or involuntary; (4) whether the program or intervention occurred in prison, after prison, or before the offense; (5) whether the program or intervention was administered by a public or private entity; (6) whether the program or intervention was overseen by multiple or single agencies; (7) what reintegration or intervention mechanism was used; and (8) who was administering the intervention/reintegration mechanism. When a program or intervention was evaluated, additional information was collected regarding the evaluation methodology, including data type (i.e., quantitative or qualitative), data collection mechanism, and data source. All data items were assessed for each study included in the review; the team met to discuss the risk of bias (i.e., selection bias, reproducibility of results) for each study; and reviewers worked independently and subsequently met to discuss the findings.

3. Results

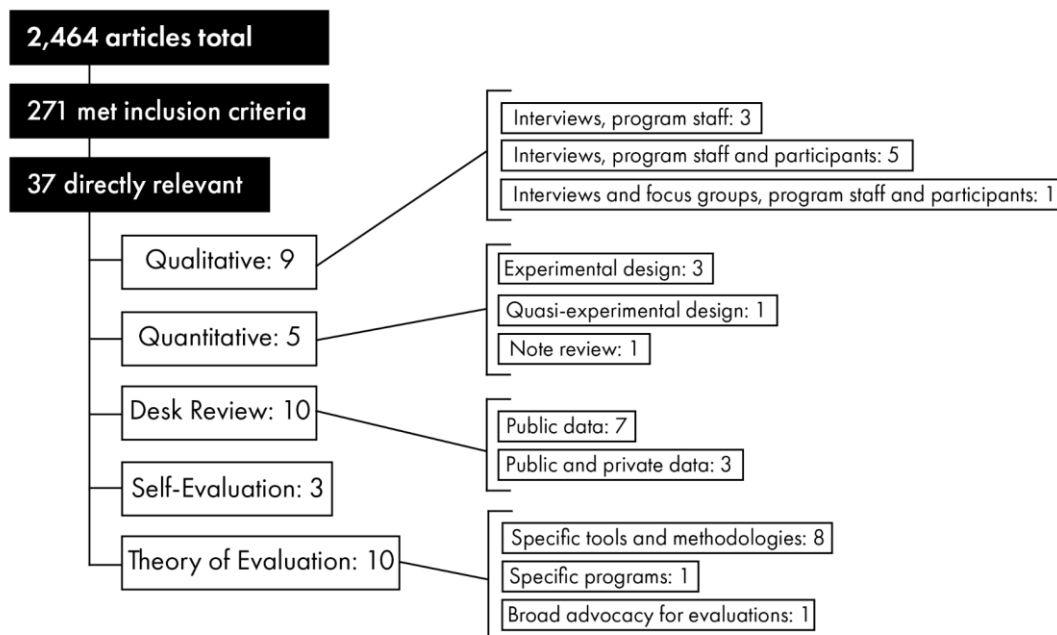
We screened 2,464 articles, from which we retrieved 271 articles (231 peer-reviewed articles, 30 articles derived from the grey literature, and ten articles found independently) related to the deradicalization, disengagement, rehabilitation, or reintegration of terrorists or extremists (Table 1).

Table 1: Literature categories

Peer-reviewed articles	231
Grey literature	30
Found independently	10
TOTAL	271

Of these articles, 37 articles were categorized as directly related and 234 were categorized as indirectly related to the topic of interest. These 37 articles seemed to significantly exceed what RAND had reported in a 2017 article when noting that they “were able to find only eight sources that assessed the impact of CVE programming” (Beaghley et al., 2017, p. ix). Moreover, since we excluded any evaluations of programs primarily focused on prevention rather than deradicalization or disengagement, we did not include all of the articles included in the RAND report (meaning that the total since RAND’s 2017 article might be even higher). As we read through the full text of the selected articles, however, it became clear that our findings were less promising than we had hoped. Below we provide a literature synthesis tabulating the results according to the type of data collection of data and method used to evaluate the programs: 1) qualitative, 2) quantitative, 3) desk review, and 4) self-assessments. In addition, we provide a synthesis of articles focused on 5) evaluation theories mentioned in articles about reintegration of terrorists or extremists. Results are described for each study within each category, as aggregation was not possible due to the heterogeneity of data collection methods.

Figure 1: Articles identified and reviewed



3.1 Qualitative methods (n=9)

Our review identified nine articles that used qualitative methodologies for evaluating reintegration programming. Eight relied primarily on first-hand data collected via interviews with program staff and/or program participants, and one used both interviews and focus groups (with program staff and program participants).

3.1.1 Interviews with program staff and/or program participants (n=8)

3.1.1.1 Interviews with program staff only (n=3)

Three of the eight evaluations that used interviews as their primary data collection methodology included interviews with only program staff (i.e., staff members, program administrators, government liaisons, etc.). Of these, two evaluations were based in the Netherlands (Schuurman & Bakker, 2015; van der Heide & Schuurman, 2018), where

researchers conducted semi-structured interviews with program staff and partner agencies in order to gather opinions about the program's implementation and effectiveness. The interviews were conducted over time in order to assess the development of the program. In the 2018 van der Heide and Schuurman article, the interview data were supplemented with a review of internal documents to collect information on program participants. The third evaluation that fell into this category focused on an initiative in Indonesia where data were also derived from interviews with program staff (Anindya, 2019). The number of respondents interviewed in these evaluations ranged from 7 to 29 (though the number of *interviews* was even higher, as the 29 participants participated in a total of 72 interviews over a 27-month period). That said, while each of the evaluations included interviews with program staff and individuals who were connected to the program (e.g., those at partner agencies), the Indonesian interviews were primarily conducted with P/CVE experts that were not affiliated with the program being evaluated, raising concerns about their understanding of the program and the validity of the results.

3.1.1.2 Interviews with program staff and program participants (n=5)

Five studies included interviews with both program participants and program staff. The number of interviewees ranged from 20 (Azam & Fatima, 2017) to 40 (Alonso & Díaz Bada, 2016). Of the four studies that broke down the distribution of interviewees' identities, three interviewed more staff than participants (Cherney, 2020; Cherney, 2018; Alonso & Díaz Bada, 2016) and one interviewed more participants than staff (Azam & Fatima, 2017); three specified that they interviewed both current and former program participants (Cherney, 2018; Christensen, 2015; Azam & Fatima, 2017); and one noted that participants interviewed had varying degrees of program involvement, with "some who had two to four engagements, to clients who had been engaged in the intervention for up to two years" (Cherney, 2018, p. 8). Finally, three studies specified that the interviews conducted were semi-structured in nature (Alonso & Díaz Bada, 2016; Christensen, 2015; Azam & Fatima, 2017), though none of the articles included interview guides or the questions asked during the evaluation.

3.1.1.3 Interviews and focus groups: program staff and program participants (n=1)

One evaluation used a mixed-methods approach that relied on both interviews and focus groups. The article evaluated two programs in Indonesia for combatants involved in the Poso conflict. Researchers interviewed over 100 participants and “key informants,” and conducted 16 focus group discussions (McRae, 2010, p. 426).

3.1.2 Results of qualitative evaluations

The vast majority of evaluations that used qualitative methods reported mixed or vague appraisals of the programs.

3.1.2.1 Interviews with program staff only (n=3)

The three evaluations that used interviews of program staff members noted a few positive elements but were overall somewhat negative about the programs. The evaluation of the Dutch NCTV-RN reintegration program noted that the initiative had some strengths, particularly the focus on disengagement as a foundation for reintegration and the format of the program that allowed participants to work closely with staff members; however, the program manifested challenges in its implementation (Schuurman & Bakker, 2015). Problems with management and cooperation between stakeholders, as well as a lack of clarity on how to define success, created friction in the creation and ongoing operation of the program. Ultimately, the researchers did not comment on the overall effectiveness of the program, as no participants had yet completed the full reintegration process.

A later evolution of this program was evaluated by van der Heide and Schuurman, leading to similarly mixed results (van der Heide & Schuurman, 2018). Again, researchers found that the theoretical underpinnings of the program structure with regard to disengagement and deradicalization were sound, but the program faltered in its implementation. The researchers suggested that these issues stemmed from a broad lack of empirical data. Overall, the evaluation indicated the program was a modest success, though with the caveat that data on participants’ rates of recidivism were only collected for a short period of time after the completion of the program.

Anindya offered a preliminary assessment of an intervention program in Indonesia, emphasizing the fact that the program was still in its early stages (Anindya, 2019). The article did not discuss rates of recidivism, having no base rate to compare to, and noted that while program staff have been successful in building trust with participants, issues with interagency cooperation have posed a challenge.

3.1.2.2 Interviews with program staff and program participants (n=5)

The evaluations conducted through interviews with both program participants and staff yielded similarly vague results. Cherney's evaluation of the prison-based Proactive Integrated Support Model (PRISM) intervention in Australia presented some benefits reported by program participants, but noted that the program itself is not tailored to extremist offenders (Cherney, 2018; Cherney, 2020). Cherney noted that the voluntary nature of the program meant that participants were more inclined to express their thoughts and beliefs, which facilitated the cognitive opening and self-reflection that program staff flagged as vital to the success of the intervention. That said, he also noted that the program's voluntary nature presented recruitment challenges, as potential participants expressed concern for the confidentiality and legal implications of what they shared in the program, as well as overall suspicion of the program and its staff's motives.

Azam and Fatima's evaluation of the Mishal deradicalization program in Pakistan deemed the program's religion module ineffective and concluded that "disengagement supersedes deradicalization at *Mishal*" (2017, p. 22). However, their ultimate assessment of the program was relatively balanced, stating, "The program has room for improvement and may deliver even better results with more financing and improved human resource [*sic*]," but that "even reducing the risk of reengagement for lower level militants can be classified as a success and worthy attempt" (Azam & Fatima, 2017, p. 23).

The Christensen evaluation of EXIT Sweden provided the vaguest program evaluation, describing the impact of the program but not providing an assessment of that impact (likely in part due to the difficulty in retaining data in accordance with Sweden's strict data protection requirements). The article stated that the process of former extremists becoming EXIT coaches reorients their mindset, as "[they] have gradually come to perceive

their story through EXIT and Fryshuset's frame of interpretation, transforming their extremist self-understanding and identity into that of a coach and social worker" (Christensen, 2015, p. 129).

The only clear judgment of a program was the evaluation conducted by Alonso and Díaz Bada (2016), which presents a starkly negative assessment of the impact of former ETA terrorists in deradicalizing other disengaging ETA terrorists. They noted that the deradicalization program sets a low threshold for "demonstrat[ing] a real distancing from violence" and that program participants' interactions with victims focus on forgiveness in a superficial way, thus warning that the program participants that then move into deradicalizing other former terrorists may not be given enough time to deradicalize before entering deradicalization work themselves (Alonso & Díaz Bada, 2016, p. 999). The authors ultimately characterized program elements as "imperfect... and some of them even counterproductive," as they do not adequately hold former terrorists accountable for their actions (Alonso & Díaz Bada, 2016, p. 1000).

3.1.2.3 Interviews and focus groups: program staff and program participants (n=1)

The evaluation that used both interviews and focus groups had mixed results. McRae described small gains, but not overall effectiveness. He found that some potential security threats that the program aimed to address (such as idleness and robustness of combatant networks) were not significant in the context where the program was implemented (McRae, 2010). The program had the clearest impact on only one of the planners' goals, police-combatant contact, which only improved somewhat as a result of the intervention program (McRae, 2010).

3.2 Quantitative methods (n=5)

Relative to the qualitative evaluations discussed above, we found comparatively few evaluations that utilized quantitative methods to assess reintegration programs (four evaluations described in five articles). Perhaps more noteworthy is that three of the four quantitative evaluations focused on Sri Lankan interventions that targeted the Liberation Tigers of Tamil Eelam (LTTE; one evaluation in Kruglanski et al., 2014a, Kruglanski et al.,

2014b, and Dugas & Kruglanski, 2014; two evaluations in Webber et al., 2018), a militant separatist group of which only a small portion (the Black Tigers) consistently used terrorist tactics (i.e., suicide bombings). Moreover, these three evaluations focused on post-2009 efforts—i.e., after the LTTE had been militarily defeated—raising the question of whether any observed success should be attributed to the programming or to the dissolution of the organization. The other evaluation that used quantitative data focused on a program that was not targeted at specific terrorist groups, but more broadly at individuals convicted of terrorist offenses and prisoners deemed to be at risk of radicalization (Cherney & Belton, 2021). In other words, while we identified four evaluations that relied on quantitative data or methods, it is not clear what percentage of these were evaluating programs designed to facilitate the reintegration of terrorists or extremists (as some of the programs included LTTE members who might better be classified as militant separatists, and one of the programs targeted at-risk individuals who had not yet radicalized to violence).

3.2.1 Study designs

3.2.1.1 Experimental design (n=3)

Perhaps unsurprisingly, only one of the evaluations identified in this literature review relied on experimental design (though it was discussed in three articles). This evaluation collected data from former LTTE members at the beginning of the intervention and nine months into it (Kruglanski et al., 2014a; Kruglanski et al., 2014b; Dugas & Kruglanski, 2014). It measured changes in survey participants' degree of organizational embeddedness, attitudes toward program staff, and willingness to engage in armed conflict in conjunction with LTTE's separatist aims. The evaluation included both an experimental group (n=1906) and a small control group (n=152), with the control group receiving very few of the intervention's components (Dugas & Kruglanski, 2014). The control group was "matched on many relevant characteristics to the rehabilitation sample," but was also slightly older and less educated than the experimental group, as well as being deemed by the Sri Lankan army to be higher risk due to their more active involvement with the LTTE (Dugas & Kruglanski, 2014, p. 435).

3.2.1.2 Quasi-experimental design (n=1)

Though we identified only one evaluation that relied on an experimental design, we did identify two evaluations that employed a quasi-experimental design (though the two were addressed in a single article). The first and more rigorous evaluation—the only evaluation assessing experimental and control groups over time—consisted of collecting survey data at the beginning, midpoint, and end of a one-year period (Webber et al., 2018). These data, importantly, were collected from two distinct groups of former LTTE members: (1) the roughly 80% of respondents (n=490) receiving the full extent of the intervention’s programming and (2) the approximately 20% of respondents (n=111) receiving only some of the intervention’s components (Webber et al., 2018). The second evaluation in this category included a control group but collected data just one time. This evaluation, however, relied on the comparison of survey responses from former LTTE members (n=179) who had completed a prison-based rehabilitation program with those from a control group (n=144) of analogous non-LTTE community members who had not participated in the program (Webber et al., 2018).

3.2.1.3 Note review (n=1)

The fourth quantitative evaluation—the sole quantitative study that did not rely on a survey for data collection—entailed coding case notes and risk assessment results from the PRISM intervention based in the Australian state of New South Wales (Cherney & Belton, 2021). Independent researchers (i.e., those not employed by the program) compiled a list of potential indicators of positive change mentioned in the disengagement literature and identified and coded such indicators within each set of case notes available for each participant. Thus, the researchers tracked variables such as level of psychological coping skills, commitment to extremist beliefs, and engagement in prosocial activities for the duration of the individuals’ participation in the program. Despite the study design seeking to standardize results so as to make them quantifiable, multiple degrees of subjectivity remained; the raw data analyzed were provided by the program—as case notes and risk assessments were completed by PRISM staff—and the coding of changes to the aforementioned variables

was subject to the researchers' discretion. This study also had a small n, with researchers coding both PRISM case notes and risk assessments for just 11 program participants, as well as the PRISM case notes of three additional participants. However, this small n resulted from the size of the intervention program itself; at the time of the article's writing, only 18 former extremists and individuals deemed to be at risk of radicalization had participated in PRISM programming.

3.2.2 Results of quantitative evaluations

The results of these evaluations were mostly positive. Both survey-based evaluations conducted by Webber et al. (2018; assessing an LTTE deradicalization program in Sri Lanka) were deemed a success. The longitudinal survey found that participants' extremist attitudes decreased over time and dropped more substantially for the participants who received the full extent of the intervention's programming. The one-time survey measured the long-term impact of the intervention and found that this reduction in extremist attitudes endured, and in fact left program participants with less extreme attitudes than the control group. This study also found that those who maintained connections with their previous social network had greater levels of extremist attitudes (and thus were deemed at risk for re-radicalization), but that most of the former participants did not maintain these ties. The preliminary results of the third evaluation of an LTTE deradicalization program were also encouraging, finding that participants' changes in attitude were "authentically effected [*sic*] by the rehabilitation programs that were implemented" (Kruglanski et al., 2014a, p. 195). The evaluation that used note review as its methodology provided less of an evaluation of the program's effectiveness, not claiming causation but noting that "the more often a person is engaged in PRISM over time the more likely they are to exhibit behaviors and attitudes that demonstrate disengagement" (Cherney & Belton, 2021, p. 637).

3.3 Desk Reviews (n=10)

A total of ten articles employed textual analysis as their primary research method. These desk reviews fell into two categories: those that relied exclusively on public

information (program websites, news articles, etc.) and those that also incorporated information not previously available to the public.

3.3.1 Public information (n=7)

We found seven desk reviews that relied overwhelmingly on publicly available information. These studies did not collect new qualitative or quantitative data, but instead used existing data, news articles, and other documents in their analysis. One review supplemented its public dataset with publicly available government-provided data on participants and practitioners involved in the program (Daugherty, 2019). Another used similar sources, including media, government, and institutional data (e.g., Amnesty International, Human Rights Watch) reports (Onapajo & Ozden, 2020). The majority of the evaluations did not explicitly specify a methodology (Fekete, 2015; Cano Paños, 2018; bin Hassan & Pereire, 2006; Ehiane, 2019; Williams & Lindsey, 2014). Though each of these evaluations were conducted by researchers external to the program or programs being evaluated, reliance on public data (some of which was provided by the governments or organizations that operated the programs) introduced the potential for bias.

3.3.2 Public and private data (n=3)

Three articles analyzed information not available to the public, in addition to publicly available data. In one study, researchers obtained documents from the government overseeing the program being evaluated (Andrews, 2020). These documents complemented public sources from the program's public website, scholarly work, and other public reports. In a second evaluation, the researchers used similar sources in addition to an interview with a practitioner in the program. While the author had planned on carrying out a qualitative study with further interviews, many of the possible respondents were imprisoned by the government, and the study was instead based primarily on text-based sources (San, 2020). The final study also supplemented its primary dataset of public information with interviews. Researchers interviewed 14 individuals involved in the evaluated program, as well as an unspecified number of academics, experts, and NGO workers (Parrin, 2016). Notably, each study supplemented public information with private data that were internally gathered and

provided by the institutions operating the program in question. Although the program assessments were conducted by external researchers, the use of internal data introduces the risk of bias into the findings.

3.3.3 Results of desk reviews

Similar to the qualitative evaluations, the desk reviews reported mixed evaluation results. Daugherty highlighted both successes and failures of Exit Sweden but noted that many of the participants had only recently started the program, making long-term evaluation impossible at that time. Both bin Hassan and Pereire's assessment of the Singaporean approach and William and Lindsey's evaluation of the Saudi program also pointed out strengths as well as areas for improvement. Onapajo and Ozden were more critical of the Nigerian program, remarking that the successes of the program have been self-reported by the government. The researchers' own evaluation noted several issues, mainly in the program's structure; Ehiane also critiqued the Nigerian program on structure and priorities. Parrin's evaluation of the Somalian program was similarly critical but included the caveat that the program was still in its early stages. One desk review was highly critical of Sweden and Germany's Exit programs, taking issue with the structure, ideological underpinnings, and perceptions of the program (Fekete, 2015). Others focused on more specific issue areas, such as gender (Andrews, 2020) or applicability to other models and settings (San, 2020). Numerous studies noted a lack of data, particularly in terms of long-term post-release monitoring and recidivism rates, which posed a barrier to effective evaluations (Daugherty, 2019; Ehiane 2019; Onapajo & Ozden, 2020; San, 2020).

3.4 Self-Evaluations (n=3)

While most of the assessments we found were written by external researchers, we did encounter three articles written by current or former practitioners within the programs being assessed. The first was a self-description of the Saudi program, which presented a solely internal perspective claiming the efficacy of the therapies used. The author, also the founder of the program, relied heavily on data gathered through experience working in the program, using an exploratory research design (Alyami, 2015). While the article included comments

from participants about their experiences, it does not appear that researchers conducted standardized, formal interviews. Success and recidivism rates are self-reported over an unspecified time period, and the definition of successful reintegration appears to be that an individual is not incarcerated again. The second study relied entirely on disengagement statistics and data regarding family participation in the intervention, both of which were provided to the researchers (one of whom was a program staff member) by the police department running the program (Bastug & Evlek, 2016). Thus, neither the data nor the data analysis came from an impartial third party. The third study was conducted by the program's designers and operators but provided some of the most granular data of any of the evaluations we found. The evaluation, which considered a pilot program of interventions through one-to-one messaging on social media, included rates of engagement by participants and whether participants responded to different types of messages. Researchers also reviewed the content of the messages sent by participants and the experiences of the intervention providers (former extremists who took on the position voluntarily) (Frenett & Dow, 2015).

3.4.1 Results of self-evaluations

The results of these evaluations were overwhelmingly positive. The Saudi evaluation cited an 86.17% success rate and a 13.83% recidivism rate (both provided by Interior Ministry, which ran the program), and though it presents viewpoints that are both supportive and critical of the “soft approach” taken in the program being evaluated, the critiques are ultimately diluted or rejected (Alyami, 2015, p. 1). The Turkish evaluation offered slightly more nuance, noting that results depended to a degree on the type of extremist group. It also described some of the program's limitations and identified some changes that could have been made to improve the program. Overall, though, the evaluation found the program to be effective according to the metric that “more than sixty percent of the militants with whom program officers met decided to leave the group” (Bastug & Evlek, 2016, p. 41). The final evaluation acknowledged that it was too early to assess long-term behavioral shifts but noted that the pilot program illustrated that one-to-one messaging was a promising method for deradicalization (Frenett & Dow, 2015).

3.5 Theory of Evaluation (n=10)

In addition to the aforementioned program evaluations, we encountered ten articles that examined reintegration program evaluation from a theoretical perspective. Eight of those articles (addressing five unique evaluation frameworks) offered a distinct evaluation tool or methodology that could be applied to reintegration program evaluation, one designed an evaluation for a specific reintegration program, and one presented a series of recommendations for P/CVE program evaluation writ large.

3.5.1 Specific tools & methodologies (n=8)

While eight articles recommended models for use in evaluating reintegration programming, two advocated specifically for Multi Attribute Utility Technology (MAUT), an evaluation method considered to be useful in the assessment of social programs for which experimental studies are difficult or otherwise untenable (Horgan & Braddock, 2010; Marsden, 2015). Horgan and Braddock also argued that MAUT can potentially “be a framework for guiding the development of future such initiatives that draw lessons from existing programs” (2010, p. 269). The method entails identifying and weighing stakeholders’ objectives and goals, followed by assessing the degree to which the program meets such goals. Specifically, the process consists of four steps: 1) identifying what will be evaluated and what the goal of the evaluation will be, 2) identifying stakeholders, 3) eliciting from stakeholders the specific program attributes to be evaluated and categorizing these attributes, and 4) determining the relative importance of each attribute and category. For the categorization component of step three, researchers organize the attributes into a hierarchical structure called a “value tree,” coming up with overarching categories that attributes fall under. Thus, in step four, weights are added to each attribute and category of the value tree (according to relative importance). Horgan and Braddock argued that “there must be one version of the overall value tree that can be applied to all risk reduction initiatives so as to evaluate them on equivalent terms,” noting that if a given attribute is irrelevant to a particular reintegration program, it can be given a weight of 0 (2010, p. 283). Thus, MAUT represents a promising method for standardizing reintegration program evaluation results.

Another approach-focused article provided a looser framework for program evaluation; the methodology closely mirrors MAUT but is not given a name (Mitchell, 2016). The study identified the individual components that are present within various reintegration programs and cross-referenced those data with a list of programs deemed to be most successful (no measure of success is presented in the article) in order to determine which individual components should be included in reintegration programs.

The second distinct methodology, the Innovative Moments Coding System (IMCS), tracks turning points in an individual's self-conception, thus enabling what the authors highlight as a necessary shift toward evaluating change as a result of interventions (da Silva et al., 2019). The methodology tracks "Innovative Moments" (IMs), which are instances when an intentional intervention enables the program participant to "develop a more flexible and alternative framework of meaning construction," and follows the different levels of IMs (with higher levels indicating deeper self-reflection) over the course of the intervention (da Silva et al., 2019, p. 206). The IMCS consists of two coders independently 1) reviewing video, audio, or a transcript of a therapy session or other intervention; 2) pinpointing the individual's problematic self-narrative; 3) identifying IMs; then 4) categorizing IMs. This process evaluates both the quantity and quality of the IMs; this knowledge establishes a fuller picture of the transformation of the individual's self-narrative and how it occurred, thus facilitating analysis of the degree to which the intervention was responsible for this change.

The third type of evaluation, the realistic evaluation method, examines the impact that context has on the program's outcomes (Gielen, 2018). It entails four steps: 1) using theoretical literature to establish a conceptual framework for developing hypotheses regarding the impact of context on program outcomes; 2) multimedia data collection; 3) analysis of the context in which the program takes place, mechanisms at work, and outcomes; and 4) refinement of the conceptual framework to establish a theoretical model of "what works, for whom, how and in what circumstances" (Gielen, 2018, p. 456).

A fourth type of evaluation was offered by Williams and Kleinman (2014), who presented a utilization-focused evaluation method, which operates under the premise that evaluations should be designed with an eye toward stakeholders' needs and the utility of the evaluation's results. As such, their article focused primarily on the evaluation design process

rather than prescribing a particular evaluation method. A utilization-focused approach must first identify stakeholders—and especially key stakeholders—and consult them regarding evaluation questions and other decisions in the evaluation process (Williams & Kleinman, 2014, p. 120). The selection of evaluation personnel must also be carefully considered in order to facilitate cultural awareness in the evaluation design and process (Williams & Kleinman, 2014, p. 121). These stakeholder and personnel identification processes lead into the development of the evaluation design, which entails a consensus-building exercise among stakeholders regarding the statement of the problem and the evaluation aims (Williams & Kleinman, 2014, pp. 121-122). Once consensus is reached, the evaluators must gather information about the program from a variety of perspectives in order to develop a nuanced understanding of the logic of the program’s design, incorporating “both the mechanics and the assumptions of how and why the program ostensibly works” (Williams & Kleinman, 2014, pp. 123-124). Only at this point can the evaluation methods be solidified, taking into account the evaluation’s goals and the program logic. Finally, program evaluators implement the evaluation, analyze the data collected, and communicate their findings and recommendations to stakeholders (Williams & Kleinman, 2014, pp. 130-132). In the appendices of their article, Williams and Kleinman also provided a thorough impact analysis checklist for evaluators to use when designing, implementing, and presenting the findings of a utilization-focused evaluation, as well as a self-assessment for evaluation designers and program evaluators (2014, pp. 139-161).

Finally, Koehler advocated for a different category of evaluation altogether: a structural integrity evaluation (Koehler, 2017b; Koehler, 2020a). This type of evaluation, which accounts for the relative paucity of available data on reintegration programs and the difficulty in analyzing data and attributing causality, relies on the idea that a program’s “theory of change” is highly indicative of its potential for success and focuses on assessing easily measurable components of a program’s capacity and content in order to evaluate the program’s structural integrity (Koehler, 2020a, pp. 35-36). Koehler’s structural integrity framework calls for the collection and analysis of data relating to six fields—running and developing a program, personnel and organization, participant classification, care and advisory services, quality assurance, and transparency—and includes both questions aimed at

ascertaining the status of several listed indicators of structural integrity within each field and a “De-radicalisation Programme Integrity Evaluation Checklist” with checklists assessing the program with regard to each field (Koehler, 2017b; Koehler, 2020a). Finally, Koehler provided guidelines for the drafting of a structural integrity evaluation report, which can be tailored to the interests of particular stakeholders (Koehler, 2020a, p. 85).

3.5.2 Specific programs (n=1)

One article presented an unimplemented study design for the Saudi Arabian Risk Reduction Initiative (Kutner, 2016). This study design recommended examining an individual component of the reintegration program: religious re-education. The hypothetical study implemented an experimental design, wherein some program participants would receive religious re-education while a control group would not. The study design offered a variety of evaluation mechanisms, including covert observation of participants in the experimental group by program staff and a questionnaire loosely based on the VERA-2R.

3.5.3 Broad advocacy for evaluations (n=1)

Finally, one article advocated broadly for more systematic program evaluation (Bellasio et al., 2018). It suggests that program evaluators focus on programs’ economic efficiency, conducting cost-benefit analyses to determine the highest-value programs. It also recommended a set of evaluation conditions, “including the use of empirical data, multiple methodologies and adequate stakeholder engagement methods” (Bellasio et al., 2018, p. 79), while encouraging experimental study designs whenever possible. The report described a “promising evaluation” to have met the minimum criteria of “capturing multiple perspectives (to limit bias), using grids, rubrics, scores or a set of indicators (to enhance transparency and objectivity), and providing recommendations (to increase the probability that evaluation results are used)” (Bellasio et al., 2018, p. 77). As such, the report stressed the importance of program evaluators clearly presenting their evaluation methodologies to facilitate other researchers’ assessment of the evaluation.

The report acknowledged that reintegration program evaluation is in its nascent stages as a field, as is reintegration programming itself. Thus, the authors suggested that, in this

initial phase, researchers should focus primarily on accumulating knowledge rather than definitively assessing program efficacy. They posit that real-time evaluations could facilitate knowledge accumulation while also speeding up program improvement by providing program providers with immediate feedback.

4. Discussion

This literature review confirms that despite an increase in literature regarding the evaluation of programs designed to reintegrate terrorists or extremists, we know little about which programs (or program components) work. In an ideal situation, we would look at the characteristics of the evaluated reintegration programs, the methods used in the evaluations, and the outcomes of the evaluations. With these data, we would then attempt to identify the program features most frequently correlated with success (e.g., inclusion of a social worker, intervention in prison) and the evaluation features most frequently correlated with empirically valid results. Unfortunately, due to the heterogeneity of approaches, type of interventions, context and data collection methods, it is impossible to aggregate results to derive what does and does not work.

4.1 Barriers to evaluation

Though the number of evaluations has increased over the past ten years, there are still serious challenges to conducting evaluations. First, terrorist violence is relatively uncommon compared to other forms of crime, so a relatively small number of people are viable candidates for reintegration programming. With so few data points, it is difficult to measure the effectiveness of a given program with any certainty. A second barrier to evaluation is poor access to program data. For academics and groups outside of the program or government, it is often difficult to obtain outcome data or observe the interventions. Without this information, it is challenging—if not impossible—for researchers to design comprehensive and accurate research studies on the effectiveness of reintegration programming. Third, there are significant logistical barriers that hinder the evaluation of these programs. For example, funding cycles for CVE programs and evaluations are not always aligned; funding for

evaluation is often inadequate to support robust, empirical studies; and most evaluations cannot be used to support long-term conclusions about deradicalization within a group or community. And finally, potential evaluators must contend with the diverse and often contradictory goals of the programs. There does not appear to be any consensus within the field over what a successful program looks like, and each program has its own metrics for evaluating success. Some interventions focus solely on preventing recidivism (Basit, 2015), while others encourage ideological disengagement and concrete changes in attitudes or beliefs (Cherney, 2020). Even within single countries, reintegration programs might have different goals or measures of success. These inconsistencies make it difficult for scholars to compare the success of one program to that of another, further complicating the development of a robust literature on the reintegration process.

Given this, it should come as no surprise that evaluations of programs designed to reintegrate terrorists or extremists are few and far between. That said, it should be noted that our review had limitations: we did not have access to evaluation reports that may have been generated for government agencies; we excluded non-English literature; and it is possible (even likely) that there are numerous completed evaluations that yielded negative results and were consequently not published. Despite the limitations, however, our literature review highlighted a number of issues that lead naturally to some straightforward recommendations for those working on this issue.

4.2 Recommendations for practice

Before offering recommendations, we would like to address the elephant in the room. Obviously, the field would benefit from consensus on the use of metrics for measuring success. A common definition of what success looks like would make possible analysis across studies—facilitating not only cross-cultural analysis, but also meta-analyses of the literature writ large—but in the short term, we think this is an unreasonable expectation. It is not only the case that tertiary prevention programming pursues a variety of distinct objectives (e.g., deradicalization, disengagement, rehabilitation, reintegration), but also that the words used to describe these objectives often have contested definitions. Moreover, reintegration programs themselves are often influenced or shaped by political actors and community activists who

intend well but have little understanding of the dynamics at play. Thus, the objective of a single program might actually be CVE- or PVE-adjacent at best, a program “that may (or may not) be labeled as CVE programming, and that is intended to produce outcomes associated with the prevention of violent extremism, but that is not necessarily supported by peer-reviewed literature or rigorous evaluation” (Savoia et al., 2020, p. 18). Finally, we acknowledge that achieving consensus on measures of success may never be a viable option given the methodological diversity of programming in this space. Additionally, it would also be better if there were more external, independent, and empirical evaluations of reintegration programming. The reality, though, is that the impediments to evaluation are often outside the control of the researchers eager to do such analysis.

These recommendations are consequently more modest insofar as we hope to highlight three key issues that would be easy to implement, and that would improve the quality of data in this space (even in the absence of a standardized definition of success or an increase in funded evaluations).

(1) All articles describing program evaluation should include a clear definition of success.

There is no question that these definitions will vary, and that some will be articulated by the programs while others will be articulated by the evaluators. Most articles we read failed to include any form of definition of success, though some articles did reference *indicators* of success (Kruglanski et al., 2014b; Anindya, 2019; Cherney, 2018). Only a few articles actually included a version of a definition of success, with one article noting that success entails effective “[demotivation] of individuals from participating in the radicalization process” and another two defining success as avoidance of recidivism but not defining recidivism (Bastug & Evlek, 2016; Alyami, 2014; Schuurman & Bakker, 2016). Standardization, as noted above, would be ideal; however, the inclusion of a definition of success (even if the definition is unique to the program) would greatly facilitate the comparison of programs and evaluations.

(2) All articles describing program evaluation should include a baseline set of descriptive data. Our initial intention in executing this literature review was to create a robust map of the programs that had been evaluated. This map was meant to capture a wide range

of variables including population serviced (e.g., age, gender, education level), geographic area serviced (e.g., urban, rural), characteristics of administering organizations (e.g., government, private), and types of services provided (e.g., mental health, education, spiritual health). When available, we also intended to include data on study design, sample size, length of the observational period, and outcome measures being used to evaluate the programs. One of the most frustrating realities of the literature that we reviewed, though, was the inconsistent nature of program descriptions and the evaluation data being reported. As one example, many of the articles describe the programming provided as “individualized programming” without mentioning the interventions (Schuurman & Bakker, 2015; Campelo et al., 2018; Cano Paños, 2018; Cherney, 2020). It would be helpful if these programming packages were disaggregated and described, even if the program can only be evaluated as a whole. Given this, though we did attempt to code the articles as described above, we were ultimately unable to create this map because the data simply were not available. Thus, we think the field would benefit if articles discussing actual reintegration programming (to include articles that discuss program evaluations) always included the following information: the name of the program, the specific type of programming or intervention(s) provided (e.g., financial aid, psychological counseling), whether the program is voluntary or involuntary, which private or public organization(s) administer the program, the number of individuals who have participated in the program, and the number of participants in the program who were successfully reintegrated as a result of the program.

- (3) All articles describing program evaluation should clearly outline their methodologies when presenting their evaluation results. Several of the evaluations that we encountered omitted key details regarding the data collection methods being used, including studies that did not include a breakdown of interviewees’ identities (e.g., program participant or staff, current versus former program participant) or that failed to describe the methodology at all (Christensen, 2015; Alonso & Díaz Bada, 2016; McRae, 2010). Though these details may not be essential to the findings of a single evaluation, they are critical to facilitating the maturity of the broader discussion about

program effectiveness. Such details make it possible for other researchers to recreate—and thus validate—the evaluations conducted, and also enable researchers to learn from, adapt, and expand upon one another’s evaluation methodologies.

The evaluation of reintegration programming is still a relatively underdeveloped area of research, and while we agree with calls for increased investment, greater use of empirical data, and more external evaluations, we have focused here on more modest recommendations that can be implemented immediately. In other words, we believe it is possible to improve the quality of the evaluation work conducted in this field and our understanding of program effectiveness by implementing relatively small changes that will make it possible to put distinct evaluations of distinct programs in conversation with one another. This change would not produce magical or transformative results, but it could meaningfully move the needle in both the short- and mid-term on our understanding of how to facilitate the reintegration of terrorists or extremists.

Acknowledgements

This project was made possible by funding and support from the National Institute of Justice (Applying a Developmental Evaluation Approach to Address Community Safety and Health Challenges of Reintegration Programs in the USA - Award Number 2019-ZA-CX-001). The content of this manuscript, as well as the views and discussions expressed, are solely those of the authors and do not necessarily represent the official views of any of the above institutions, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government.

References

- Alonso, R., & Díaz Bada, J. (2016). What Role Have Former ETA Terrorists Played in Counterterrorism and Counterradicalization Initiatives in Spain?. *Studies in Conflict & Terrorism*, 39(11), 982-1006.
- Alyami, A. M. S. (2015). The role of art therapy in counterterrorism: The Saudi experience. *The Arts in Psychotherapy*, 44, 1-10. <http://dx.doi.org/10.1016/j.aip.2015.04.003>.
- Altier, M. B., Leonard Boyle, E., Shortland, N. D., & Horgan, J. G. (2017). Why they leave: An analysis of terrorist disengagement events from eighty-seven autobiographical accounts. *Security Studies*, 26(2), 305-332.
- Altier, M. B., Thoroughgood, C. N., & Horgan, J. G. (2014). Turning away from terrorism: Lessons from psychology, sociology, and criminology. *Journal of peace research*, 51(5), 647-661.
- Andersen, H. S. (2010). *The challenge of extremism: Examples of deradicalisation and disengagement programmes in the EU*. Ministry of Refugee, Immigration and, and Integration Affairs.
- Andrews, S. (2020). Prevent tragedies: A case study in female-targeted strategic communications in the United Kingdom's Prevent counter-terrorism policy. *Journal for Deradicalization*, (24), 1-39.
- Anindya, C. R. (2019). The deradicalisation programme for Indonesian deportees: A vacuum in coordination. *Journal for Deradicalization*, (18), 217-243.
- Ashour, O. (2007). Lions tamed? an inquiry into the causes of de-radicalization of armed Islamist movements: The case of the Egyptian Islamic Group. *The Middle East Journal*, 61(4), 596-625. <https://doi.org/10.3751/61.4.12>.
- Awan, I. (2013). Muslim prisoners, radicalization and rehabilitation in British prisons. *Journal of Muslim Minority Affairs*, 33(3), 371-384. <https://doi.org/10.1080/13602004.2013.853979>.
- Azam, Z., & Fatima, S. B. (2017). Mishal: a case study of a deradicalization and emancipation program in Swat Valley, Pakistan. *Journal for Deradicalization*, (11), 1-29.
- Barkindo, A. & Bryans, S. (2016). De-Radicalising Prisoners in Nigeria: developing a basic prison based de-radicalisation programme. *Journal for Deradicalization*, (7).
- Basit, A. (2015). Countering Violent Extremism: Evaluating Pakistan's Counter-Radicalization and De-radicalization Initiatives. *IPRI Journal*, 2, 44-68.

Bastug, M. F., & Evlek, U. K. (2016). Individual disengagement and deradicalization pilot program in Turkey: methods and outcomes. *Journal for Deradicalization*, (8), 25-45.

Beaghley, S., Helmus, T. C., Matthews, M., Ramchand, R., Stebbins, D., Kadlec, A., & Brown, M. A. (2017). *Development and pilot test of the RAND Program Evaluation Toolkit for Violent Extremism*. RAND Corporation.

https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1799/RAND_RR1799.pdf.

Bellasio, J., Hofman, J., Ward, A., Nederveen, F., Knack, A., Meranto, A. S., & Hoorens, S. (2018). *Counterterrorism evaluation: Taking stock and looking ahead*. RAND Corporation.

Bertram, L. (2015). How Could a Terrorist be De-Radicalised?. *Journal for Deradicalization*, (5), 120-149.

bin Hassan, M. H. & Pereire, K. G. (2006). An ideological response to combating terrorism — the Singapore perspective. *Small Wars and Insurgencies*, 17(4), 458-477. DOI: 10.1080/09592310601060815.

Braun, K. (2018). ‘Home, Sweet Home’: Managing Returning Foreign Terrorist Fighters in Germany, The United Kingdom and Australia. *International community law review*, 20(3-4), 311-346.

Bubolz, B. F., & Simi, P. (2015). Leaving the world of hate: Life-course transitions and self-change. *American Behavioral Scientist*, 59(12), 1588-1608.

Campelo, N., Bouzar, L., Oppetit, A., Pellerin, H., Hefez, S., Bronsard, G., Cohen, D., & Bouzar, D. (2018). Joining the Islamic State from France between 2014 and 2016: An observational follow-up study. *Palgrave Communications* 4(137), 1-10.
<https://doi.org/10.1057/s41599-018-0191-8>.

Cano Paños, M. A. (2018). Fighting the Jihadist threat beyond criminal law: Analysis of programs to prevent radicalization and de-radicalization at the European Level. *Revista de Estudios en Seguridad Internacional*.

Chermak, S., & Gruenewald, J. A. (2015). Laying a foundation for the criminological examination of right-wing, left-wing, and Al Qaeda-inspired extremism in the United States. *Terrorism and Political Violence* 27(1), 133-159.

Cherney, A. (2018). Supporting disengagement and reintegration: Qualitative outcomes from a custody-based counter radicalisation intervention. *Journal for Deradicalization*, (17), 1-27.
<https://journals.sfu.ca/jd/index.php/jd/article/view/174>.

Cherney, A. (2020). Evaluating interventions to disengage extremist offenders: A study of the Proactive Integrated Support Model (PRISM). *Behavioral Sciences of Terrorism and Political Aggression* 12(1), 17-36. <https://doi.org/10.1080/19434472.2018.1495661>.

Cherney, A., & Belton, E. (2021). Evaluating case-managed approaches to counter radicalization and violent extremism: An example of the Proactive Integrated Support Model (PRISM) intervention. *Studies in Conflict & Terrorism* 44(8), 625-645. <https://doi.org/10.1080/1057610X.2019.1577016>.

Christensen, T. W. (2015). How extremist experiences become valuable knowledge in EXIT programmes. *Journal for Deradicalization*, (3), 92-134.

Corner, E., & Gill, P. (2019). Psychological distress, terrorist involvement and disengagement from terrorism: A sequence analysis approach. *Journal of Quantitative Criminology*, 36(3), 499–526. <https://doi.org/10.1007/s10940-019-09420-1>.

da Silva, R., Fernández-Navarro, P., Gonçalves, M. M., Rosa, C., & Silva, J. (2019). Tracking narrative change in the context of extremism and terrorism: Adapting the innovative moments coding system. *Aggression and Violent Behavior*, 47, 204-214. <https://doi.org/10.1016/j.avb.2019.05.002>.

Daugherty, C.E. (2019). Deradicalization and disengagement: Exit programs in Norway and Sweden and addressing neo-Nazi extremism. *Journal for Deradicalization*, (21), 219-260.

De Jong, J. T. V. M. (2010). A public health framework to translate risk factors related to political violence and war into multi-level preventive interventions. *Social Science & Medicine*, 70(1), 71-79.

Disley, E., Weed, K., Reding, A., Clutterbuck, L., & Warnes, R. (2011). *Individual disengagement from Al Qa'ida-influenced terrorist groups: A rapid evidence assessment to inform policy and practice in preventing terrorism*. RAND Corporation.

Doosje, B., Moghaddam, F. M., Kruglanski, A. W., De Wolf, A., Mann, L., & Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11, 79-84.

Dugas, M., & Kruglanski, A. W. (2014). The quest for significance model of radicalization: Implications for the management of terrorist detainees. *Behavioral Sciences and the Law* 32, 423-439.

Ehiane, S. O. (2019). De-radicalisation and disengagement of the extremist group in Africa: the Nigerian experience. *Journal of African Foreign Affairs*, 6(2), 123-138.

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

Fekete, L. (2015). Rewiring fascists: Behind the cult of Exit. *Race & Class*, 56(3), 91-101. <https://doi.org/10.1177/0306396814556226>.

Ferguson, N. (2016). Disengaging from terrorism: A Northern Irish experience. *Journal for Deradicalization*, (6), 1-28.

Frenett, R. & Dow, M. (2015). *One to one online interventions – A pilot CVE methodology*. Institute for Strategic Dialogue.

Ganor, B. & Falk, O. (2013). De-Radicalization in Israel's prison system. *Studies in Conflict & Terrorism*, 36(2), 116–131. <https://doi.org/10.1080/1057610x.2013.747071>.

Gielen, A.-J. (2018). Exit programmes for female jihadists: A proposal for conducting realistic evaluation of the Dutch approach. *International Sociology*, 33(4), 454–472. <https://doi.org/10.1177/0268580918775586>.

Götsch, K. (2017). Austria and the threats from Islamist radicalisation and terrorist involvement: An overview of governmental and non-governmental initiatives and policies. *Journal for Deradicalization*, (12), 169-191.

Grip, L., & Kotajoki, J. (2019). Deradicalisation, disengagement, rehabilitation and reintegration of violent extremists in conflict-affected contexts: A systematic literature review. *Conflict, Security & Development*, 19(4), 371–402. <https://doi.org/10.1080/14678802.2019.1626577>.

Hasisi, B., Carmel, T., Weisburd, D., & Wolfowicz, M. (2020). Crime and terror: Examining criminal risk factors for terrorist recidivism. *Journal of Quantitative Criminology*, 36(3), 449-472.

Heinke, D. H., & Raudszus, J. (2018). Germany's Returning Foreign Fighters and What to Do About Them. *Coolsaet and Renard (see note 16)*, 47-48.

Hellmuth, D. (2015). Countering jihadi radicals and foreign fighters in the United States and France: très similaire. *Journal for Deradicalization*, (4), 1-43.

Horgan, J. (2009). Deradicalization or disengagement? A process in need of clarity and a counterterrorism initiative in need of evaluation. *Revista De Psicología Social*, 24(2), 291–298. <https://doi.org/10.1174/021347409788041408>.

Horgan, J., & Braddock, K. (2010). Rehabilitating the terrorist?: Challenges in assessing the effectiveness of de-radicalization programs. *Terrorism and Political Violence*, 22, 267-291. DOI: 10.1080/09546551003594748.

Hwang, J. C., Panggabean, R., & Fauzi, I. A. (2013). The disengagement of jihadis in Poso, Indonesia. *Asian Survey*, 53(4), 754-777.

Khan, A. u. (2017). *Pakistan's Strategy of Countering Violent Extremism: Need for Soft Power*. Issi.org. Retrieved January 27, 2022, from https://issi.org.pk/wp-content/uploads/2017/10/10-SS_Asad_ullah_Khan_No-3_2017.pdf.

Koehler, D. (2015). De-radicalization and disengagement programs as counter-terrorism and prevention tools. Insights from field experiences regarding German right-wing extremism and jihadism. *Countering radicalisation and violent extremism among youth to prevent terrorism*, 118, 120-150.

Koehler, D. (2017a). How and why we should take deradicalization seriously. *Nature Human Behaviour* 1, 1-4.

Koehler, D. (2017b). *Structural quality standards for work to intervene with and counter violent extremism*. German Institute on Radicalisation and De-radicalisation Studies and Ministry of the Interior, Digitisation and Migration Baden-Württemberg. <http://girds.org/publications/handbook-for-structural-quality-standards-in-deradicalization-work>.

Koehler, D. (2020a). *Stand-alone de-radicalisation programme evaluation tool for stakeholders*. DARE: Dialogue About Radicalisation and Equality. https://www.dare-h2020.org/uploads/1/2/1/7/12176018/dare-de-radicalisation_evaluation_tool.pdf.

Koehler, D. (2020b). Terminology and definitions. In S. J. Hansen (Ed.), *Routledge handbook of deradicalisation and disengagement* (pp. 11-25). Routledge.

Konze, A. (2016). Deradicalisation of foreign fighters. *Criminal Justice and Security in Central and Eastern Europe*, 351(352), 277.

Kruglanski, A. W., Gelfand, M. J., Belanger, J. J., Gunaratna, R., & Hettiarachchi, M. (2014a). De-radicalising the Liberation Tigers of Tamil Eelam (LTTE). In A. Silke (Ed.), *Prisons, terrorism and extremism: Critical issues in management, radicalisation and reform* (pp. 183-196). Routledge.

Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hettiarachchi, M., & Gunaratna, R. (2014b). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Advances in Political Psychology* 35(1), 69-93.

Kutner, S. (2016). The call for component analyses of the Saudi Arabian Risk Reduction Initiative: An examination of religious re-education's role in the deradicalization and disengagement process. *Journal for Deradicalization*, (7), 107-123. <https://journals.sfu.ca/jd/index.php/jd/article/view/59/50>.

Leistedt, S. J. (2016). On the radicalization process. *Journal of forensic sciences*, 61(6), 1588-1591.

Marsden, S. V. (2015). Conceptualising 'success' with those convicted of terrorism offences: Aims, methods, and barriers to reintegration. *Behavioral Sciences of Terrorism and Political Aggression* 7(2), 143-165.

McManus, A. L. (2019). Egypt's foreign terrorist fighters: Cycles of violence. In T. Renard (Ed.), *Returnees in the Maghreb: Comparing policies on returning foreign fighters in Egypt, Morocco, and Tunisia* (pp. 9-23). Egmont Institute.

McRae, D. (2010). Reintegration and localised conflict: security impacts beyond influencing spoilers. *Conflict, Security & Development*, 10(3), 403-430.

Meloy, J. R., & Gill, P. (2016). The lone-actor terrorist and the TRAP-18. *Journal of Threat Assessment and Management* 3(1), 37-52.

Mitchell, S. (2016). Deradicalization: Using triggers for the development of a US program. *Journal for Deradicalization*, (9), 101-125.
<https://journals.sfu.ca/jd/index.php/jd/article/view/74>.

Morrison, J. F., Silke, A., Maiberg, H., Slay, C., & Stewart, R. (2021). A Systematic Review Of Post-2017 Research On Disengagement And Deradicalisation. *Full Report. Crest*.

Onapajo, H. & Ozden, K. (2020). Non-military approach against terrorism in Nigeria: deradicalization strategies and challenges countering Boko Haram. *Security Journal*, 33, 476-492. <https://doi.org/10.1057/s41284-020-00238-2>.

Osman, S. (2014). *Radicalisation, recidivism and rehabilitation: Convicted terrorists a*. Taylor & Francis. Retrieved January 27, 2022, from <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203584323-15/radicalisation-recidivism-rehabilitation-sulastri-osman>.

Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews *BMJ* 2021; 372.

Parrin, A. (2016). Creating legal framework for terrorism defectors and detainees in Somalia. *Columbia Journal of Transnational Law* 55(1), 228-276.

Porges, M. L. (2010). Deradicalisation, the Yemeni way. *Survival*, 52(2), 27-33.

Richards, J. (2017). Demobilising and disengaging violent extremists: Towards a new UN framework. *Stability: International Journal of Security and Development*, 6(1).
<https://doi.org/10.5334/sta.543>.

- Richards, J. (2018). High risk or low risk: Screening for violent extremists in DDR Programmes. *International Peacekeeping*, 25(3), 373–393. <https://doi.org/10.1080/13533312.2018.1440177>.
- Rushchenko, J. (2019). Terrorist recruitment and prison radicalization: Assessing the UK experiment of ‘Separation Centres.’ *European Journal of Criminology*, 16(3), 295–314. <https://doi.org/10.1177/1477370819828946>.
- San, S. (2020). Counter-terrorism policing innovations in Turkey: a case study of Turkish National Police CVE experiment. *Policing & Society*, 30(5), 583-598.
- Savoia, E., McBride, M., Stern, J., Su, M., Harriman, N., Aziz, A., & Legault, R. (2020). Assessing the impact of the Boston CVE pilot program: A developmental evaluation approach. *Homeland Security Affairs* 16(6), 1-50. <http://www.hsaj.org/articles/16166>.
- Schmid, A. P. (2013). Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review. *International Centre for Counter-Terrorism — The Hague* 4, no. 2.
- Schuurman, B. (2018). *Becoming a European homegrown jihadist: A multilevel analysis of involvement in the Dutch Hofstadgroup, 2002-2005*. Amsterdam University Press.
- Schuurman, B., & Bakker, E. (2015). Reintegrating jihadist extremists: Evaluating a Dutch initiative, 2013–2014. *Behavioral Sciences of Terrorism and Political Aggression*, 8(1), 66–85. <https://doi.org/10.1080/19434472.2015.1100648>.
- Silke, A. (2001). The devil you know: Continuing problems with research on terrorism. *Terrorism and Political Violence*, 13(4), 1-14.
- Silke, A., & Schmidt-Petersen, J. (2017). The golden age? What the 100 most cited articles in terrorism studies tell us. *Terrorism and Political Violence*, 29(4), 692-712.
- Sivenbring, J., & Andersson Malmros, R. (2020). Mixing logics: Multiagency approaches for countering violent extremism.
- Souleimanov, E. A., & Aliyev, H. (2020). Ideology and disengagement: A case study of nationalists and Islamists in Chechnya. *Europe-Asia Studies*, 72(2), 314-330.
- Spalek, B. (2016). Radicalisation, de-radicalisation and counter-radicalisation in relation to families: Key challenges for research, policy and practice. *Security Journal*, 29(1), 39-52.
- Suarda, I. G. W. (2016). A literature review on indonesia’s deradicalization program for terrorist prisoners. *Mimbar Hukum-Fakultas Hukum Universitas Gadjah Mada*, 28(3), 526-543.
-

Sumpter, C. (2017). Countering violent extremism in Indonesia: priorities, practice and the role of civil society. *Journal for Deradicalization*, (11), 112-147.

van der Heide, E. J., & Schuurman, B. W. (2018). Reintegrating terrorists in the Netherlands: Evaluating the Dutch approach. *Journal for Deradicalization*, (17), 196-239.

Webber, D., Chernikova, M., Kruglanski, A. W., Gelfand, M. J., Hetiarachchi, M., Gunaratna, R., Lafreniere, M.-A., & Belanger, J. J. (2018). Deradicalizing detained terrorists. *Political Psychology*, 39(3), 539-556. doi: 10.1111/pops.12428.

Williams, M. J., & Kleinman, S. M. (2014). A utilization-focused guide for conducting terrorism risk reduction program evaluations. *Behavioral Sciences of Terrorism and Political Aggression*, 6(2), 102-146. <https://doi.org/10.1080/19434472.2013.860183>.

Williams, M. J., & Lindsey, S. C. (2014). A social psychological critique of the Saudi terrorism risk reduction initiative. *Psychology, Crime & Law*, 20(2), 135-151. <https://doi.org/10.1080/1068316X.2012.749474>.

Webber, D., & Kruglanski, A. W. (2018). The social psychological makings of a terrorist. *Current Opinion in Psychology*, 19, 131–134. <https://doi.org/10.1016/j.copsyc.2017.03.024>.

Windisch, S., Scott Ligon, G., & Simi, P. (2017). Organizational [Dis]trust: Comparing Disengagement Among Former Left-Wing and Right-Wing Violent Extremists. *Studies in Conflict & Terrorism*, 42(6), 559–580. <https://doi.org/10.1080/1057610x.2017.1404000>.

Wittendorp, S., Bont, R. D., Bakker, E., & Roy van Zuijdewijn, J. D. (2017). Measures against Jihadist Foreign Fighters: A policy comparison between the Netherlands, Belgium, Denmark, Germany, France, the UK and the US (2010 to 2017). Measures against Jihadist Foreign Fighters: A policy comparison between the Netherlands, Belgium, Denmark, Germany, France, the UK and the US (2010 to 2017).

Yehoshua, S. (2014). The Israeli Experience of Terrorist Leaders in Prison: Issues in radicalisation and de-radicalisation. In *Prisons, Terrorism and Extremism* (pp. 144-156). Routledge.

Appendix 1

Australian Strategic Policy Institute
Brookings Institution
Combating Terrorism Center at West Point
Carnegie Endowment for International Peace
The Center for Strategic and International Studies (CSIS)
Council on Foreign Relations (CFR)
Counter-Terrorism Committee of the United Nations
Defense Technical Information Center (DTIC)
EXIT-Deutschland
European Union
European Union Institute for Security Studies
Program on Extremism at George Washington University
German Institute of Radicalization and Deradicalization Studies (GIRDS)
Global Counterterrorism Forum
International Centre for Counter-Terrorism (The Hague)
The International Centre for the Study of Radicalisation and Political Violence
Institute for Strategic Dialogue (ISD)
International Institute for Counter-Terrorism
International Peace Institute
Leiden University
Middle East Institute (MEI)
National Institute of Justice (NIJ)
Organization for Security and Cooperation in Europe (OSCE)
RAN (Radicalization Awareness Network)
RAND Corporation
Resolve
TerRa (Terrorism and Radicalisation)
Congressional Testimony
The Soufan Group
The Unity Initiative
National Consortium for the Study of Terrorism and Responses to Terrorism (START)
UN Office of Drugs and Crime
US Director of National Intelligence (DNI)
US Department of Homeland Security (DHS)
US Government Accountability Office (GAO)
US National Counterterrorism Center (NCTC)
United States Institute of Peace (USIP)
Violence Prevention Network
War on the Rocks
Washington Institute for Near East Policy (WINEP)
Wilson Center

Egmont Institute
Heinrich Boll, S. Rajaratnam School
The International Centre for the Study of Radicalisation (ICSR)
Danish Institute for International Studies
Norwegian Institute of International Affairs
Swedish National Council for Crime Prevention (Brottsförebyggande rådet, BRÅ)

About the JD Journal for Deradicalization

The JD Journal for Deradicalization is the world's only peer reviewed periodical for the theory and practice of deradicalization with a wide international audience. Named an [“essential journal of our times”](#) (Cheryl LaGuardia, Harvard University) the JD's editorial board of expert advisors includes some of the most renowned scholars in the field of deradicalization studies, such as Prof. Dr. John G. Horgan (Georgia State University); Prof. Dr. Tore Bjørgo (Norwegian Police University College); Prof. Dr. Mark Dechesne (Leiden University); Prof. Dr. Cynthia Miller-Idriss (American University Washington D.C.); Prof. Dr. Julie Chernov Hwang (Goucher College); Prof. Dr. Marco Lombardi, (Università Cattolica del Sacro Cuore Milano); Dr. Paul Jackson (University of Northampton); Professor Michael Freeden, (University of Nottingham); Professor Hamed El-Sa'id (Manchester Metropolitan University); Prof. Sadeq Rahimi (University of Saskatchewan, Harvard Medical School), Dr. Omar Ashour (University of Exeter), Prof. Neil Ferguson (Liverpool Hope University), Prof. Sarah Marsden (Lancaster University), Prof. Maura Conway (Dublin City University), Dr. Kurt Braddock (American University Washington D.C.), Dr. Michael J. Williams (The Science of P/CVE), Dr. Mary Beth Altier (New York University) and Dr. Aaron Y. Zelin (Washington Institute for Near East Policy), Prof. Dr. Adrian Cherney (University of Queensland), Dr. Håvard Haugstvedt (Center for Research on Extremism, University of Oslo).

For more information please see: www.journal-derad.com

Twitter: @JD_JournalDerad

Facebook: www.facebook.com/deradicalisation

The JD Journal for Deradicalization is a proud member of the Directory of Open Access Journals (DOAJ).

ISSN: 2363-9849

Editor in Chief: Daniel Koehler