Routledge
Taylor & Francis Group

A utilization-focused guide for conducting terrorism risk reduction program evaluations

Michael J. Williams[a]* and Steven M. Kleinman[b]

[a]*Interdisciplinary Social Psychology Department, University of Nevada, Mailstop 300, Reno, NV 89557, USA;* [b]*The Soufan Group, 410 Park Avenue, 15th Floor, New York, NY 10022, USA*

The present work employs a utilization-focused evaluation perspective to ask the big question regarding so-called deradicalization programs: how to evaluate the degree to which a given terrorism risk reduction initiative reduces post-detainment terrorism engagement. Its dual objectives are: (a) to provide a roadmap for conducting such an impact analysis with a utilization-focus, and (b) to highlight some of the unique challenges (both methodologically and theoretically) that face evaluators in the context of evaluating terrorism risk reduction initiatives. Additionally, the appendices of this work contain both a process checklist for conducting an impact analysis of such initiatives, and an evaluation self-assessment tool.

Keywords: terrorism risk reduction; deradicalization; terrorism; evaluation; countering violent extremism

In 2010, the US Department of defense released the disturbing news that one in five Guantanamo Bay detainees (74 total) had resumed terrorist activity post-release (Mount, 2010). Though deeply troubling, that statistic is not especially surprising, given that Guantanamo Bay merely incarcerates detainees and does not include a rehabilitative component (Williams & Lindsey, 2013). In short, the capture and detention of suspected or confirmed terrorists is, at best, a short-term response to curbing terrorism, not a long-term solution (Soufan et al., 2013).

Several nations have responded to the needs both to release certain detainees, and to maintain national security, by implementing so-called terrorist "deradicalization" programs[1] (Shane, 2009). However, the effectiveness of such programs remains empirically unproven, and post-detainment terrorism engagement has been found among those who have, and have not, participated in them (Morris, Eberhard, Rivera, & Watsula, 2010). Fortunately, many officials that have implemented deradicalization programs agree that systematic program evaluations are important (Soufan et al., 2013). Unfortunately, no one has done those (Soufan et al., 2013).

There has been similar agreement about the need for establishing meaningful metrics that can provide critical feedback, both to program managers and to senior officials, regarding the effectiveness of established de-radicalization programs. Many of these same individuals, however, view the field as still in its infancy and, as a result, believe there is insufficient data upon which to build these metrics. Often, the only

*Corresponding author. Email: mjwilliams@unr.edu

real "metric" currently used is whether or not there has been a terrorist attack since the program was initiated (Soufan et al., 2013).

There are many types of program evaluations, to furnish several kinds of answers, and any of those may be pursued with what may be called a utilization focus (Patton, 2008). Specifically, any evaluation can – and, arguably should – be conducted for (and with) specific intended users, for their specific, intended purposes (Patton, 2008). The present work employs a utilization-focused program evaluation perspective to ask the big question regarding so-called deradicalization programs: how to evaluate the degree to which a given terrorist deradicalization program reduces post-detainment terrorism engagement. As such, the present work represents a guide for conducting an impact analysis on that topic, and it has two objectives: (a) to provide a roadmap for conducting such an analysis with a utilization-focus, and (b) to highlight some of the unique challenges (both methodologically and theoretically) that likely face evaluators in the context of evaluating deradicalization programs. Additionally, the appendices of this work contain both a process checklist for conducting an impact analysis of terrorism risk reduction initiatives, and a self-assessment tool for evaluators of such initiatives (see Appendices 1 and 2, respectively).

An assumption and terms

An assumption of this guide is that the program being evaluated is preexisting, and fully functioning in accordance with the mission statement set forth in the founding documentation for a given program. Although that assumption may seem obvious to some (after all, how can one evaluate a non-existent program?), that is a narrow view of the scope of program evaluation. As mentioned, there are several types of evaluations; one of which (formative evaluation) is specifically intended to guide programs through developmental stages (Patton, 2008; Wholey, Hatry, & Newcomer, 2010). In addressing ways to improve evaluation of anticrime programs, Lipsey, Petrie, Weisburd, and Gottfredson (2006) point out that, whenever possible, even nascent programs should be developed in ways that incorporate plans for impact evaluations: to build in mechanisms for quality control, service improvement, and to document program outcomes (including program successes; Wholey et al., 2010). Although this guide is explicitly intended to guide impact evaluations of extant programs, it is hoped that this guide also may inform those engaged in deradicalization program development. By envisioning components required for an impact evaluation (e.g. outcome measures), program developers can integrate systems into their programs that will facilitate subsequent impact evaluations. Nevertheless, given that the present work is a guide for evaluating the impact of a given program, it is assumed that the program is up and running.

Before conducting an impact evaluation, research on evaluations of criminal justice programs highlights the necessity that guarantees must be in place, to assure that evaluators will have access to relevant outcome data (Lipsey et al., 2006). Among the pitfalls related to data access is that individuals' criminal offense records may be stored in de-centralized locations, and may not be available to researchers without legal authorizations (Lipsey et al., 2006). Furthermore, for ethical reasons, program managers may be unwilling to provide evaluators with certain outcome data that they feel is either confidential or proprietary (Lipsey et al., 2006).

Further considerations regarding the "evaluability" (Wholey et al., 2010) of an extant program include, of course, sufficient funding for personnel (evaluators,

program staff, and – perhaps – participants), but also sufficient time allotted for the evaluation, and that there are sufficient numbers of participants who have been, or will be, tracked for sufficient time to assess with reasonable confidence the program's impact (Lipsey et al., 2006). In summary, aside from an appropriate research design, preconditions for an effective evaluation include (a) that the program is operational, (b) that researchers have access to sufficient amounts (and kinds) of data, and (c) that there are sufficient resources (time, money, personnel) available.

The term "deradicalization program" is often misused, in light of the finding that many rehabilitated terrorists retain aspects of former radical ideologies (Horgan, 2009a, 2009b; Horgan & Taylor, 2011). Instead, some favor the term "disengagement program:" suggesting that the objective of such programs is not to change former terrorists' radical ideologies, but merely to persuade them to remain disengaged from their former terrorist activities (Horgan, 2009a, 2009b). However, disagreement remains regarding the most appropriate term (Hannah, Clutterbuck, & Rubin, 2008; Horgan & Taylor, 2011; Rabasa, Pettyjohn, Ghez, & Boucek, 2010). The present work opts for terminology, forwarded by Horgan and Braddock (2010), to describe so-called deradicalization programs as "terrorism risk reduction initiatives:" a term that encompasses the nature of such programs, without implying – nor precluding – ideological change on behalf of former terrorists.

The second term to make clear is what constitutes the primary outcome/dependent variable to evaluate. In short, what criterion should be measured, by which we may estimate a given program's success? The answer is both simple and complex; the simple answer is one must measure "post-detainment terrorism engagement". The complexity comes in defining such engagement (Barrett & Bokhari, 2009; Rabasa et al., 2010; Williams & Lindsey, 2013). Presently, consensus has not been reached – either among the research community or among nations who implement terrorism risk reduction initiatives – regarding successful terrorist rehabilitation (Horgan & Braddock, 2010). Whereas the detonation of a deadly explosive device against civilians for political purposes easily can be considered an act of terrorism, what of funding terrorism, or the production of terrorist propaganda, or simply verbally encouraging others to join a terrorist organization? Clearly, there are myriad outcome variables that could be measured as indicators of post-detainment terrorism engagement. Of course, the measurement of multiple indicators might provide a clearer picture of the effectiveness of a given terrorism risk reduction initiative: but which to measure, and which outcome variable is most indicative of an initiative's success? Further complicating that question is – regardless of what outcomes are measured – should success be measured by an absolute value (e.g. 10 incidents of post-detainment terrorism engagement per year), the percentage of such engagement for a given year, or change over time (e.g. a 10% reduction of post-detainment terrorism engagement compared to the previous year; Wilson, 1993). Aside from such aggregate/macro-level questions, criminal justice research regarding prediction of rehabilitation successes highlights that important questions also arise regarding post-detainment terrorism engagement rates based on offender characteristics (e.g. age, marital status, employment status, type of terrorism offense; Winokur, 2002). The question becomes not simply "does the program work", but for whom does it work the best (and worst). The answer to all of these questions, regarding what to measure and how to measure post-detainment terrorism engagement, is at the very heart of utilization program evaluation: let the stakeholders decide.

The stakeholders

Stakeholders are those who will use the evaluation's findings, support or maintain the initiative, or who are affected by the initiative's activities or evaluation results (United Way, 2008). As such, they are those who are most impacted by a given program: which is why they are entitled to deciding upon the outcome variable(s) they believe are most worthy of measurement (United Way, 2008). From the perspective of researchers who aspire to compare the effectiveness of different terrorism risk reduction initiatives (hereafter "initiatives" or "programs"), it would be ideal if international consensus could be reached regarding standardized definitions of risk reduction outcome variables: especially those that define post-detainment terrorism engagement. However, it is important to note that – although the international research community could be considered a stakeholder – the interests of that community seem less compelling than the interests of the citizens of the nations within whose borders the ostensibly rehabilitated terrorists will be released, and who may incur the greatest casualties should that rehabilitation fail. Therefore, internationally standardized definitions of risk reduction outcome variables, including a definition of post-detainment terrorism engagement, may elude the research community: which is perfectly acceptable from a utilization-focused evaluation perspective, provided that a given evaluation's definitions serve the self-identified needs of the stakeholders (Patton, 2008).

Identifying and consulting stakeholders

Given the paramount importance of serving stakeholders, one of the first parts of a utilization-focused evaluation is the identification of all reasonably affected stakeholders, which include: those who will primarily use the evaluation's results, support or otherwise maintain the program, or who are clearly affected by the program's activities (Patton, 2008; United Way, 2008; Wholey et al., 2010). This is important for at least two reasons. First, stakeholders are uniquely positioned, as those with indigenous or otherwise "insider" information, to accurately inform the decisions that must be made throughout the evaluation (Soufan et al., 2013). Second, involving stakeholders in decision-making processes tends to increase their "buy in" of the evaluation: that they will be more likely to understand the purpose of the evaluation, to cooperate in its implementation, to understand its results, and to make use of those results (Patton, 2008; Wholey et al., 2010). In addition, well-engaged stakeholders can determine and prioritize key evaluation questions and increase the general public's perceived credibility of the results (United Way, 2008). Such buy-in is central to utilization-focused evaluations: what has been referred to as the process of participatory evaluation planning (Centers for Disease Control and Prevention, 2011a).

Of particular importance in identifying stakeholders is identifying what could be called either "lead" or "key" stakeholders. These are individuals toward the top of a given organization who have the political clout and/or the social capital to rally other stakeholders to the evaluation's cause, and who can help maintain stakeholders' engagement in the evaluation throughout the process. In short, commitment at the top of the organization should be considered essential (United Way, 2008). In accord with this notion, research on successful criminal justice program evaluations maintains that the likelihood of such evaluations succeeding is greatly reduced when they are imposed from outside the organization (e.g. as a requirement from funders; Lipsey et al., 2006).

In the case of risk reduction initiatives, key stakeholders can include (but are not limited to) the following: national policy-makers, security organizations (e.g. police and intelligence agencies), social services (e.g. welfare agencies), educational institutions, mental health professionals, community leaders (especially in cases where a "community" is largely defined by a specific ethnic or national identity), reformed ex-terrorists, religious scholars, administrators who facilitate interagency networking, and family members of program participants (Soufan et al., 2013). Additionally, community members from locations where rehabilitated program participants will be released should be considered stakeholders, and their needs – including informational needs – should be incorporated in the evaluation plan (International Peace Institute, 2010). Specifically, despite a perhaps high-quality rehabilitation of program participants, if the community is unconvinced that participants are no longer a significant threat, the program will lack credibility (International Peace Institute, 2010). Consequently, program participants might find it difficult to reintegrate into an unwelcoming community (International Peace Institute, 2010).

It is recommended that evaluators begin by brainstorming prospective stakeholders according to the purposes of the evaluation. As mentioned, the stakeholders are those who will primarily use the evaluation's results, support or otherwise maintain the program, or who are clearly affected by the program's activities. After listing prospective stakeholders, it might be useful for evaluators to organize the stakeholders in a hierarchical network, to understand which are highest-ranking and most-connected; likely those persons are key stakeholders (see Wholey et al., 2010, for a complete description of this process).

Selecting the evaluation personnel

Aside from experience, methodological know-how (in both quantitative and qualitative methods, as well as facilitating focus groups and group decision-making), one of the considerations regarding the selection of evaluation personnel should be their level of cultural awareness. In principle, "who could argue against taking into account the cultural context when designing and conducting an evaluation?" (Frierson, Hood, & Hughes, 2010, p. 64). As mentioned, cultural responsiveness includes gaining input from stakeholders, and it also includes selecting evaluation personnel who have cultural awareness not only of the program and its stakeholders, but of their own cultural biases. To be clear, there are no culture-free evaluators (Frierson et al., 2010). Nevertheless, part of the beauty of utilization-focused evaluations is that evaluators' biases, including cultural biases, are mitigated to the extent that stakeholders drive both the agenda and decision-making. However, evaluators' cultural awareness may be an asset to a given evaluation, if for no other reasons than it may facilitate their communication with stakeholders and help the evaluators to recognize culture-specific challenges that the evaluation must surmount (e.g. what are the most culturally accepted ways to collect data from men vs. women in a given culture when using male vs. female data collectors).

Defining the problem and identifying goals

After the essential task of identifying key stakeholders, evaluators should gather them to begin structuring the evaluation. Among the chief items on the agenda of that

meeting is for the stakeholders to come to consensus on a statement of the problem (United States Government Accountability Office, 2012). Additionally, they must come to mutually agreed-upon terms about the goals of the evaluation (United States Government Accountability Office, 2012).

In defining those goals, consideration should be paid both to the magnitude of post-detainment terrorism engagement that would be considered successful, and the timeframe for that effect to become manifest. Of course, when discussing terrorism prevention, the notion of anything less than total success is, perhaps (understandably) anathema to most; however, it is also unrealistic for that standard to be expected into the indefinite future. Therefore, programmatic goals should be time-bound. Regarding effect sizes, Lipsey et al. (2006) point out that, for large criminal populations, a reduction of post-detainment criminal activity amounting to a mere 1/10th of its standard deviation would produce a large social benefit.

It is important to note that stakeholders' goals might not be related solely to post-detainment terrorism engagement. Rather, they might be concerned with whether the initiative's resources are being used efficiently, or whether there are any unintended side effects of the initiative, or whether the initiative is being carried out as planned (United States Government Accountability Office, 2012). Though by no means exhaustive, the following questions, suggested by the W.K. Kellogg Foundation (1998), can help guide stakeholders in identifying their goals both for their program's outcomes, and for the information that the evaluation could provide. What do stakeholders need to know more about? More specifically, what decisions do stakeholders need to make about which they seek more information to inform those decisions? To what uses will they put the answers to those questions? Specifically, who will make the decisions and when? How are decisions made within stakeholders' respective organizations? How do the stakeholders know if the initiative is running as planned? Also, as mentioned, it is important to understand what subpopulations of offenders are most responsive to the initiative. As implied by these questions, the most important research questions among those related to post-detainment terrorism engagement are those that have policy-relevant implications (Lipsey et al., 2006).

The answers to questions such as those questions will help to form the basis of the goals both for the program and for the evaluation (W.K. Kellogg Foundation, 1998). Those prospective goals represent the myriad interests of the stakeholders that might be in competition with one another, perhaps for no other reason than the limited financial resources and time afforded to any given evaluation might not permit each question to be answered. Therefore, after all of stakeholders' prospective goals have been identified, likely they will need to be prioritized (W.K. Kellogg Foundation, 1998). To accomplish that objective, evaluators ought to be well versed in orchestrating group decision-making. A thorough discussion of those skills and processes is beyond the scope of the present work, though it should be mentioned that such processes might include collating and posting the stakeholders' goals, then having each stakeholder vote for what they feel are (for example) the top two or three most pressing goals (see Patton, 2008). The evaluator's objective at this point is to focus the evaluation to those goals that are both shared by the greatest number of stakeholders, and that can be plausibly accomplished within the financial and temporal confines of the present evaluation (Centers for Disease Control and Prevention, 2011a).

Describing the program

Having completed the previous stage in the evaluation, though perhaps tempting, evaluators should not proceed to develop the evaluation's research design without continuing to gather information about the initiative's effectiveness (Wholey et al., 2010). Such information gathering includes evaluators' first-hand contact with the initiative's activities and their independent review of literature regarding the initiative (Wholey et al., 2010). The objective is for evaluators to ascertain what is known about the initiative's resources (including staff), activities, and technologies that ostensibly contribute to the program's outcomes (Centers for Disease Control and Prevention, 2011b; Wholey et al., 2010). Research on theory-based criminal justice evaluations emphasizes that, given this information will be used to develop a theory-based evaluation, another critical objective of the information-gathering phase is to surface key assumptions about how the program is thought to affect change (Cooper & Worrall, 2012).

During this information-gathering process, evaluators should consider interviewing program staff and program participants (McNamara, 2006). Ideally, evaluators should speak with "satisfied" staff and "successful" participants, in addition to speaking with staff that have quit the program and participants who have "failed". Important insights regarding the program might be gained by understanding not only its successes, but its shortcomings and failures (McNamara, 2006).

Assessment protocols

Among the chief pieces of information to gather regarding how an initiative functions is its risk assessment protocol (Latessa & Lovins, 2010). Fundamentally, the question is: how are risk determinations made, at every phase of the program? Specifically, the question becomes on what evidence (if any) are risk assessments made. Research from the past five decades has demonstrated that actuarial (i.e. statistical) methods of predicting post-detainment offending tend to be more accurate than subjective clinical assessment (Latessa & Lovins, 2010). However, actuarial methods are only as good as the evidence upon which they are based. Therefore, questions remain regarding that evidence. For example, to what extent are the actuarial samples similar to the program's participants (Latessa & Lovins, 2010)? To what extent are the cultural and legal structures from which the actuarial assessments were derived similar to those under evaluation (Latessa & Lovins, 2010)? Additionally, as a matter of implementation fidelity, it is important to gather information regarding how well, and consistently, the risk assessment protocols are followed. Related questions include, to what degree are risk assessors trained in the use of the protocols?

Not only are objective assessments important for gauging program participants' risk, but for assigning appropriate treatments. For example, intake/triage assessments are important, if for no other reason, than the counter intuitive reason that subjecting relatively low risk offenders to intensive rehabilitation treatments can increase their risk of post-detainment offending (Lowenkamp & Latessa, 2004). Additionally, research on the Saudi terrorism risk reduction initiative suggests that treatment triage should take into consideration participants' motives for their offenses (Williams & Lindsey, 2013).

Confirming the ends, ways, and means of the initiative and the evaluation

In evaluation parlance, ends, ways, and means are typically referred to as outcomes, outputs, and inputs respectively (Patton, 2008). It should go without

saying, but – for an initiative to succeed – those elements must logically cohere. However, in actuality, it is possible for a given program's inputs, outputs, and outcomes to be poorly or illogically coordinated, such that either the program does not work as intended, or that it might produce desired results, but for reasons other than those assumed by the program's managers (Patton, 2008).

At this stage, it is highly recommended that evaluators graphically represent the initiative's resources (inputs), activities (outputs), and outcomes in a form typically referred to as a logic model (Patton, 2008; United States Government Accountability Office, 2012; Wholey et al., 2010). The purpose of the logic model is to lay bare both the mechanics and the assumptions of how and why the program ostensibly works (Patton, 2008; Wholey et al., 2010). In doing so, flaws or weaknesses in the program's logic are likely to be more readily perceived both by evaluators and by stakeholders (Patton, 2008).

Theory of change

Additionally, logic models afford evaluators a bird's-eye-view of the initiative that can assist them in identifying the initiative's theory of change (Brookings Institution, 1998; Patton, 2008). As the name implies, the theory of change is the most plausible reason why a given initiative produces its results (Patton, 2008; Wholey et al., 2010). By no means should it be assumed that up-and-running risk reduction initiatives are based upon coherent theories of change. For example, a recent descriptive study of initiatives in Singapore, Indonesia, Northern Ireland, Great Britain, and France found that, although most had goals, few had clearly defined, objectively based strategies to achieve them (Soufan et al., 2013).

The importance of identifying the initiative's theory of change cannot be overstated, for two basic reasons. First, for a program to be improved, the mechanisms of how the program works must be reasonably understood. To do otherwise is analogous to trying to tune an automobile without understanding how the automobile functions. Second, for the evaluation to inform usefully those who would use the evaluation's findings, it must measure useful constructs. In other words, the evaluation must take the pulse of the program from its vital components. For a discussion of the development of logic models, see Patton (2008) and Wholey et al. (2010).

By no means must there be only one theory of change in effect for a given initiative (Patton, 2008). To the contrary, most initiatives attempt to rehabilitate their participants using several methods, to influence participants in several ways, from psychological and religious counseling to socioeconomic and environmental interventions (Soufan et al., 2013). Consequently, each of those rehabilitative methods might have different theories of change that (one may hope) work in concert. Furthermore, assuming that multiple theories of change may be in effect, it is important to consider how they interact. For example, a study by the RAND corporation found that rehabilitative components intended to reorient participants' extremist ideologies (e.g. through discussions with counter ideologues), had little effect unless those initiatives also included components that addresses participants other emotional and practical needs (Rabasa et al., 2010).

Just as there may be several theories of change in play for a given initiative, evaluators may create several logic models to lay bare the assumptions of the workings of the initiative. The logic model process is especially important for evaluators to examine the assumptions underlying a given program's success and failures considering that, to

date, there is little social science research to inform evaluators regarding the theoretical strengths and weaknesses of risk reduction initiatives. Nevertheless, Williams and Lindsey (2013) provide a theoretical critique of one of the most comprehensive initiatives (Saudi Arabia's) that may offer evaluators useful insights in identifying and critiquing initiatives' theories of change.

As touched upon, one of the primary uses of logic models is to represent graphically to stakeholders how a given initiative seems to function (including, perhaps, why it does not function as intended). The initial logic model(s) presented to stakeholders ought to be considered merely preliminary (Patton, 2008). The first drafts of logic models are vehicle for evaluators to come to shared understandings with stakeholders regarding the ends, ways, and means of the initiative: which also includes an understanding of the theory (or theories) of change and assumptions underlying the functioning of the initiative (Patton, 2008).

Choosing appropriate methods

After the evaluation's goals (i.e. stakeholders' informational needs) have been prioritized (through the aforementioned group processes), and after the stakeholders and evaluators have come to agreement regarding a reasonably accurate description of the program's ends, ways, and means, evaluators must bring their training in research methods to bear on developing appropriate methods to provide stakeholders with their sought-after answers. However, just as there are no "culture-free" evaluators, evaluators tend to be biased in favor of certain research designs over others (e.g. experimental vs. quasi-experimental vs. correlational; Lum & Yang, 2005). Indeed, a meta-analysis of criminal justice evaluations revealed that researchers' major field of study (e.g. psychology vs. sociology vs. criminology) was associated with their tendency to use experimental (vs. non-experimental) methods (Lum & Yang, 2005). Therefore, evaluators' methodological biases might disadvantage an evaluation if those predilections are not subordinated to stakeholders' information needs. Of course, a thorough discussion of research methods is well beyond the scope of any single article. Therefore, the present work continues by highlighting special methodological concerns that evaluators will likely need to consider in evaluating the effectiveness of a given risk reduction initiative.

Comparison groups

The effectiveness of an initiative is always relative to some type of benchmark (Centers for Disease Control and Prevention, 2011a). As mentioned, the dependent variable of post-detainment terrorism engagement could be measured in several ways including, for example, the change in the absolute value of post-detainment terrorism engagement incidents over time, or the change in percent of such incidents over time. Alternatively, even within a given initiative, a measure of the effectiveness of novel risk reduction method X could be compared to the effectiveness of standard method Y.

Although randomized experimental methods are widely considered the best for linking cause to effect in criminal justice evaluations (Lum & Yang, 2005), there is debate regarding their ethical use in such evaluations (Killias, 2006). On the one hand, public policy deserves to be based upon compelling evidence linking an initiative's activities to its impact. However, given the extraordinarily high stakes of failure

regarding terrorism risk reduction initiatives, legislators can scarcely justify the trial-and-error approach inherent to randomized experimentation (Killias, 2006).

To reconcile these opposing views of experimentation in criminal justice settings, it is important to note that, strictly speaking, randomized experiments are impossible across jurisdictions, because program types would be confounded with their jurisdictions. In short, one cannot randomly assign those who engage in post-detainment terrorism to the jurisdictions in which they engage in such activity. Specifically, there might be circumstances unique to a given jurisdiction that affect the likelihood of a given risk reduction initiative participant engaging in post-detainment terrorism. Therefore, cross-jurisdiction evaluations are – by their very nature – at best, suited to quasi-experimental designs. Consequently, best practice to test relative effectiveness across jurisdiction is to use those quasi-experimental designs strongest at inferring causation (e.g. propensity score matching, regression discontinuity, or time series designs; Shadish, Cook, & Campbell, 2001). Additionally, to enhance the confidence one may have in generalizability of such an evaluation's results, appropriate statistical tests are those based upon "nested"/random-effects models (Heck, Thomas, & Tabata, 2007; Mertler & Vannatta, 2010). In its report on improving evaluations of anticrime programs, The National Research Council recognized that well-conducted quasi-experimental methods are valid alternatives to randomized experiments (2005).

Therefore, randomized experiments are possible only within a given risk reduction initiative (e.g. to test effects of experimental component X vs. control component Y): which – regarding their ethical use – might be best suited to testing less socially critical outcome measures than post-detainment terrorism engagement rates. However, the case can be made that a theoretically informed novel intervention – hypothesized to be superior to previous interventions, but as-yet unproven – should be tested experimentally (i.e. the novel intervention vs. the status quo). In other words, it would be socially irresponsible to deploy an (albeit well intentioned), novel intervention on an entire population of program participants, given that the intervention's effectiveness has yet to be demonstrated. In such cases, the ethical decision would be to test the novel intervention with but a sample of program participants, then – if the intervention is deemed successful – deploy it with the remainder of program participants deemed suitable for it.

As mentioned, stakeholders should be the deciders regarding the outcomes to measure. Nevertheless, evaluators can – and should – guide decisions regarding how to orchestrate comparisons that are likely to offer compelling results both for stakeholders and (if results are to be made public) to the general public (Centers for Disease Control and Prevention, 2011a). As such, evaluators must concern themselves with drawing samples of observations (i.e. cases) that will support the types of causal inferences that stakeholders wish to make. Although a thorough discussion of sampling methodology is beyond the scope of the present work, it should be noted that research questions related to demographics of program participants (e.g. for what age group is the initiative most successful), stratified sampling techniques might need to be employed. Readers interested in such sampling methodologies are referred to Groves et al. (2009).

Eligibility criteria

Additionally, evaluators must collaborate with stakeholders to determine the selection eligibility criteria. For example, should the evaluation be based upon the initiative's

effectiveness for all prospective participants (i.e. all of the detainees who could elect to participate in the rehabilitation program), or should it evaluate the effectiveness only of those who opt into the rehabilitation program? In the former case, the program's effectiveness could be masked insofar as the number of successfully rehabilitated participants may be low relative to the overall detention population. In the case of the latter, selection bias is a concern, because those who opt into the program are (ostensibly) more motivated to rehabilitate themselves than the general detention population. Naturally, such motivation could be expected to influence their willingness to rehabilitate themselves, independent of the effectiveness of the initiative's rehabilitation methods per se (Wholey et al., 2010).

Also, it should be noted that a large power differential exists between program staff and program participants, which has implications for measuring the impact of a given risk reduction initiative. From the perspective of program staff, it can be assumed that their interests are aligned (understandably) with those of national security and demonstrating the effectiveness of their risk reduction initiative: perhaps more so than with rehabilitation of participants per se. Although those three objectives are by no means mutually exclusive (and, ideally, are one and the same), there may be a tendency for initiatives to permit only those prospective participants deemed at relatively low risk of committing post-detainment terrorism to participate. Therefore, a given program could demonstrate low post-detainment terrorism engagement rates, (perhaps) not because the program's interventions are especially effective, but that the participants already were at a low risk of reoffending. Although low risk offenders should be considered eligible for rehabilitation, even hard-core militants have been successfully rehabilitated (Rabasa et al., 2010). Therefore, it seems that initiatives need not preclude higher risk participants from participating in rehabilitation interventions, but that it should be done in ways that do not jeopardize the progress of lower risk participants, or that unduly compromise legitimate security concerns.

Likewise, an initiative's conservative tendencies may be manifest in unreasonably stringent risk assessments, such that few participants become eligible for parole. This would tend to have a similar impact on lowering post-detainment terrorism rates as rehabilitating only the lowest risk offenders: independent of whether the rehabilitation interventions per se are effective. Of course, it is the prerogative of a given nation to establish risk assessment criteria as it wishes. Nevertheless, this underscores the importance of instating valid and objective risk assessment protocols to prevent what could be considered conservative assessment "mission creep:" whereby, across time, risk assessments become more conservative due to assessors' subjective biases, or political pressures placed upon them, to parole fewer participants (Latessa & Lovins, 2010).

Considering that eligibility criteria, for both rehabilitation interventions and parole, differ across jurisdictions (Rabasa et al., 2010; Soufan et al., 2013), those factors must be accounted for, if one intends to make cross-national comparisons. In short, those programs that appear, at first glance, to have the lowest post-detainment terrorism engagement rates may be those that work with the least risky participants, and/or have the most restrictive parole standards. This suggests the value of implementing international standards for assessing participants' eligibility for rehabilitation interventions and parole. This is not to suggest that standardized eligibility criteria should displace those of a given nation, but that such standardized criteria should be measured in addition to those of interest to host nations. By doing so, those standardized eligibility

criteria could be variables accounted for in propensity score matching designs, to facilitate cross-jurisdiction impact comparisons.

Given the aforementioned power differential between program staff and program participants, another implication for measuring impact can be derived from participants' perspectives. Specifically, it can be assumed that they participate in hopes of promoting their release. Therefore, it can also be assumed that non-repentant participants would be motivated to fake their way through rehabilitation. This highlights the notion that participants' "deradicalization" is a less reliable outcome variable than their "disengagement" (Horgan & Braddock, 2010).

Another implication for impact measurement stems from participants who achieve parole, and who publicly comply with non-violent doctrines, but who continue secretly to harbor violence-justifying ideologies. Although such public compliance is recognized by some government to be a sufficient rehabilitation outcome (Barrett & Bokhari, 2009; Boucek, 2009), public compliance of a given belief tends to be less stable over time than private (i.e. genuine) acceptance of that belief (Horgan, 2009a; Kelman, 1958). Therefore, to account for post-detainment terrorism engagement that might occur relatively long after release, this highlights the importance of relatively long-term post-release monitoring and accompanying longitudinal data collection. Actuarial formulas, designed to assess the risk that participants will engage in post-detainment terrorism engagement at any given post-release point in time have yet to be developed (Williams & Lindsey, 2013). If no such timeline can be developed, it begs the question whether terrorism parolees should be subject to lifelong monitoring/data collection (Williams & Lindsey, 2013).

Archival and monitoring data

Alternatively, it should not be assumed that original data collection will be required to conduct the evaluation. Though perhaps unlikely, it could be that a given initiative's records contain sufficient data to answer stakeholders' research questions. However, such archival data are not without causes for concern; among them is that quality control of the data collection is thereby wrested from evaluators' hands (Groves et al., 2009). For example, one concern – both for archival data and original data collections – is the Hawthorne effect: to what extent have participants demonstrated improvements merely by virtue of their awareness that they were being observed (Roethlisberger & Dickson, 1939). Nevertheless, archival data could be useful, if not as a primary source of data, then as a supplementary source that could be used in triangulating the results from an original data collection.

Regardless of whether an evaluation is conducted with original and/or archival data, evaluators should recommend the ongoing collection, and periodic analysis, of program monitoring data. This recommended practice is based primarily upon the goal of empowering stakeholders to maintain quality control, improve service delivery, and fine tune program efficiency (Wholey et al., 2010). Additionally, it will allow stakeholders the opportunity to compare their programs' outcomes over time. That will allow a more detailed triangulation of findings than can be done within a single study: allowing stakeholders to have increased confidence in findings that are congruent across separate waves of data. Finally, in the present age of accountability, and finite financial resources for public programs, it behooves stakeholders to collect ongoing data about their programs, in order to document and justify their programs' ongoing worthiness for continued funding.

Process evaluation

Every impact evaluation should also be, in part, a process evaluation: designed to understand both what is being done by the initiative, and how faithfully/consistently its components are implemented. As mentioned, some data supporting this process evaluation will have been collected during the phase of "describing the program". However, ongoing collection of process-relevant data is warranted, if for no other reason than to confirm that the program continues to operate in similar ways during the evaluation proper as during the initial information-gathering phase. Program staff or participants may conduct themselves (perhaps unwittingly) differently, if they are aware that they are being studied (as mentioned, the so-called Hawthorne effect; Roethlisberger & Dickson, 1939).

Additionally, in discussing criminal justice evaluations, Lipsey et al. (2006) point out that, if an evaluation finds that the program has a beneficial impact, the process evaluation component affords description of the program's activities in sufficient detail, so that they may be modeled at other sites. Alternatively, if the evaluation reveals a null (or otherwise undesirable) impact, the process evaluation might shed light on the weak link(s) in the process that prevented the initiative from working as intended (Lipsey et al., 2006). Such documentation of an initiative's activities should be considered integral pieces of theory-driven crime reduction initiatives (Cooper & Worrall, 2012).

Dissemination and publicity

At first glance, publicity might not appear germane to the topic of choosing appropriate research methods, especially given that the evaluation results of a terrorism risk reduction initiative might not be made public. However, certainly the results will be disseminated to key stakeholders, and – whether public or private – it should go without saying that evaluations should strive to present results that are accurate, valid, and believable (Centers for Disease Control and Prevention, 2011a; Patton, 2008; Wholey et al., 2010). The combination of those three elements represents not only good science, but good showmanship: in the sense that results – even if accurate and valid – are relatively useless if they are not embraced by their audience (W.K. Kellogg Foundation, 1998). Consequently, in developing appropriate research methods, evaluators ought to consider the kinds of evidence that are likely to be compelling to stakeholders and – if applicable – to the broader public.

As mentioned, the primary research question likely to be addressed in evaluating a given risk reduction initiative is whether it is effective (i.e. effective per stakeholders' criteria). Empirically, such a question must be answered by quantitative methods, because effectiveness of something is always gauged in comparison to something else, and quantitative methods are the only way to make such comparisons in an objective way. Nevertheless, it would be a mistake for evaluators to assume that quantitative data will be most compelling to stakeholders and the broader public (Patton, 2008). To illustrate, if not to support, quantitative findings, qualitative data collection and analysis can be incorporated into the research design. For example, to highlight empirical findings regarding the strengths and weaknesses of a given program, excerpts of interviews with those who were successfully or unsuccessfully rehabilitated could be used to exemplify those strengths and weaknesses. Such narratives are likely to have staying power in the minds of audience members, which can make impersonal numerical

data come to life in the words of those about whom the numerical data refer (Patton, 2008). Furthermore, public declarations from former extremists (especially from leaders), denouncing their formerly violent ways, could be counted among an evaluation's outcome measures.

To be sure, the potential constructive influence on risk reduction and rehabilitation programs generated by the contributions offered by former extremists remains unclear. This was a topic specifically examined by The Soufan Group in the course of a multi-year, global study of programs designed to counter violent extremism (2013). They found that among European-based programs, one of the "most consistently contentious issues is [the] recognition, resourcing, and involvement of former terrorists and prisoners, particularly in counter-narrative efforts" (Soufan et al., 2013). At the same time, a highly public role has been played by former extremists within (or parallel to) the government-sanctioned risk reduction programs in Southeast Asia (SEA), most notably Singapore, Indonesia, and Malaysia. A case in point involves Nasir Abas, a former leader of the al Qaeda-affiliated extremist group, Jemaah Islamiya. Following years of involvement in violent attacks carried out in the SEA region, Abas became a relatively well-known figure for speaking out against extremism as a viable means for achieving sociopolitical ends. This is the primary theme of his book, *Inside Jemaah Islamiyah* (Abas, 2011), where his target audience has specifically included individuals either already involved in extremist activities or contemplating such a path. While Abas has largely operated independently from government programs, he has been sought by regional government officials to advise and support ongoing risk reduction programs.

There are seemingly endless varieties of qualitative data that can be collected, analyzed, and reported in a program evaluation. Although a discussion of qualitative methodologies is beyond the scope of the present work, interested readers are referred to Patton (2008) and Wholey et al. (2010). Nevertheless, the trend in evaluation is toward a synthesis of mixed-method research designs, and evaluators would be well advised to consider how qualitative methods could play a part in their evaluation designs of risk reduction initiatives (Bouffard & Little, 2004; Droitcour, 1997; Wholey et al., 2010).

Implementation

Training data collection staff

Whether or not data will be collected exclusively by the evaluators, it is advisable for evaluators to conduct adequate training of data collection staff (Centers for Disease Control and Prevention, 2011b). Although this requirement is, perhaps, a smaller task if the evaluators are the sole data collectors, it is important nevertheless to promote – among other salubrious effects – uniformity of collection methods among the evaluators.

In addition to uniformity of data collection methods, there are at least two other beneficial effects of proper data collection training. First, assuming that training includes a component regarding data storage, it is reasonable to assume that such training will reduce the likelihood of data loss or breaches in data security (Centers for Disease Control and Prevention, 2011b). Of course, in the context of evaluating risk reduction programs, the importance of data security (which might be a matter of national security) can scarcely be overstated. Second, adequate training in collection methods – ideally,

methods that are culturally aware – can safeguard against social missteps of several kinds, ranging from breaches of etiquette to breaches of ethics (Centers for Disease Control and Prevention, 2011b).

Regarding ethics, to protect the confidentiality of both program staff and program participants, it is important that data are collected and reported in accord with standards at least as stringent as those asserted by the US Office for Human Research Protections (Code of Federal Regulations, 2011). At the data collection phase, this includes such practices as preserving participants' anonymity by de-identifying their data. At the data reporting phase, this includes practices such as reporting findings only in the aggregate.

Additionally, it is conceivable that ethical dilemmas could arise if proper disclosures are not made to participants in advance of data collection. For example, during the course of an evaluation, if evaluators acquire information suggestive of pending acts of terrorism, they might be caught between the duty to protect participants' anonymity and the duty to counter violent extremism. To avoid such dilemmas, evaluators should include disclosures in their informed consent materials that make clear to participants that evaluators would be required to disclose to the proper authorities any information they reasonably believe to be related to pending illegal activities.[2]

Pilot testing

Evaluations should not be implemented without pilot testing the data collection instruments (e.g. surveys; Wholey et al., 2010). This may be self-evident in the case of using newly constructed instruments in which their validity has yet to be established, but it is equally important when using previously validated instruments in cultures other than those in which they were developed (Baugh & Guion, 2006; Centers for Disease Control and Prevention, 2011b). The purpose of such pilot testing is to assess the cross-cultural validity of those measures (Frierson et al., 2010).

For example, one of the primary cross-cultural challenges when using survey methods is to assess whether respondents in other cultures interpret the meaning of survey items similarly as respondents in the country in which the survey was developed (Baugh & Guion, 2006; Groves et al., 2009). This is not merely a concern regarding whether questions have been properly translated into other languages, but whether the meaning/intent of the question is judged similarly by respondents cross-culturally (Baugh & Guion, 2006; Groves et al., 2009). To assess how a given set of survey questions is interpreted by respondents from different cultures, evaluators would be well advised to conduct focus groups with samples of respondents (Baugh & Guion, 2006; Groves et al., 2009) to develop any survey that will be launched in pilot testing.

Analysis

An analysis plan should be developed prior to data collection, though it may be revised based upon unexpected results from the pilot-testing phase (United States Government Accountability Office, 2012). The intent of the analysis plan should focus on providing stakeholders with information according to their needs identified during the logic-modeling phase of the evaluation. As mentioned, qualitative analyses are likely to be an important part of a mixed-method research design. Regardless of design, evaluators should be cautious to avoid the pitfall of allotting too little time for data analysis relative to the time allotted for data collection (Wholey et al., 2010). Additionally, it bears

repeating that it is important to analyze not simply "does the program work", but to what extent does it work as a function of participants' individual difference variables (e.g. type of offense, motive(s) for the offense, age, offense history; Winokur, 2002).

Making recommendations

In keeping with the spirit of utilization-focused program evaluation, recommendations should be made based upon stakeholders' stated objectives (McNamara, 2006; Patton, 2008). Additionally, it is considered a professional courtesy to permit key stakeholders the opportunity to review preliminary findings prior to formal dissemination of the evaluation (Wholey et al., 2010). Although the idea of briefing stakeholders of the findings, prior to their official dissemination, may appear to invite stakeholders' influence to sway (or otherwise conceal) perhaps unflattering findings, that is neither the intent, nor a scientifically valid option, regarding this procedure. Instead, this procedure permits stakeholders an opportunity to alert evaluators to details that might have been overlooked or misunderstood by the evaluators (Wholey et al., 2010). Of course, even after publication of the findings – regardless of the integrity with which they were reported – findings can become politicized or otherwise spun to make claims other than those clearly intended, or supported, by the evaluation. Of course, this is beyond evaluators' control.

In accord with making recommendations based upon stakeholders' objectives, Horgan and Braddock (2010) recommend several practices based upon multi-attribute utility technology (popularized by Edwards and Newman (1982)). Among those recommendations is for evaluators to recognize that although a given initiative may have several objectives, those objectives are seldom of equal importance (Horgan & Braddock, 2010). Consequently, evaluators ought to organize and frame their recommendations with respect to stakeholders' hierarchy of objectives.

Additionally, recommendations should be made that can directly inform policy decisions about the initiative (Horgan & Braddock, 2010; Lipsey et al., 2006).

Communicating findings/recommendations

It is virtually impossible to make evaluation results too user-friendly, and it is important to recognize that high-quality evaluation methods are wasted if they are not readily understandable and compelling to stakeholders (W.K. Kellogg Foundation, 1998). For example, key points should be attractively formatted to stand out (e.g. via text boxes), and prose should be relatively jargon-free (W.K. Kellogg Foundation, 1998). Also, as mentioned, qualitative data (e.g. narratives) are likely to be a compelling way to highlight the stories told by the quantitative data. The strategy that should permeate the communication of findings and recommendations should be to link results back to the evaluation plan (Centers for Disease Control and Prevention, 2011b).

In presenting the relationship of the evaluation's results to the evaluation plan, it is not only permissible, but the hallmark of an exceptionally well-considered evaluation, if the results beg more questions than they answer (Centers for Disease Control and Prevention, 2011b). Such questions might highlight areas for improvement not only in the initiative, but in the evaluation process itself (Centers for Disease Control and Prevention, 2011b). Suggested improvements to the evaluation should not be considered stains on the current evaluation, but part of an iterative process of evaluation that should be

integrated into any organization devoted to self-improvement (United States Government Accountability Office, 2012).

Future directions

Aside from such questions as which nation has the most effective risk reduction initiative, and why, a seemingly basic – though as-yet unanswered – question is: are risk reduction initiatives more or less effective than no risk reduction initiative at all (Morris et al., 2010). Moreover, it is not yet clear whether risk reduction through a formal terrorism risk reduction initiative is any more effective than more conventional counterterrorism operations (Soufan et al., 2013). Post-detainment terrorism engagement rates have yet to be compared between nations that have vs. have not implemented risk reduction initiatives (Morris et al., 2010). In short, the perhaps iatrogenic effects of risk reduction initiatives have yet to be assessed.

Standard outcome measures

Another direction for the future of terrorism risk reduction program evaluation pertains to making consistent the outcome variables measured across evaluations. As mentioned, the appropriate outcome variables for the evaluation of terrorism risk reduction initiatives are by no means standardized (Horgan, 2009a). This is justified insofar as the outcome variables of interest in any given utilization-focused evaluation should be determined by the stakeholders (Patton, 2008). Nevertheless, it would serve both of the security interests of the nations who conduct terrorism risk reduction initiatives, and the research interests of the evaluation community, if a set of standard outcome variables (i.e. their definitions/operationalizations) could be derived. Additionally, in discussing the need for evaluation of criminal justice programs, Wilson (2006) points out that generalizability of findings is established, in part, through comparisons across locations.

   This is not to suggest that standardized outcome variables should supplant those of interest to any given stakeholder, but – as with regard to standardized criteria for assessing participants' eligibility for rehabilitation interventions and parole – that such standardized outcome variables should be measured in addition to those of interest to the stakeholders. The purpose of such standardization would permit cross-national, cross-cultural meta-analyses of risk reduction initiatives (Williams & Lindsey, 2013). By affording the opportunity for meta-analyses, the programs of various locations might be evaluated for their relative effectiveness.

   However, it is important to note that there are limits – the extent to which are currently unknown – regarding cross-national, cross-cultural comparisons of terrorism risk reduction initiatives. For example, some nations are markedly different from others with respect to their legal landscapes. Specifically, Saudi Arabia's terrorism risk reduction initiative dictates, to some extent, which participants are forbidden to fraternize following their release from that program (Boucek, 2009). Though such restrictions on participants' post-release civil liberties may contribute to the effectiveness of the Saudi program, restriction would be legally unacceptable in some nations. Therefore, any cross-jurisdiction comparisons of terrorism risk reduction initiatives must deeply explore, and make explicit, socio-cultural-legal differences between the jurisdictions before making claims (if any) regarding the generalizability of the findings.

Essential components

Additionally, standardization of outcome variables could enable analysis of the degree of the effectiveness of program components by comparing programs that have contrasting components (Williams & Lindsey, 2013). Although replete with methodological challenges, a crude estimate of the additive or multiplicative effects of program components could be derived from cross-national comparisons of initiatives (Williams & Lindsey, 2013). Specifically, if the outcome variables of post-detainment terrorism engagement are measured uniformly across locations, then similar components of different risk reduction initiatives could be compared for their relative effectiveness (Romaniuk & Chowdhury-Fink, 2012; Williams & Lindsey, 2013). Additionally, what might emerge is an understanding of the essential kernels of a program: the essential components of a successful terrorism risk reduction initiative that must be implemented with high fidelity lest the program suffer an exponential decrease in effectiveness (Patton, 2008; Williams & Lindsey, 2013).

An alternative approach to identifying essential kernels of a program, suggested by the literature on evaluating criminal justice programs, is – within a given jurisdiction – to begin paring away those components of an intervention thought to be relatively unimportant contributors to the program's outcomes (Wilson, 2006). Then, one may note whether post-detainment terrorism engagement rates remain the same (Wilson, 2006). However, that logic implies that the process should continue until there is a rise in such engagement: which is likely to be socially unacceptable, in the case of terrorism engagement. Therefore, it seems this approach is less applicable to assessing the essential components of terrorism risk reduction initiatives than to assessing other aspects of such initiatives (e.g. their efficiency, or levels of participant/staff satisfaction).

Stakeholder training

It should be noted that while the role of the stakeholder is of vital importance – as is the feedback received from these interested parties – it must not be assumed that these individuals and agencies have a sufficient grounding in the research, the history, or the challenges relating to establishing, operating, and/or evaluating any manner of disengagement program. The simple reality is that one cannot assume that with authority comes the requisite expertise. This highlights the need for evaluators to provide a sufficient level of background information (perhaps even training) for stakeholders in a manner that would better equip them to understand more precisely what they might expect from these programs, while empowering them to provide more meaningful feedback over time. To that end, the appendices of this work (containing both a process checklist for conducting an impact analysis of terrorism risk reduction initiatives, and a self-assessment tool for evaluators of such initiatives; see Appendices 1 and 2, respectively) could be adapted for dissemination to stakeholders. Doing so could benefit them by giving stakeholders an overview of the evaluation process, and by helping them to assess their knowledge and understanding of specific areas of program evaluation. As a thumbnail sketch of the process checklist, Table 1 displays its major headings. It should be noted that this table is not intended as a substitute for the complete checklist, and interested readers are encouraged to see the appendices included in the online version of this article, or to obtain them by contacting the first author.

Table 1.   Process checklist for an impact analysis of terrorism risk reduction initiatives: major headings.

Prior to First Group Meetings with Stakeholders
Identify lead/key stakeholders.
Select the evaluation personnel.
First Group Meeting with Stakeholders
Request a briefing from the initiative's staff regarding the initiative's basic mission and
    activities.
Facilitate stakeholders' consensus regarding a statement of the problem(s).
Facilitate stakeholders' consensus regarding goals for the evaluation.
Facilitate stakeholders' prioritization of their goals/informational needs.
Deliberate about primary outcome/indicator/dependent variables to measure.
Following the First Group Meeting with Stakeholders
Continue to gather information about the initiative.
Compose first draft of the initiative's logic model(s).
Reconvene the Stakeholders
Walk stakeholders through a presentation of the logic model.
Design the evaluation
Design appropriate research methods.
Train data collection personnel.
Pilot test measures and procedures.
Commence substantive evaluation
Conduct measurements.
Data analysis.
Communicate findings & make recommendations
Preliminary presentation.
Formal Presentation.

Conclusion

Research suggests that the process of evaluation itself, regardless of the substantive findings, will tend to improve a given program (unless the evaluation is conducted in a way that overburdens program staff; Patton, 2008). Specifically, organizations tend to benefit merely by engaging in the activities endemic to most evaluations (e.g. coming to terms about program goals, deciding upon appropriate outcomes to measure, and collaborating in the development of logic models; Patton, 2008). By engaging in such activities, it is thought that organizations can gain heightened self-awareness, collective vision, and enhanced group cohesion (Patton, 2008; Wholey et al., 2010). In the extreme, evaluations can still be worthwhile whether or not the process continues to the data collection, analysis, or reporting phases (Patton, 2008). Therefore, though the present work highlights many challenges to the evaluation of terrorism risk reduction initiatives, those challenges should by no means preempt evaluation in this domain. To the contrary, even evaluations modest in scope, or those that reach highly tentative findings, seem likely to have beneficial effects upon the initiatives fortunate enough to undertake evaluation (Patton, 2008). In doing so, the primary objective of utilization-focused evaluation will have been met nonetheless: to serve the stakeholders.

Acknowledgments

Notes

1. The term "deradicalization" has long been in use to describe various efforts to diminish, deter, and/or reverse the process(es) an individual might experience on the path toward violent extremism. A problem naturally arises with the root word, "radical," as this is not an inherently pejorative term (to be sure, radical ideas have always lingered at the leading edge of science, technology, and philosophy.) We address this terminology later in this paper.
2. Both the role and the professional standing of any member of the evaluation team will be a factor in adjudicating disclosure-related issues. For example, a licensed psychologist might encounter professional ethics concerns that must be accounted for in advance. The evaluation effort thus would be well served by the establishment of clear guidelines with respect to how the leadership (i.e., the government or non-governmental organization sponsoring the program and/or its evaluation) views the balance between security interests and ethical standards.
3. Though ideal if all members possess the following qualities, it may be sufficient to have these qualities represented by the team as a whole: provided that the team has excellent communication and it collaboratively defers to those with the requisite expertise as demanded by the exigencies of the situation.
4. Any request for such a briefing should be made well in advance of the first stakeholder meeting, to permit staff ample time to prepare the briefing.
5. Both disengagement and deradicalization can be measured. One need not opt for one instead of the other.
6. Consider measuring all of the above, if it is not cost/labor prohibitive. Doing so might serve as-yet unidentified stakeholder needs. Additionally, over time it would serve the interest of cross-national comparisons of effective counter-extremism programming. Given the transnational nature of much terrorism, such comparisons can be considered in the interests of every stakeholder worldwide. However, stakeholders' explicitly stated needs remain paramount. Therefore, they retain the prerogative to decide what outcomes to measure.
7. It may be that all of these can be measured. Though, measuring change over time requires a longitudinal design that may be time/cost prohibitive.
8. It is possible that an initiative's theory of change might be a) relatively incoherent, and/or b) poorly correspond to initiative activities.

Notes on contributors

Michael J. Williams is a National Science Foundation graduate research fellow, and a doctoral candidate at the University of Nevada, Reno's interdisciplinary social psychology program. His research focuses on countering violent extremism, intelligence interrogations (with current funding from the High-Value Detainee Interrogation Group/the FBI), and program evaluation. Currently, he is co-leading evaluations of two programs designed to counter violent extremism: one in collaboration with the LAPD, and one of a community-based, Muslim-led, grassroots CVE initiative (evaluation funded by the National Institute of Justice).

Steven M. Kleinman is the Director for Strategic Research for The Soufan Group. A career intelligence officer with more than 28 years of operational and leadership experience, he graduated from U.C., Davis, with a B.A. in psychology; the National University, with an M.S. in forensic sciences; and the National Defense Intelligence College, with an M.S. in strategic intelligence.

References

Abas, N. (2011). *Inside Jemaah Islamiyah: A former member's true story*. Jakarta: Graha Grafindo.
Barrett, R., & Bokhari, L. (2009). Deradicalization and rehabilitation programmes targeting religious terrorists and extremists in the Muslim world: An overview. In T. Bjorjo & J. Horgan (Eds.), *Leaving terrorism behind* (pp. 170–180). New York, NY: Routledge.

Baugh, E., & Guion, L. A. (2006). Using culturally sensitive methodologies when researching diverse cultures. *Journal of Multi-disciplinary Evaluation, 4*. Retrieved from http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/74/88

Boucek, C. (2009). Extremist re-education and rehabilitation in Saudi Arabia. In T. Bjorjo & J. Horgan (Eds.), *Leaving terrorism behind* (pp. 212–223). New York, NY: Routledge.

Bouffard, S., & Little, P. (2004). Detangling data collection: Methods for gathering data. *Harvard Family Research Project*. Retrieved from http://www.hfrp.org/publications-resources/browse-our-publications/detangling-data-collection-methods-for-gathering-data

Brookings Institution. (1998). *Learning what works: Evaluating complex social interventions.* Retrieved from http://www.brookings.edu/~/media/research/files/articles/1998/3/01politics%20mann/learningwhatworks

Centers for Disease Control and Prevention. (2011a). Developing an effective evaluation plan: Setting the course for effective program evaluation. Retrieved from http://www.cdc.gov/obesity/downloads/CDC-Evaluation-Workbook-508.pdf

Centers for Disease Control and Prevention. (2011b). *Implementing evaluations: Learning and growing through evaluation module 2*. Retrieved from http://www.cdc.gov/asthma/program_eval/LG-Mod2_DraftFinal_Allsections_Wordaym.pdf

Code of Federal Regulations. (2011). *Title 45 public welfare department of health and human services part 46 protection of human subjects.* Washington, DC: U.S. Government printing office. Retrieved from http://www.hhs.gov/ohrp/policy/ohrpregulations.pdf

Cooper, J. A., & Worrall, J. L. (2012). Theorizing criminal justice evaluation and research. *Criminal Justice Review, 37*(3), 384–397. doi:10.1177/0734016812442940

Droitcour, J. A. (1997). Cross design synthesis: Concept and application. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: SAGE.

Edwards, W., & Newman, J. R. (1982). *Multiattribute evaluation*. Beverly Hills, CA: SAGE.

Frierson, H. T., Hood, S., & Hughes, G. B. (2010). A guide to conducting culturally-responsive evaluations. In J. Frechtling (Ed.), *The 2010 user-friendly handbook for project evaluation* (pp. 75–96). Arlington, VA: National Science Foundation.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). New Jersey: Wiley & Sons.

Hannah, G., Clutterbuck, L., & Rubin, J. (2008). *Radicalization or rehabilitation: Understanding the challenge of extremist and radicalized prisoners*. Santa Monica, CA: RAND Corp.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2007). *Multilevel and longitudinal modeling with IBM SPSS*. New York: Routledge.

Horgan, J. (2009a). Individual disengagement: A psychological analysis. In T. Bjorjo & J. Horgan (Eds.), *Leaving terrorism behind* (pp. 170–180). New York, NY: Routledge.

Horgan, J. (2009b). *Walking away from terrorism*. New York, NY: Routledge.

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. doi:10.1080/09546551003594748

Horgan, J., & Taylor, M. (2011). Disengagement, de-radicalization, and the arc of terrorism: Future directions for research. In R. Coolsaet (Ed.), *Jihadi terrorism and the radicalisation challenge: European and American experiences*. Farnham, Surrey: Ashgate.

International Peace Institute. (2010). *A new approach? Deradicalization programs and counterterrorism*. Retrieved from http://www.ipinst.org/media/pdf/publications/a_new_approach_epub.pdf

Kelman, H. (1958). Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution, 1*, 51–60.

Killias, M. (2006). Improving impact evaluations through randomised experiments: The challenge of the National Research Council report for European criminology. *Journal of Experimental Criminology, 2*(3), 375–391. doi:10.1007/s11292–006-9016-7

Latessa, E. J., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims & Offenders, 5*(3), 203–219. doi:10.1080/15564886.2010.485900

Lipsey, M., Petrie, C., Weisburd, D., & Gottfredson, D. (2006). Improving evaluation of anti-crime programs: Summary of a National Research Council report. *Journal of Experimental Criminology, 2*(3), 271–307. doi:10.1007/s11292-006-9009-6

Lowenkamp, C. T., & Latessa, E. J. (2004). Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. In *Topics in community corrections* (pp. 3–8). Washington, DC: Department of Justice.

Lum, C., & Yang, S.-M. (2005). Why do evaluation researchers in crime and justice choose non-experimental methods? *Journal of Experimental Criminology, 1*(2), 191–213. doi:10.1007/s11292-005-1619-x

McNamara, C. (2006). *Field guide to nonprofit program design, marketing and evaluation* (4th ed.). Minneapolis, MN: Authenticity Consulting, LLC. Retrieved from http://www.authenticityconsulting.com/pubs/PG_gdes/PG-toc.pdf

Mertler, C. A., & Vannatta, R. A. (2010). *Advanced and multivariate statistical methods: Practical application and interpretation* (4th ed.). Los Angeles, CA: Pyrczak.

Morris, M., Eberhard, F., Rivera, J., & Watsula, M. (2010). *Deradicalization: A review of the literature with comparison to findings in the literatures on deganging and deprogramming*. Institute for Homeland Security Solutions. Retrieved from http://sites.duke.edu/ihss/files/2011/12/Morris_Research_Brief_Final.pdf

Mount, M. (2010). *Report: 20 percent of released detainees returning to terrorism. CNN*. Retrieved from http://www.cnn.com/2010/POLITICS/01/11/detainees.terror.retu

National Research Council of the National Academies. (2005). *Improving evaluation of anticrime programs*. Washington, DC: The National Academies Press.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: SAGE.

Rabasa, A., Pettyjohn, S. L., Ghez, J. J., & Boucek, C. (2010). *Deradicalizing Islamist extremists*. Santa Monica, CA: Rand Corp. Retrieved from http://www.rc.rand.org/pubs/monographs/MG1053/

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.

Romaniuk, P., & Chowdhury-Fink, N. (2012). *From input to impact: Evaluating terrorism prevention programs*. Center on Global Counterterrorism Cooperation. Retrieved from http://www.globalct.org/wp-content/uploads/2012/10/CGCC_EvaluatingTerrorismPrevention.pdf

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin.

Shane, S. (2009, January 23). Obama orders secret prisons and detention camps closed. *The New York Times*. Retrieved from http://www.nytimes.com/2009/01/23/us/politics/23GITMOCND.html

Soufan, A., Fallon, M., Freedman, D., Borum, R., Horgan, J., Gelles, M., . . . McManus, B. (2013). *The QIASS countering violent extremism (CVE) risk reduction project*. Qatar International Academy for Security Studies.

United States Government Accountability Office. (2012). *Applied research and methods: Designing evaluations*. (Report GAO-12-208G). Retrieved from http://www.ignet.gov/goa_588146.pdf

United Way. (2008). *Outcome measurement resource network*. Retrieved from http://www.nationalserviceresources.org/filemanager/download/ProgramMgmt/Outcome_Measurement_Showing_Results_Nonprofit_Sector.pdf

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation* (3rd ed.). San Francisco, CA: Josey-Bass.

Williams, M. J., & Lindsey, S. C. (2013). A social psychological critique of the Saudi terrorism risk reduction initiative. *Psychology, Crime & Law, 19*(3), 1–17. doi:10.1080/1068316X.2012.749474

Wilson, J. Q. (1993). *Performance measures for the criminal justice system*. Washington, DC: Bureau of Justice Statistics, Article prepared for the U.S. Department of Justice, Bureau of Justice Assistance, pp. 157–167.

Wilson, J. Q. (2006). The need for evaluation research. *Journal of Experimental Criminology, 2*(3), 321–328. doi:10.1007/s11292-006-9011-z

Winokur, K. P. (2002). What works in juvenile justice outcome measurement: A comparison of predicted success to observed performance. *Federal Probation, 66*(2), 50–55.

W.K. Kellogg Foundation. (1998). *W.K. Kellogg Foundation Evaluation Handbook*. Retrieved from http://www.wkkf.org/~/media/62EF77BD5792454B807085B1AD044FE7/EvaluationHandbook.pdf

Appendix 1

Process checklist for an impact analysis of terrorism risk reduction initiatives

Prior to First Group Meetings with Stakeholders

☐ **Brainstorm/ identify all affected stakeholders**, according to the evaluation's purpose, those who will:
>   Primarily use the evaluation's results
>   Carry out the initiative
>   Be directly or indirectly affected by initiative activities

☐ Organize this list in a hierarchical network to identify lead/key stakeholders.

☐ **Identify lead/key stakeholders**: Individuals at the top of the political and programmatic organizations.
>   ☐ Engage in dialogs with these individuals to assess their willingness and ability to pursue an ongoing collaboration in the evaluation process. Specifically, assess whether they are able to participate in regular meetings with the other stakeholders, and that evaluation staff will have access to information listed below under "Continue to gather programmatic information"
>   ☐ If commitment at the top is lacking, the evaluation may have to be deferred or abandoned.

☐ **Select the evaluation personnel**[3]
>   *Core competencies*
>   ☐   *Cultural knowledge: pertinent to norms of all stakeholders' cultures.*
>   ☐   *Quantitative and Qualitative data collection/analysis skills, including . . .*
>   ☐   *Observational methods*
>   ☐   *Focus group facilitation / Small group processes*
>   ☐   *Interviewing*
>   ☐   *Survey design*
>   Perhaps, also including . . .
>   ☐   *Ethnographic methods*
>   ☐   *Archival data methods*
>   ☐   *Psychophysiological methods*
>   Presentation skills
>   ☐   *Public speaking*
>   ☐   *Fluency with presentation software & desktop publishing*

☐ Deliberate with evaluation personnel to identify sources of personal socio-cultural biases.
>   ☐ Deliberate with evaluation personnel on how to minimize the influence of these biases.

First Group Meeting with Stakeholders

☐ **Request a briefing from the initiative's staff regarding the initiative's basic mission and activities.**[4]

☐ **Facilitate stakeholders' consensus regarding a statement of the problem(s).**

☐ **Facilitate stakeholders' consensus regarding goals for the evaluation**.

☐ Poll stakeholders for informational needs
>   *What decisions do stakeholders need to make?*
>   *Who will make those decisions?*
>   *How will the decisions be made (i.e., by*
>   *executive decision or through committee vote;*
>   *how will differences in opinion be settled;,*
>   *how will the decisions be disseminated)?*

When do those decisions need to be made?
What information do they need to guide those decisions?
How are decisions made within stakeholders' respective organizations?
In addition to decision-making, to what uses will they put that information?

How will stakeholders know if the initiative is running as planned?

☐ Consider having stakeholders deliberate in small groups about their information needs.
☐ What metrics will be used to measure progress toward stated goals?

☐ **Facilitate stakeholders' prioritization of their goals/informational needs.**
    ☐ Consider having stakeholders vote (individually, perhaps by secret ballot) for their most important informational needs (e.g., top 2-3 items).
☐ **Deliberate about primary outcome/indicator/dependent variables to measure** (i.e., criteria that are indicative of the initiative's success, per its stated objectives).
☐ Consider to what extent de-radicalization vs. disengagement is the initiative's objective[5]
☐ Consider, with stakeholders, a variety of indicators that could be measured, pertinent to their stated objectives: *e.g. to what extent the initiative's participant:*
    Commit politically/ideologically motivated violence
    Plan politically/ideologically motivated violence
    Finance terrorism
    Organize terrorist groups
    Provide support to terrorist acts (e.g., arranging for transportation, communication, and safe haven)
    Collaborate in production/dissemination of terrorist propaganda
    Inciting others to terrorism[6]
    ☐ Consider, with stakeholders, what type of information regarding the above would be most useful
    e.g., Frequency counts of the above events
        Percent of initiative participants who engage in the above events
        Percent change over time (quarterly, yearly, etc.) of the above events.[7]
  ☐ Consider secondary outcomes/indicators/dependent variables to measure
    Psychological factors (e.g., Attitude change) Post-release social adjustment

Following the First Group Meeting with Stakeholders

☐ **Continue to gather information about the initiative**
☐ Independent literature review: both academic, and materials in use by the initiative
☐ First-hand research (e.g., observations, interviews), to gain a sound working knowledge of the initiative's . . .
    ☐ *Staff: both those who are/were satisfied, and those were dissatisfied/quit*
    ☐ *Participants: both those who are/were successful, and those who failed/quit*
    ☐ *Activities*
    ☐ *Technologies (including assessment protocols)*

☐ **Compose first draft of the initiative's logic model(s).**
    ☐ Identify initiative's theory (or theories) of change.[8]

Reconvene the Stakeholders

☐ **Walk stakeholders through a presentation of the logic model**
    ☐ Gain stakeholders' feedback regarding the logic model's assumptions, theories, inaccuracies, and overlooked components.
    ☐ Revise logic model, based upon shared understandings of the initiatives ends (outputs), ways (activities), means (inputs), theory(ies) of change, and associated assumptions. If necessary, consider multiply logic models to reflect different phases of a program over time.

☐ Discuss feasible timeline for evaluation.
☐ Do not underestimate time needed for data analysis
☐ Incorporate time for revision/re-piloting of data collection instruments.

Design the evaluation

☐ **Design appropriate research methods**
   Must be steadfastly oriented toward answering stakeholders' prioritized information needs.
   ☐ Consider appropriate comparison group(s).
      ☐ *Consider if/how quasi-experimental designs (e.g., propensity score matching, or regression discontinuity designs) might be employed to answer questions regarding causality.*
   ☐ Consider appropriate sampling methodology.
      ☐ *Consider stratified sampling, if informational needs involve demographic sub-populations.*
      ☐ *Consider, in collaboration with stakeholders, sample eligibility/selection criteria.*
   ☐ Consider a mixed method design approach, including …
   ☐ *Consider how qualitative data may inform stakeholders.*
   ☐ *Consider how both longitudinal and cross-sectional data may inform stakeholders.*
   ☐ *Consider how archival data might may inform stakeholders.*
   ☐ *Consider how to integrate a process evaluation into the design.*
   ☐ Consider what kinds of information are likely to be most compelling (i.e., considered most accurate, valid, and believable/vivid) to …
      Stakeholders
      The general public (if applicable)
   ☐ Compose data analysis plan
   ☐ Appraise stakeholders (as appropriate, both methodologically and politically) of the research design.

☐ **Train data collection personnel**
☐ Stress uniformity in collection procedures.
☐ Address proper data storage. (Build in frequent, secure data backups.) Attune collection procedures to cultural norms.
☐ Apprise stakeholders (as appropriate, both methodologically and politically) of progress.

☐ **Pilot test measures and procedures**
   For novel surveys and novel self-report measures:
   ☐ Consider developing measures with a focus group of participants from the target population (to improve cross-cultural validity of those questions).
   For all other surveys and self-report measures:
      ☐ Translate and back-translate measures to gauge whether questions are worded appropriately.
      ☐ Incorporate methods (e.g., interviews, focus groups), to gauge whether survey questions are understood as intended by respondents.
      ☐ If applicable, analyze instruments' measurement properties *(e.g., reliability analysis, factor analysis).*
      ☐ Revise, and re-pilot measures, as necessary, until …
         a) Measured constructs are understood equivalently cross-culturally
         b) Conventionally acceptable levels of reliability are achieved, and
         c) The theoretically expected number, and kind, of factors are captured by the instruments.
☐ Appraise stakeholders (as appropriate, both methodologically and politically) of progress.

Commence substantive evaluation

☐ **Conduct measurements**
   ☐ Periodically review data (prior to final analysis), to identify systematic completion errors or systematically missing data.

☐ If methodologically appropriate, modify procedure to improve data collection methods.
☐ Periodically consult with data collection staff to maintain high/uniform collection standards.
☐ Periodically apprise stakeholders (as appropriate, both methodologically and politically) of progress.

☐ **Data analysis**
☐ Integrate (triangulate) data from all forms of data collection.
☐ Make clear both the consistencies and contradictions in the findings.
☐ Make clear (at a minimum) statistical confidence levels, confidence intervals, and effect sizes.

Communicate findings & make recommendations

Preliminary presentation
☐ Provide stakeholders with a preliminary/courtesy review of the findings.
☐ Make all materials/presentations as user-friendly, and visually attractive, as evaluators' imagination and resources permit.
☐ Avoid jargon.
☐ Focus presentation on the specific needs/interests of the target audience (i.e., a specific stakeholder or group of stakeholders)
☐ Request stakeholders' feedback regarding overlooked/misunderstood details. Integrate that feedback into formal report of the findings.

Formal Presentation
☐ Present findings in multiple formats: written, verbal (in-person/presentation), and/or multi-media (e.g., film, photographic, narrative) formats.
☐ Consistently, frequently link findings back to the evaluation plan (i.e., to stakeholders' previously-stated informational needs).
☐ Pose new questions suggested by the evaluation.
☐ *Pertaining to the initiative.*
☐ *Pertaining to the evaluation itself.*
☐ Organize recommendations based upon stakeholders' stated objectives and programmatic priorities.

Appendix 2

Self-assessment for evaluations of terrorism risk reduction initiatives

Increasing Your Evaluation Capacity: How Do You Rate Yourself?

This self-assessment can be used to understand your knowledge level of specific areas of program evaluation. Individual responses of this self-assessment are not meant to be totaled; rather this should be used as a checklist to determine what your individual level of competency is in relation to a specific evaluation content area. In general, scores of "0" indicate a pre-novice level of competency for that topic; scores of "1" indicate a novice level of competency; scores of "2", an advanced beginner level; and scores of "3", a practitioner level. Demonstrated outcomes for each of these competency levels, as well as educational resources for each of the evaluation topic areas contained in this assessment can be found following this self-assessment.

How do you rate yourself?
0- I know nothing about this
1- I understand the basic concept - **(novice level)**
2- I can implement this concept with assistance - **(advanced beginner level)**
3- I can implement this concept independently and/or teach it to others - **(practitioner level)**

I. Program Planning for Program Evaluation	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Know the terms or components of a logic model	0	1	2	3
b. Develop a logic model or other theory of change for program planning	0	1	2	3
c. Create evaluations that match a program logic model or program theory	0	1	2	3

II. Focusing an Evaluation	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Determine whether a program is a good candidate for evaluation (interest, resources, expertise, capacity)	0	1	2	3
b. Determine the purpose of evaluation (stakeholders, audience, etc.)	0	1	2	3
c. Know when to use different types of evaluation (process, outcome, etc.)	0	1	2	3
d. Develop evaluation questions from a logic model	0	1	2	3
e. Develop an evaluation plan (indicators, data sources, etc.)	0	1	2	3
f. Manage an evaluation (conduct, budget, create timeline, monitor, critique)	0	1	2	3
g. Develop a plan for the continual refinement of the evaluation process	0	1	2	3

III. Evaluation Questions and Designs	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Develop evaluation questions that match the goals of the evaluation	0	1	2	3
b. Match evaluation questions to levels of logic model (inputs, outputs, outcomes)	0	1	2	3
c. Define and distinguish indicators for success (e.g., metrics)	0	1	2	3
d. Generate appropriate evaluation questions based on audience, culture, program context, purpose, stakeholders	0	1	2	3
e. Knows different types of evaluation designs (pre-post, longitudinal, retrospective)	0	1	2	3
f. Match evaluation design to evaluation questions (what needs to be known)	0	1	2	3
g. Adapt designs to limitations (funding, time, resources, expertise)	0	1	2	3

IV. Evaluation Methods	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Understand when to use qualitative method	0	1	2	3
b. Understand when to use quantitative method	0	1	2	3
c. Describe the strengths and limitations of different qualitative methods	0	1	2	3

(Continued.)

IV. Evaluation Methods	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
d. Describe the strengths and limitations of different quantitative methods	0	1	2	3
e. Understand the interactive relationship between qualitative and quantitative methods.	0	1	2	3
f. Apply appropriate methods to answer evaluation questions	0	1	2	3
g. Develop survey questions	0	1	2	3
h. Develop protocols for focus groups and interviews	0	1	2	3
i. Develop observation protocol	0	1	2	3
j. Write methods section for evaluation report and Institutional Review Board (IRB)	0	1	2	3

V. Collecting and Handling Data	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Understand institutional requirements for collecting data with human subjects (IRB)	0	1	2	3
b. Conduct focus groups and interviews	0	1	2	3
c. Understand strategies for effective data collection (consent, timing, facilitation, setting, non disruptive, working with special populations)	0	1	2	3

(Continued.)

V. Collecting and Handling Data	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
d. Process, handle and store data (working with data sets, creating data code books, transcripts)	0	1	2	3
e. Critique tools and instruments (for reliability and validity)	0	1	2	3
f. Use technology (web-based surveys, photo techniques)	0	1	2	3
g. Understand the criteria used to select a specific data collection method that best suits the circumstances	0	1	2	3

VI. Analyzing and Interpreting Data	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Understand basic concepts in analyzing and interpreting qualitative data (e.g. triangulation, member checks)	0	1	2	3
b. Select and apply descriptive statistics (e.g. frequencies, means, standard deviation, range)	0	1	2	3
c. Understand assumptions, properties and limitations of inferential statistics (e.g. parametric/non-parametric data, data diagnostics)	0	1	2	3
d. Select and conduct appropriate procedures for data analysis (includes qualitative and quantitative software packages)	0	1	2	3

(Continued.)

VI. Analyzing and Interpreting Data	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
e. Interpret findings and construct conclusions	0	1	2	3
f. Identify limitations of results	0	1	2	3
g. Identify alternative means/models for analyzing and interpreting data	0	1	2	3

VII. Communicating Evaluation Results	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Identify specific information needs of a given audience (to include anticipating information requirements beyond that stated at the beginning of the evaluation)	0	1	2	3
b. Match content of evaluation report to audience needs	0	1	2	3
c. Know standard content of evaluation reports	0	1	2	3
d. Develop different types of evaluation reports (full report, executive summary, impact statement/success story, marketing materials, media strategies, scholarly dissemination)	0	1	2	3
e. Develop program recommendations and commendations	0	1	2	3

VIII. Knowledge of social science re: violent extremism	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. Pathways toward radicalization	0	1	2	3
b. Pathways toward deradicalization	0	1	2	3
c. Awareness of the key factors involved in both (e.g., socioeconomic, political, ethnic, religious, historical, cultural, geographic)				

IX. History & culture of violent extremism	I know nothing about this	I understand the basic concept	I can implement this concept with assistance	I can implement this concept independently and/or teach it to others
a. History or anti-West Islamic extremism	0	1	2	3
b. Perspectives of contemporary extremists	0	1	2	3
c. Current risk reduction initiatives	0	1	2	3

This assessment is based upon the National 4-H Learning Priorities Evaluating for Impact Committee (Arnold, M. E., Calvert, M. C., Cater, M., D., Evans, W., LeMenestrel, S., Silliman, B., & Walahoski, J. S. (2008). *Evaluating for impact: Educational content for professional development.* Washington, DC: National 4-H Learning Priorities Project, Cooperative State Research, Education, & Extension Service, USDA).

Competency 1: Program planning for effective program evaluation

The first competency area involves understanding and using program development logic models to plan and implement programs. Particular emphasis is placed on identifying short-, medium-, and long-term outcomes and their importance in setting the stage for effective evaluation of the program. At this level, individuals also learn to identify other areas of a logic model that can also provide important evaluation information.

Novice level: Individuals are familiar with terminology and reasoning of one or more logic models.

Demonstrated competencies:

- Define and give examples of each level of the Targeting Outcomes of Programs model or UWEX Logic Model.
- Explain logic model steps used in an established program (e.g. talk through a planned program).
- Describe the role of the logic model in the larger process of program development (e.g. note context of assessment, implementation, and evaluation).
- Explain role and significance of program evaluation standards of utility, feasibility, propriety, and accuracy.
- Explain the role and importance of ethics in evaluation.
- Describe how to use results of evaluation to modify or extend a planned program (e.g. criteria, processes).
- Explain the role of reassessment in the process of evaluation and reprogramming.

Advanced beginner level: Individuals are skilled in generating, explaining, and assessing the results of a logic model.

Demonstrated competencies:

- Develop a small-scale logic model (with assistance).
- Explain program/logic model to immediate stakeholders (e.g. those directly involved in the program).
- Explain role of program/logic model in the larger county (state) plan-of-work (in the case of integrated programming, link a specific program to larger goals).
- Explain subtopics of evaluation standards of utility, feasibility, propriety, and accuracy, with examples related to logic model for program.
- Explain lessons learned after completing sample program evaluation and interpret changes in program or evaluation to immediate stakeholders.

Practitioner level: Individuals are skilled in interpreting a logic model and engaging stakeholders in using the model to improve programs.

Demonstrated competencies:

- Guide a stakeholder group in development of a logic model for a program.
- Explain program and logic model to a broad group of stakeholders.
- Lead or contribute to development of a logic model for a larger system.
- Explain (at planning and/or reassessment) the link between specific program outcomes and strategies used to accomplish those outcomes.
- Explain the relevance and importance of standards for a specific program logic model, with examples from practice.
- Explain the relevance and importance of standards for a specific program logic model, with examples from practice.
- Explain specific processes and products of a program evident in the evaluation of a specific program.

Resources to assist with this competency area:

American Evaluation Association. (2008). *Guiding principles for evaluators*. Retrieved June 9, 2008, from http://www.eval.org/Publications/GuidinqPrinciplesPrintable.asp

Boone, E. J., Safrit, R. D., & Jones, J. (2002). *Developing programs in adult education: A conceptual programming model* (2nd ed.). Evanston, IL: Waveland Press.

Centers for Disease Control, Evaluation Working Group. (2005). Retrieved June 9, 2008, from http://www.cdc.gov/eval/standard.htm

Douglah, M. (1998). *Developing a concept of extension program evaluation*. Retrieved January 4, 2008, from http://learninqstore.uwex.edu/pdf/G3658-7.PDF

Edna McConnell Clark Foundation. (2006). *Evaluating your organizational capacity: A self-assessment tool*. Retrieved June 9, 2008, from http://www.emcf.org/pub/readingroom/mckinsevselfassessment.htm

Ohio State Extension. (2008). *Successful assessment methods and measurement in evaluation (SAMMIE)*. Retrieved January 4, 2008, from http://sammie.osu.edu/

Ramlow, M. E. (2007). *Program evaluation standards*. Retrieved December 27, 2007, from www.eval.org/EvaluationDocuments/progeval.html

Rockwell, K., & Bennett, C. (2008). *Hierarchy for targeting outcomes and evaluating their acheivement (TOP)*. Retrieved January 4, 2008, from http://citnews.unl.edu/TOP/enqlish/

Stufflebeam, D. L. (2002). The CIPP model checklist. Retrieved January 4, 2008, from http://www.wmich.edU/evalctr/checklists/checklistmenu.htm#models

University of Wisconsin Extension. (2008). *Logic model evaluation*. Retrieved January 4, 2008, from www.uwex.edu/ces/pdande/evalloqicmodel.html

University of Wisconsin Extension. (2008). *Program development*. Retrieved January 4, 2008, from www.uwex.edu/ces/pdande/proqdev/index.html

W.K. Kellogg Foundation. (1998). *W.K. Kellogg foundation evaluation handbook*. Retrieved from http://www.oip.usdoi.gov/BJA/evaluation/links/WK-Kelloqq-Foundation.pdf

Competency 2: Focusing an evaluation

When faced with conducting a program evaluation many people are not sure how to begin. Often times, this uncertainty results in an evaluation that has not been sufficiently planned and focused. Careful evaluation planning is directly connected to the quality of the evaluation results and is a critical first step in the evaluation process. This competency area covers the different purposes and types of program evaluation. A special emphasis is placed on developing evaluation questions that are linked to the program's theory or framework. Developing indicators and identifying data sources is also critical at this phase. Developing and following an evaluation protocol, timeline and project management plan also is important.

Novice level: Learner understands the elements of planning and focusing an evaluation

Demonstrated competencies:

- The purposes of evaluation.
- Process and outcome evaluation.
- Developing evaluation questions.
- Identifying indicators of change.
- Identifying data sources.
- Managing an evaluation.

Advanced beginner level: Should be able to create and use an evaluation plan to carry out an evaluation

Demonstrated competencies:

- Create and evaluation plan.
- Conduct the program evaluation.
- Monitoring of the evaluation.
- Evaluation of the evaluation.

Practitioner level: Should be able to demonstrate skill in planning and implementing a program evaluation.

Demonstrated competencies:

- Plan and conduct a complete evaluation with minimal guidance.

Resources to assist with this competency area:

Baumberger, M., Rugh, J., & Mabry, L. (2006). *First clarify the purpose: Scoping the evaluation. Real world evaluation*. Thousand Oaks, CA: Sage.

Davidson, E. J. (2005). What is evaluation, defining the purpose of the evaluation, identifying evaluation criteria, organizing the criteria and identifying potential sources of evidence. In *Evaluation methodology basics: The nuts and bolts of sound evaluation* (pp. 1–66). Thousand Oaks, CA: Sage.

Henderson, K. A., & Bialeschki, M. D. (2002). *Evaluating leisure services: Making enlightened decisions*. State College, PA: Venture.

McLaughlin, J. A., & Jordan, G. B. (2004). Using logic models. In J. S. Wholey, H. P. Hatry, & K. E. Newcomb (Eds.), *Handbook of practical program evaluation* (2nd ed.). San Francisco, CA: Jossey Bass.

Mertens, D. M. (2005). Evaluation. In *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed., pp. 47–88). Thousand Oaks, CA: Sage.

Patton, M. Q. (1997). Focusing evaluations: Choices, options, and decisions. In *Utilization-focused evaluation: The new century text* (3rd ed., pp. 177–194). Thousand Oaks, CA: Sage.

Preskill, H., & Russ-Eft, D. (2005). Focusing the evaluation. In *Building evaluation capacity: 72 activities for teaching and training* (pp. 75–100). Thousand Oaks, CA: Sage.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). Identifying issues and formulating questions: An overview of program evaluation, tailoring evaluations. In *Evaluation: A systematic approach* (7th ed., pp. 67–100). Thousand Oaks, CA: Sage.

Stecher, B. M., & Davis, W. A. (1987). Thinking about the focusing process, thinking about client concerns and evaluation approaches, how to formulate an evaluation plan. In *How to focus a program evaluation* (pp. 9–88). Thousand Oaks, CA: Sage.

Taylor-Powell, E., Steele, S., & Douglah, M. (1996). *Planning a program evaluation*. Madison, Wl: University of Wisconsin Cooperative Extension Publication G3658-1.

University of Wisconsin Extension. (2008). Logic model evaluation. Retrieved January 4, 2008, from www.uwex.edu/ces/pdande/evaluation/evalloqicmodel.html

Walker, R., & Wiseman, M. (2006). Managing evaluations. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 360–383). Thousand Oaks, CA: Sage.

Western Michigan University Program Evaluation Center. (n.d.). *The program evaluation standards*. Retrieved September 8, 2008, from http://www.wmich.edu/evalctr/jc/PGMSTNDS-SUM.htm

Competency 3: Evaluation design

Evaluations are only as good as the questions that drive them, so developing effective questions and strategies for each level of impact is a critical skill in program development. This competency section focuses on evaluation questions appropriate to quantitative and qualitative methods and their relation to outcome indicators and results. A special emphasis is placed on types of evaluation questions, and matching questions to indicators and outcomes. Learning the terminology and types of questions that can be used in evaluation is important. Advanced beginners are encouraged to work with a mentor or team to develop viable questions for real programs and understand the logical links between outcome goals and questioning strategies.

Novice level: Individuals are familiar with key concepts of design: evaluation questions and indicators, timelines.

Demonstrated competencies:

- Should be able to explain types of evaluation questions – open-ended, close-ended – content, process.
- Distinguish between short-, mid-, and long-term outcome questions.

- Define and distinguish indicators.
- Know several quantitative and qualitative methods.
- Explain and follow an external timeline.

Advanced beginner level: Have developed skills in generating, explaining, and assessing questions and indicators and be able to match them to appropriate evaluation methods.

Demonstrated competencies:

- Able to generate evaluation questions.
- Able to link indicators to framework stages.
- Select and use at least one quantitative and one qualitative indicator in consultation with mentor.
- Explain and follow external timeline.

Practitioner level: Skilled in preparing questions and evaluation methods to effectively assess and improve programs.

Demonstrated competencies:

- Able to generate and edit questions on own, based on program objectives.
- Able to adapt questions to audience and methods.
- Able to link appropriate indicators to program outputs and outcomes.
- Able to select and use more than one method on own.
- Able to adjust or augment to changes in plan or program.

Resources to assist with this competency area:

Bamberger, M., Rugh, J., & Mabry, L. (2006). *Real world evaluation*. Thousand Oaks, CA: Sage.

Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions*. San Francisco, CA: Jossey-Bass.

CYFERNet. (2008). *Evaluation: Evaluation design and methods*. Retrieved August 27, 2008, from http://cvfernet.ces.ncsu.edu/cvfres/browse3.php?catid=610&cateqorvname=Evalu ation+Desiqn+and+Methods&search=Evaluation&subcat=Desiqning+a+Proqram+E valuation&searchtype=browse

CYFERNet. (2008). *Evaluation: Evaluating outcomes for early childhood, school-age, and teens*. Retrieved August 27, 2008, from http://cvfernet.ces.ncsu.edu/cvfres/browse2.php?search=Evaluation

Department of Agricultural and Extension Education. (2008). *AEE 577 Evaluation in agricultural and extension education, Class II: Approaches and models of evaluation*. Retrieved January 4, 2008, from www.cals.ncsu.edu/aqexed/aee577/Class%20ll/aee577class2.html

Douglah, M. (1998). *Developing a concept of extension program evaluation*. Retrieved January 4, 2008, from http://learninqstore.uwex.edu/pdf/G3658-7.PDF

Durfee, W., & Chase, T. (2003). *Brief tutorial on Gantt charts*. Retrieved August 27, 2008, from http://www.me.umn.edu/courses/me4054/assianments/qantt.html

Earthman, E., Richmond, L. S., Peterson, D. J., Marczek, M. S., & Betts, S. C. (1999). *Adapting evaluation measures too hard to reach audiences*. Retrieved August 27, 2008, from http://aq.arizona.edu/fcs/cvfernet/evaluation/adapeval.pdf

Hewitt, B. (2007). *Business with CSREES, FY 2007–2011 POW planning. Module 3: The planned programs section*. Retrieved August 27, 2008, from www.csrees.usda.qov/business/reportinq/planrept/traininqfv0711.html

Ohio State Extension. (2008). *Successful assessment methods and measurement in evaluation (SAMMIE)*. Retrieved January 4, 2008, from http://sammie,osu.edu/

Rockwell, K., & Bennett, C. (1995). *Targeting outcomes of programs*. Retrieved August 27, 2008, from http://citnews.unl.edu/TOP/index.html

Rockwell, K., & Bennett, C. (2008). *Hierarchy for targeting outcomes and evaluating their acheivement*. Retrieved January 4, 2008, from http://citnews.unl.edu/TOP/enqlish/

Silliman, B. (2007). *Critical indicators of youth development outcomes*. Retrieved December 27, 2007, from www.nationa!4-hheadquarters.qov/librarv/lndicators 4H MM.pdf

Stufflebeam, D. L. (2002). *The ClPP model checklist*. Retrieved January 4, 2008, from www.wmich.edu/evalctr/checklists/cippchecklist.htm#process

United Way. (2008). *Outcome measurement resource network*. Retrieved January 7, 2008, from http://national.unitedwav.orq/outcomes/resources/mpo/examples.cfm

University of Arizona. (2007). *Beyond basics: Evaluating community-based Programs. Module II: Selected design concepts*. Retrieved August 27, 2008, from http://ag.arizona.edu/fcs/cyfernet/cyfar/PartsFrameset.htm

University of Wisconsin Extension. (2008). *Logic model evaluation*. Retrieved January 4, 2008, from www.uwex.edu/ces/pdande/evaluation/evalloqicmodel.html

University of Wisconsin Extension. (2008). *Program development*. Retrieved January 4, 2008, from www.uwex.edu/ces/pdande/proqdev/index.html

W.K. Kellogg Foundation. (2008). *Evaluation questions*. Retrieved August 27, 2008, from http://www.wkkf.orq/Default.aspx?tabid=90&CID=281&ltemlD=2810011&NID=2820011&LanauaaelD=0

Competency 4: Evaluation methods

This competency involves gaining knowledge of specific methods often used in evaluations. Individuals know about quantitative and qualitative methods and their appropriate use.

Novice level:

Demonstrated competencies:

- Able to explain the difference between qualitative and quantitative methods.
- Able to identify appropriate qualitative methods for evaluation.
- Able to identify appropriate quantitative methods for evaluation.
- Able to learn the difference between inferential and descriptive statistics.

Advanced beginner level:

Demonstrated competencies:

- Knows the relationship between evaluation questions and evaluation methods.
- Can apply appropriate methods to specific evaluation questions.
- Knows basic descriptive statistics (e.g. means, median, mode, range, SD, etc.).

Practitioner level:

Demonstrated competencies:

- Able to choose and use appropriate methods for evaluation question.
- Able to apply appropriate methods to specific evaluation questions.
- Knows basic inferential statistics (e.g. *t*-tests, one-way ANOVA, etc.).

Resources to assist with this competency area:

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Baugh, E., & Guion, L. A. (2006). Using culturally sensitive methodologies when researching diverse cultures. *Journal of Multi-disciplinary Evaluation, 4*. Retrieved from http://evaluation.wmich.edu/imde/JMDENum004.html

Baumberger, M., Rugh, J. M., & Mabry, L. (2006). *Real world evaluation*. Thousand Oaks, CA: Sage.

Bouffard, S., & Little, P. (2004, August). *Detangling data collection: Methods for gathering data*. Cambridge, MA: Harvard Family Research Project, Harvard University.

Cohen, C. (2006). Evaluation learning circles: A sole proprietor's evaluation capacity-building strategy. *New Directions in Evaluation, 111*, 85–93.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed-method approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Diem, K. (2002). Using research methods to evaluate your extension program. *Journal of Extension, 40*(6). Retrieved from http://www.ioe.orq/ioe/2002december/a1.shtml

Fitzpatrick, J., Sanders, J., & Worthen, B. (2003). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). New York: Allyn & Bacon.

Greene, J., & Caracelli, V. J. (1997). Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms. *New Directions for Evaluation, 74*, 1–95.

Henderson, K. A. (2006). *Dimensions of choice: Qualitative approaches to parks, recreation, sport, and leisure research*. State College, PA: Venture.

Henderson, K. A., & Bialeschki, M. D. (2002). *Evaluating leisure services: Making enlightened decisions*. State College, PA: Venture.

Kane, M., & Trochim, W. M. K. (2006). *Concept mapping for planning and evaluation*. Thousand Oaks, CA: Sage.

Ohio State Extension. (2008). *Successful assessment methods and measurement in evaluation (SAMMIE)*. Retrieved January 4, 2008, from http://sammie.osu.edu/

Patton, M. Q. (1987). An introduction to qualitative methods, when to use qualitative methods. In *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.

Patton, M. Q. (2001). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Preskill, H., & Russ-Eft, D. (2005). *Building evaluation capacity: 72 activities for teaching and training*. Thousand Oaks, CA: Sage.

Purdue University Writing Lab. (n.d.). *Sample research report*. Retrieved from http://owl.enqlish.purdue.edu/media/pdf/20070515024844669.pdf

Purdue University Writing Lab. (n.d.). *Writing your research project report*. Retrieved from http://owl.english.purdue.edu/handouts/WAC/CDFS/powerpoint/researchproject report.ppt#256

Stecher, B. M., & Davis, W. A. (1987). *How to focus an evaluation*. Newbury Park, CA: Sage.

W.K. Kellogg Foundation. (1998). *W.K. Kellogg Foundation evaluation handbook*. Retrieved from http://www.oip.usdoi.qov/BJA/evaluation/links/WK-Kelloqq-Foundation.pdf

Competency 5: Collecting and handling data

This competency focuses on understanding how to collect and manage quantitative and qualitative data. Topics include the ethics and procedures for the collection, storage and processing of data; developing a quantitative data set; data collection methods; developing a data collection methods protocol; and matching data collection methods to evaluation questions.

Novice level:

Demonstrated competencies:

- Knows ethics of data collection, processing, and storage.
- Knows different data collection methods.
- Knows about processing and handling quantitative and qualitative data.
- Knows about processing and handling quantitative and qualitative data.
- Knows about the standard parts of a written methods section for an evaluation report or article.

Advanced beginner level:

Demonstrated competencies:

- Has gained skills in preparing IRB packages.
- Can match data collection methods to different evaluation questions.
- Can apply the proper procedures for handling data using a mock data set.
- Can apply what they have learned about writing methods sections for evaluation reports or articles.

Practitioner level:

Demonstrated competencies:

- Can apply skills in preparing IRB packages to real study.
- Can apply knowledge of data collection methods to the creation of a simple data collection tool.
- Can critique other existing data collection tools.
- Can apply the proper procedures for handling data using a real data set.
- Can apply knowledge of how to write an evaluation report.

Resources to assist with this competency area:

American Evaluation Association. (2004). *Guiding principles of evaluators*. Fairhaven, MA: Author. Retrieved from http://www.eval.orq/Rublications/GuidinqPrinciples.asp

American Evaluation Association. (2004). *Guiding principles training*. Fairhaven, MA: Author. Retrieved from http://www.eval.orq/GPTraininq/GPTraininqOverview.asp

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Bamberger, M., Rugh, J., & Mabry, L. (2006). *Real world evaluation: Working under budget, time, data, and political constraints*. Thousand Oaks: Sage.

Betts, S. C., Peterson, D. J., & McDonald, D. A. (2005). More tips: What if a cooperative extension professional must work with two or more institutional review boards? *Journal of Extension, 43*(4). Retrieved from http://www.ioe.org/ioe/2005auqust/tt1.shtml

Bouffard, S., & Little, P. (2004). *Detangling data collection: Methods for gathering data*. Cambridge, MA: Harvard Family Research Project. Retrieved from http://www.qse.harvard.edu/hfrp/content/proiects/afterschool/resources/snapshot5.pdf

Brown, R., Martin, S., & Weigel, D. (2004). What cooperative extension professionals need to know about institutional review boards: Recruiting participants. *Journal of Extension, 42*(6). Retrieved from http://www.ioe.org/ioe/2004december/tt1.shtml

Conrad, F., & Schober, M. (2007). *Envisioning the survey interview of the future*. New York, NY: John Wiley & Sons.

Cooksy, L. (2005). The complexity of the IRB process: Some of the things you wanted to know about IRBs but were afraid to ask. *American Journal of Evaluation, 26*(3), 352–361.

Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons.

Fink, A. (2002). *How to manage, analyze, and interpret survey data* (2nd ed.). Thousand Oaks: Sage.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks: Sage.

Hillier, A. (2008). Childhood overweight and the built environment: Making technology part of the solution rather than part of the problem. *The ANNALS of the American Academy of Political and Social Science, 615*(1), 56–82.

Howell, J., Miller, P., Park, H., Sattler, D., Schack, T., Spery, E., ... Palmquist, M. (2005). *Reliability and validity*. Writing@CSU. Colorado State University Department of English. Retrieved June 25, 2008, from http://writinq.colostate.edu/quides/research/relval/

Leahy, J. (2004). *Using excel for analyzing survey questionnaires*. University of Wisconsin Extension Publication G3658-14. Retrieved from http://leaminqstore.uwex.edu/pdf/G3658-14.pdf

Litwin, M. (1995). *How to measure survey reliability and validity* (Survey Kit, Vol. 7). Thousand Oaks: Sage.

Lopez, M. (2002). *Youth vote national youth survey June 2002 data codebook*. Retrieved from http://www.civicvouth.org/PopUps/vouthvotenational surveyiune2002codebook.pdf

Martin, S., Weigel, D. & Brown, R. (2005). What cooperative extension professionals need to know about institutional review boards: Obtaining consent. *Journal of Extension, 43*(2), Article 2TOT1. Retrieved from http://www.ioe.org/ioe/2005april/tt1.shtml

McDonald, D. A., Peterson, D. J., & Betts, S C. (2005). More tips: What if a cooperative extension professional must work with Native American institutional review boards? *Journal of Extension, 43*(5). Retrieved from http://www.ioe.org/ioe/2005october/tt1.shtml

Morgan, D. L., & Krueger, R. A. (1998). *The Focus group kit*. Thousand Oaks: Sage.

National Cancer Institute. (2013). *Human participant protections education for research teams*. Retrieved from http://cme.cancer.gov/c01/resources.php?file=%2F2%2Fmodules%2Fmodule06%2Eswf

National Cancer Institute. (2013). *Human participants protection training*. Retrieved from http://phrp.nihtraining.com/users/login.php

National Institute of Health. (2011). *Office of human subjects research*. Retrieved from http://ohsr.od.nih.gov/

National Network of Libraries of Medicine. (2006). *Outreach activity data collection form*. Bethesda, MD: Author. Retrieved from http://nnlm.Gov/evaluation/tools/Activitvlnfo.pdf

Newman, D. L., & Brown, R. D. (1996). *Applied ethics for program evaluation*. Thousand Oaks: Sage.

North Carolina State University . (n.d.). *Research with experimental subjects*. Raleigh, NC: Author. Retrieved from http://www.ncsu.edu/sparcs/tutorial/intro.php

Ohio State Extension. (2008). *Successful assessment methods and measurement in evaluation (SAMMIE)*. Retrieved January 4, 2008, from http://sammie.osu.edu/

O'Reilly, J., Hubbard, M., Lessler, J., Biemer, P., & Turner, C. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics, 10*(2), 197–214.

Patton, M. (2001). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks: Sage.

Peterson, D. J., McDonald, D. A., & Betts, S. C. (2005). More tips: Communicating with institutional review boards over the course of your project. *Journal of Extension, 43*(6). Retrieved from http://www.ioe.orq/ioe/2005december/tt1.shtml

Preskill, H., & Russ-Eft, D. (2005). *Building evaluation capacity: 72 activities for teaching and training*. Thousand Oaks, CA: Sage.

Richards, L. (2005). *Handling qualitative data*. Thousand Oaks, CA: Sage.

Rubin, H., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.

Russ-Eft, D. F. (1980). *Validity and reliability in survey research* (Technical Report #15). Palo Alto, CA: Statistical Analysis Group in Education, American Institutes for Research in the Behavioral Sciences (ERIC Document Reproduction Service No. ED279726).

Trochim, W. K. (2006). *Research methods knowledge base: Construct validity*. Retrieved from http://www.socialresearchmethods.net/kb/constval.php

Trochim, W. K. (2006). *Research methods knowledge base: Data preparation*. Retrieved from http://www.socialresearchmethods.net/kb/statprep.php

Trochim, W. K. (2006). *Research methods knowledge base: Reliability*. Retrieved from http://www.socialresearchmethods.net/kb/reliable.php

Weigel, D., Brown, R., & Martin, S. (2004). What cooperative extension professionals need to know about institutional review boards. *Journal of Extension, 42*(5). Retrieved from http://www.ioe.org/ioe/2004october/tt1.shtml

Weigel, D., Martin, S., & Brown, R. (2005). What cooperative extension professionals need to know about institutional review boards: Risks and benefits. *Journal of Extension, 43*(1) Article 1TOT1. Retrieved from http://www.ioe.org/ioe/2005februarv/tt1.shtml

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2004). *Handbook of practical program evaluation* (2nd ed.). San Francisco: Jossey-Bass.

W.K. Kellogg Foundation. (1998). *Evaluation handbook: Analyzing and interpreting data*, pp. 87–95. Retrieved from http://www.wkkf.org/~/media/62EF77BD5792454B807085B1 AD044FE7.ashx

University of Wisconsin Extension Publication G3658-06 *Analyzing quantitative data*. Retrieved from http://learninqstore.uwex.edu/pdf/G3658-6.pdf

University of Wisconsin Extension. (2003). Publication G3658-12 *Analyzing qualitative data.* Retrieved from http://learningstore.uwex.edu/Assets/pdfs/G3658-12.pdf

Competency 6: Analyzing and interpreting data

This competency area covers basic analysis procedures available for both quantitative and qualitative data. Using statistical software (such as SPSS) individuals know how to perform descriptive and inferential analyses and know how to interpret the results.

Novice level:

Demonstrated competencies:

- Can identify appropriate procedures for analyzing data.
- Understands basic findings and can explain them to stakeholders.

Advanced beginner level:

Demonstrated competencies:

- With guidance can apply appropriate procedures to conduct data analysis.
- With guidance, can interpret findings, construct conclusions, and develop formal methods to communicate them.

Practitioner level: Individuals are skilled in analyzing quantitative and qualitative data and at interpreting findings and articulating reasonable conclusions.

Demonstrated competencies:

- Selects and conducts appropriate analysis procedures to program data.
- Can appropriately interpret findings and develop conclusions from analysis of program data.

Resources to assist with this competency area:

Betts, S. C., & Hoffman-Tepper, K. (Eds.). (2001). *Beyond Basics: Evaluating Community-Based Programs Training Curriculum*. Retrieved from http://ag.arizona.edu/sfcs/cyfernet/cyfar/Curriculum.pdf

Callor, S., Betts, S. C., Carter, R., Marczak, M., Peterson, D., & Richmond, L. S. (2000). *Children, youth, and families at risk community-based project evaluation guide*. Tucson, AZ: University of Arizona Institute for Children, Youth and Families. Retrieved from http://aq.arizona.edu/fcs/cvfernet/cvfar/ststguide.pdf

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Upper Saddle River, NJ: Pearson Education.

Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Northcutt, N., & McCoy, D. (2004). *Interactive qualitative data analysis*. Thousand Oaks, CA: Sage.

Ohio State Extension. (2008). *Successful assessment methods and measurement in evaluation (SAMMIE)*. Retrieved January 4, 2008, from http://sammie.osu.edu/

Patton, M. Q. (1997). Deciphering data and reporting results: Analysis, interpretations, judgments, and recommendations. In *Utilization-focused evaluation: The new century text* (3rd ed., pp. 301–338). Thousand Oaks, CA: Sage.

Preskill, H., & Russ-Eft, D. (2005). *Building evaluation capacity: 72 activities for teaching and training*. Thousand Oaks, CA: Sage.

Salkind, N. (2007). *Statistics for people who think they hate statistics* (3rd ed.). Thousand Oaks, CA: Sage.

Silverman, D. (2001). *Interpreting qualitative data* (2nd ed.). Thousand Oaks, CA: Sage.

Statistical Resources on the on the web. Retrieved from http://www.psvchstat.missouristate.edU/scripts/dws148f/statisticsresourcesmain.a

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Taylor Fitz-Gibbon, C., & Lyons Morris, L. (1987). *How to analyze data*. Newbury Park , CA: Sage.

Taylor-Powell, E., & Renner, M. (2003). *Analyzing qualitative data*. Madison, Wl: University of Wisconsin-Extension. Retrieved from http://learninqstore.uwex.edu/pdf/G3658-12.PDF

Tromchin, M. K. (2006). *Research methods knowledge base: Descriptive statistics*. Retrieved from http://www.socialresearchmethods.net/kb/statdesc.php

University of Kentucky College of Agriculture. (2013). *Program development and evaluation resources*. Retrieved from http://www2.ca.uky.edu/agpsd/soregion.htm

University of Wisconsin Extension. (2003). Publication G3658-12. *Analyzing qualitative data*. Retrieved from http://learningstore.uwex.edu/Assets/pdfs/G3658-12.pdf

University of Wisconsin Extension. Publication G3658-14 *Using excel for analyzing survey questionnaires*. Retrieved from http://learninqstore.uwex.edu/pdf/G3658-14.pdf

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2004). Part three. In *Handbook of practical program evaluation* (2nd ed.). San Francisco: Jossey-Bass.

Wolcott, H. F. (2001). *Writing up qualitative data* (2nd ed.). Thousand Oaks, CA: Sage.

Competency 7: Communicating evaluation results

This competency involves learning how to convert evaluation results into forms of communication that are useful to various stakeholders. The purposes of reporting, the content of a standard evaluation report, how to identify stakeholders, and how to present the results that matter most to different stakeholder groups are included.

Novice level:

Demonstrated competencies:

- Knows the purpose of planning and reporting evaluation results.
- Knows can identify different stakeholder audiences and items of importance to each audience.
- Knows the standard content of evaluation reports.

Advanced beginner level:

Demonstrated competencies:

- Individual can develop basic evaluation reports which include all standard sections.
- Individual knows the different types of evaluation reports.

Practitioner level:

Demonstrated competencies:

- Individual can develop complete evaluation reports which are adapted for specific audiences.

Resources to assist with this competency area:

Betts, S. C., & Hoffman-Tepper, K. (Eds.). (2001). *Beyond the Basics: Evaluating Community-Based Programs, University of Arizona*. Retrieved from http://ag.arizona.edu/sfcs/cyfernet/cyfar/Curriculum.pdf

Callor, S., Betts, S., Garter, R., Marczack, M., Peterson, D., & Richmond, L. (2000). *Community-based project evaluation guide, University of Arizona, Institute of Children, youth, and family.* Retrieved from http://aa.arizona.edu/fcs/cvfernet/cvfar/ststquide.pdf

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: alternative approaches and practical guidelines* (3rd ed.). Upper Saddle River, NJ: Pearson Education.

Ohio State Extension. (2008). *Successful assessment methods and measurement in evaluation (SAMMIE)*. Retrieved January 4, 2008, from http://sammie.osu.edu/

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed., Chapter 13). Thousand Oaks, CA: Sage.

Success Story Guidelines. (2003). *University of Wisconsin Extension*. Retrieved from http://www.uwex.edu/ces/prs/success.cfm

Torres, R. T., Preskill, H., & Piontek, M. E. (2005). *Evaluation strategies for communicating and reporting* (2nd ed.). Thousand Oaks, CA: Sage.

University of Kentucky. (2013). College of Agriculture, *Program Development and Evaluation Resources*. Retrieved from http://www2.ca.uky.edu/agpsd/soregion.htm

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2004). *Handbook of practical program evaluation* (2nd ed., Part Three). San Francisco: Jossey-Bass.

Competency 8: Knowledge of social science re: violent extremism

This competency involves learning about the myriad pathways into, and away from, violent extremism, at both individual and group levels.

Novice level:

Demonstrated competencies:

- Can name several pathways that tend to lead individuals (or groups) into, or away from, radicalization.
- Has had the equivalent of an introductory course on social psychology or sociology.

Advanced beginner level:

Demonstrated competencies:

- Individual understands the basic underlying psychological mechanisms thought to affect change in participants of a given risk reduction initiative.
- Individual is conversant in theories of radicalization either from a psychological, or sociological perspective.
- Individual understands some of the basic limitations of current research on radicalization and deradicalization.

Practitioner level:

Demonstrated competencies:

- Individual understands current theories, and can independently theorize, about the social psychological and sociological factors thought to compel individuals into, and away from, terrorism.
- Individual can critique risk reduction initiative from a social science perspective.
- Individual can design risk reduction initiatives, based upon social science theory and methods.

Resources to assist with this competency area:

Aronson, E., Wilson, T. D., & Akert, R. M. (2007). *Social psychology* (6th ed.). Upper Saddle River, NJ: Pearson Education.

Bongar, B., Brown, L. M., Beutler, L. E., Breckenridge, J. N., & Zimbardo, P. G. (2007). *Psychology of terrorism*. New York, NY: Oxford University Press.

Cialdini, R. (1993). *Influence: The psychology of persuasion.* New York, NY: William Morrow.

Horgan, J. (2005). *The psychology of terrorism*. New York, NY: Routledge.

Horgan, J. (2009). *Walking away from terrorism*. New York, NY: Routledge.

Horgan, J., & Taylor, M. (2011). Disengagement, de-radicalization, and the arc of terrorism: Future directions for research. In R. Coolsaet (Ed.), *Jihadi terrorism and the radicalisation challenge: European and American experiences* (pp. 173–186). Farnham, Surrey: Ashgate.

Noricks, D. (2009). Disengagement and deradicalization: Processes and programs. In P. Davis & K. Cragin (Eds.), *Social science for counterterrorism* (pp. 299–320). Santa Monica, CA: Rand Corporation.

Rabasa, A., Pettyjohn, S. L., Ghez, J. J., & Boucek, C. (2010). *Deradicalizing Islamist extremists*. Santa Monica, CA: Rand Corp. Retrieved from http://www.rc.rand.org/pubs/monographs/MG1053/

Competency 9: History and culture of violent extremism

This competency involves learning about the historical/political/ideological roots of violent extremist/terrorist groups, including their grievances, objectives, and prior interactions with state actors. Additionally, it involves knowledge of state-sponsored responses to violent extremism/terrorism.

Novice level:

Demonstrated competencies:

- Individual knows of basic grievances between some extremist groups and their targets.
- Individual knows of basic tenets of some extremist groups' philosophies/ideologies that justify violence.
- Individual knows of some of state-sponsored responses to violent extremism.

Advanced beginner level:

Demonstrated competencies:

- Individual knows of the historical/political/ideological roots of a few extremist/terrorist groups, including their grievances, objectives, and prior interactions with state actors.
- Individual has a rudimentary understanding of the philosophical/religious justifications both for and against violent extremism.
- Individual has knowledge of a few risk reduction practices in place throughout the word.

Practitioner level:

Demonstrated competencies:

- Individual knows of the historical/political/ideological roots of several extremist/terrorist groups, including their grievances, objectives, and prior interactions with state actors.
- Individual has a nuanced understanding of the philosophical/religious justifications both for and against violent extremism.
- Individual has knowledge of the vast majority of current risk reduction practices in place throughout the word.

Resources to assist with this competency area:

Ashour, O. (2010). *The deradicalization of Jihadists*. New York, NY: Routledge.

Ballen, K. (2011). *Terrorists in love: The real lives of Islamic radicals.* New York, NY: Free Press.

Hannah, G., Clutterbuck, L., & Rubin, J. (2008). *Radicalization or rehabilitation: Understanding the challenge of extremist and radicalized prisoners*. Santa Monica, CA: RAND Corp.

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. doi:10.1080/09546551003594748

Rubin, B., & Rubin, J. C. (Eds.). (2002). *Anti-American terrorism and the Middle East: A documentary reader*. New York, NY: Oxford University Press.

Soufan, A., Fallon, M., Freedman, D., Borum, R., Horgan, J., Gelles, M., … McManus, B. (2010). The QIASS countering violent extremism (CVE) risk reduction project. *Qatar International Academy for Security Studies.* Retrieved from http://www.soufangroup.com/summary.pdf

This assessment is based upon the National 4-H Learning Priorities Evaluating for Impact Committee (Arnold, M. E., Calvert, M. C., Cater, M., D., Evans, W., LeMenestrel, S., Silliman, B., & Walahoski, J. S. (2008). *Evaluating for impact: Educational content for professional development.* Washington, DC: National 4-H Learning Priorities Project, Cooperative State Research, Education, & Extension Service, USDA).