JOURNAL FOR DERADICALIZATION

Spring 2022
No. 30
ISSN: 2363-9849

Research Methods Brief: Anatomy of Process Evaluations for P/CVE

Michael J. Williams[a1]
[a]The Science of P/CVE

Abstract
Process Evaluations are evaluations focused on understanding how a program is implemented. This also can include evaluating the extent to which a program is implemented according to plan (i.e., evaluating its "program fidelity"). In short, process evaluations seek to identify a program's "moving parts" to assess the extent to which they are functioning as intended. Ideally, that includes uncovering the theoretical mechanisms—the reasons "why"—a program's outputs or outcomes are (or are not) achieved. Understanding why a program is (or is not) working as well as expected is the backbone of evidence-based P/CVE program design and evaluation, and is essential to informing sound P/CVE program management decision-making. Consequently, without exception, good P/CVE-related research, or evaluation projects—those that are scientifically grounded—must include at least some element(s) of process evaluation. This research methods brief describes the fundamental components of process evaluations, common pitfalls and means to avoid those pitfalls, within the context of P/CVE program design and evaluation.

Article History
Received Mar 8, 2021
Accepted Mar 21, 2022
Published Mar 25, 2022

Keywords: Preventing Violent Extremism, PVE, Countering Violent Extremism, CVE, P/CVE, Design, Evaluation, Process Evaluation

Introduction

Process Evaluations are evaluations focused on understanding *how* a program is implemented (Wholey et al., 2010). This also can include evaluating the extent to which a program is implemented according to plan (i.e., evaluating its program fidelity). In other words, process evaluations seek to identify a program's "moving parts" to assess the extent to which they are functioning as intended. Ideally, that includes uncovering the theoretical mechanisms—the reasons "why"—a program's outputs or outcomes are (or are not) achieved (ibid.). **Understanding *why* a program is (or is not) working as well as expected is the backbone of evidence-based P/CVE program design and evaluation, and is essential to informing sound P/CVE program management decision-making**.

[1] Corresponding Author Contact: Michael J. Williams, Email: m.williams@thescienceofpcve.org, Twitter: @MickWilliamsPhD, ORCID iD https://orcid.org/0000-0001-5630-9814

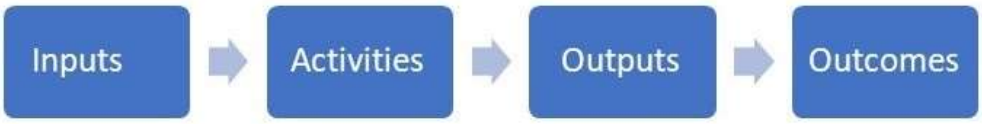
Consequently, without exception, good P/CVE-related research, or evaluation projects—those that are scientifically grounded—must include at least some element(s) of process evaluation. The purpose of doing so is to verify the extent to which each component of a program (a) was in fact performed, and (b) had the expected outcomes (that ostensibly contribute to the overall programmatic outcomes). Without such verifications, one cannot know whether a program's activities were implemented as intended, or whether they contributed, as intended, to the overall programmatic outcomes. The onus is on researchers and evaluators to demonstrate these critical points *empirically* to a rightfully skeptical audience (Sagan, 2011).

Pragmatism. As though scientific integrity was insufficient reason to engage in process evaluation, there is also a pragmatic reason for doing so. If, for example, a program does not produce the intended outcomes or impacts, or does not produce them to the desired degree, how are you—the P/CVE researchers, program evaluators, or program managers—going to explain that to the project's key stakeholders? Specifically, how are you going to explain the problem in a way that might be *useful* to those stakeholders? Process evaluations are intended, in part, to help us to understand where failure, or underperformance, occurs within a program (Wholey et al., 2010). If researchers or evaluators discover which part(s) are malfunctioning, so to speak, then program managers might have a chance to remedy the issue(s) (more on this to follow). Conversely, process evaluations can help to reveal high-performing areas of a program: areas that potentially can be leveraged to the program's advantage.

The following discussion covers the four common parts of a process evaluation's anatomy: a program's inputs, activities, outputs, and outcomes, depicted in Figure 1 below.[2]

Figure 1

Common components of process evaluations.



[2] These four components are not exclusive to process evaluations, but also are components of (for example) impact evaluations.

Inputs: Who is really in your sample?

"Inputs" of a program typically include the resources (including human resources) that perform the program's activities, and—hence—contribute to producing the program's outputs and subsequent outcomes. Therefore, for research and evaluation projects, inputs include those who participate in said projects. With respect to process evaluation, the objective is to verify who, indeed, those inputs/participants are. On the surface, this can be as basic, it would seem, as including anonymized demographic questions in a survey of those participants (e.g., to record participants' age, sex, education level, employment status, city/place of residence). However, this also should include what is easily overlooked: that those persons actually participated in your P/CVE program. For example, often surveys are administered via online platforms, and respondents are invited by email, or via links disseminated through SMS or social media; therefore, how do you know that those respondents actually have engaged with your P/CVE program? Even if respondents come from an email list of known program participants, and are sent a one-time-use link to complete the survey, how do you know that they are, indeed, the ones completing your survey? Therefore, at a minimum, part of a process evaluation must include measures to verify whether/how participants have engaged with the P/CVE program.

Activities

Measuring a program's "activities" entails more than the simple task of verifying that an activity has, indeed, occurred. For purposes of quality control, it should entail inquiry into *how* those activities were performed (Koehler & Fiebig, 2019; Koehler, 2017a, 2017b). For example, consider a P/CVE program whose operation entails some form of "call center" (or website, or social media conduit) for the organization to receive referrals. In this example, of course, we could (and should) measure the number of contacts that the organization receives, but that tells us very little about what could be called the "customer experience." As mentioned, process evaluations should focus not only on identifying the moving parts of a P/CVE program, but *why* they are/are not working well.

Therefore, in the present example, additional activity-related process evaluation data could include the following: a) the average amount of time that the organization takes to respond to such contacts (*Is it quick or sluggish?*), b) the ID of responders (*Are some of the organization's personnel faster, or in other ways better at responding to contacts*), c) what is

the sex of the new contact and does their customer experience tend to vary based upon the sex of the organization's responding representative. Note that, if desired, such indicators also could be analyzed for their correlations with other M&E variables of interest (e.g., clients' satisfaction with their care). Collection of some such indicators need not be labor intensive, but may be collected automatically (e.g., timestamps on incoming and outgoing communications, to assess "customer service" response times).

Outputs: Were the activities "effective?"

The chief objective in measuring the outputs (i.e., the immediate vs. relatively distal "outcomes" of the program's activities), is to verify that they were produced. On its face, this might sound obvious, or otherwise straight-forward, but consider the following example that highlights some of the nuance regarding this measurement issue. Consider a P/CVE program that includes a training, or other educational component, of some kind. The "activity," is the training program itself, and the "output" is that participants have learned what was covered in the training. Therefore, the output that must be measured is that participants have, indeed, sufficiently learned the material covered by the training.

At first glance, it might seem that this could be measured by some form of comprehension test, administered to participants after the training; if they score above a certain threshold (e.g., 70%, 80%, 90% correct), it could be presumed they have learned the material. However, in this example, the post-training comprehension test does not measure the effectiveness of the training: merely, that participants know the material covered by the training. It is possible that the training was ineffective, but that participants already knew the material covered by the training (or could infer the correct answers, based on how they were worded; see next heading on "Demand Characteristics"). Therefore, in this example, the process evaluation would need to measure participants both before and after the training, to demonstrate the extent to which the training was responsible for participants' learning outcomes.

As another example, experimental research on ostracism randomly assigns participants to be exposed to mild levels of ostracism (vs. inclusion): often through a game where participants are excluded (vs. included) in a ball-tossing game (Williams, 2009). Even in this paradigm, where researchers know with 100% certainty whether participants are (vs. are not) ostracized during the game, still this line of research includes a post-game questionnaire that

asks not one, but three, questions to verify whether the activity (i.e., the ball tossing paradigm) registered in the minds of participants: that it produced the presumed output (i.e., their perception that they were ostracized vs. included). Table 1, below, displays those questions.

In experimental parlance, these are known as manipulation checks (Hoewe, 2017). Without such verifications, we do not know whether a program's activities were implemented as intended, or whether they contributed (as intended) to the program's outputs and subsequent outcomes. To reiterate, the onus is on researchers and evaluators to demonstrate these critical points *empirically* to a rightfully skeptical audience (Sagan, 2011).

Table 1

Ostracism Manipulation Checks (from Williams, 2009)

For each question, please circle the number to the right that best represents the feelings you were experiencing during the game.	Not at all				Extremely
For the next three questions, please circle the number to the right (or fill in the blank) that best represents the thoughts you had during the game .					
I was ignored.	1	2	3	4	5
I was excluded .	1	2	3	4	5
Assuming that the ball should be thrown to each person equally, what percentage of the throws did you receive?	_____ %				

Outcomes: The Problems of Demand Characteristics and Expectancy Effects (and Ways to Mitigate Them)

Demand characteristics are features of a situation that make the situation appear (to those encountering the situation) to expect (or "demand") a certain response from them (Colman, 2015a). Consequently, individuals tend to modify their behavior (either consciously

or unconsciously) to conform to (or, perhaps, rebel against) those expectations (Brewer & Crano, 2000). Therefore, those individuals might, for example, try to "behave appropriately" or tell researchers or evaluators "what they want to hear."

This is a potential measurement problem in virtually all research and evaluations whereby participants are at least somewhat aware of what researchers/evaluators are assessing. This is a profoundly important matter, for example, with prison-based CVE programs whereby prisoners might participate in a program out of a belief that doing so, and telling data collectors "what they want to hear" with respect to changes in a prisoner's self-reported outcomes, will increase the prisoner's chance of earning early release or other rewards. Consequently, to minimize the effects of demand characteristics, research and evaluation designs must strive to remove incentives for respondents to report "false improvement" on any/all participant outcomes. Therefore, it must be made explicit to P/CVE program participants that there are no potential rewards, or punishments, based upon their program-related outcomes (and that their data will be anonymized).

Additionally, there are several other research design features that can help to minimize the demand characteristic of a given P/CVE program. First, is to employ a cover story that obfuscates the true or complete nature of the program. For example, a P/CVE youth mentorship program might not need to disclose to participants, prior to data collection (if ever), that P/CVE was the focus of the program. Instead, such a program could be framed in other honest, albeit more general ways, for example, in terms of "youth empowerment" or "positive youth development." This is not unethical practice; to the contrary, research on labeling theory highlights that labeling a person can affect their self-image and real-world performance outcomes (e.g., see Steele, 2010). Therefore, it would be irresponsible to risk labeling someone as a potential violent extremist by framing their participation in a program in terms of preventing or countering violent extremism.

Another design feature that can help to minimize a data collection's demand characteristic is to embed outcome measures of interest within a broader range of outcome measures (some, or all, of which might be included merely for such obfuscation). For example, a self-report survey measure of radicalization could be embedded within a larger survey of general "public opinions." That way, respondents would be less able to discern what is being measured by the primary questions of interest.

Additionally, any self-reported demographic questions should appear *after* the primary questions of interest. That way, respondents would be unable to guess, in advance, that their responses might be assessed according to their demographic characteristics (e.g., their sex, age, race, religious affiliation, etc.) and that they could be expected to respond in a certain way, based upon those characteristics.

Expectancy effects. Furthermore, if possible, data collectors should be blind to (uninformed of) the expectations/hypotheses, and/or experimental conditions, pertaining to the P/CVE program. Due to so-called "expectancy effects," if data collectors are aware of the expected outcomes of an intervention, they might (perhaps inadvertently) encourage participants to provide responses that are congruent with those expectations (Colman, 2015b; Rosenthal & Fode, 1963).

Expectancy effects are not a trivial matter. In a classic experiment, expectancy effects were induced by telling 12 research assistants what kind of behavior to expect from lab rats that were assigned to them (to run through a maze learning task; Rosenthal & Fode, 1963). All of the rodents were of the *same* genetic strain, but six lab assistants were falsely told that the rodents assigned to them had been bred for "maze brightness" (intelligence in running maze courses) and the other six were falsely told that their rodents had been bred for "maze dullness." Subsequently, the lab assistants ran the rodents in the maze-learning experiment, and—true to expectancy effects (and what is also known as the "Rosenthal effect")—the rodents that were believed to be smarter performed significantly better (i.e., actually learned faster) than those believed to be less intelligent. It was suggested that this effect was due to subtle differences in the ways that the lab assistants inadvertently handled the animals assigned to them. That incredible finding has been replicated in another of Rosenthal's studies, but in a "real-world" context: a field experiment on elementary school teachers' expectancies and their students' IQ scores (Rosenthal & Jacobson, 1966).

Finally, another design feature that can help to minimize the demand characteristic of a data collection is to collect distal data. In other words, collect data at one or more time points that are relatively far in the future relative to when their participation in the program occurred (e.g., one to six months after their participation). In short, a survey administered immediately after a participant engages with a P/CVE program would have a relatively larger demand

characteristic than if the survey was administered, for example, one month after their participation.

Combining evaluation types

As mentioned, all good P/CVE research and evaluations require some foray into process evaluation. This rightfully implies that process evaluations can be combined with other forms of research and evaluations, to serve the informational needs of key stakeholders. This does not necessarily require a multiplication of effort (e.g., two types of evaluations—a process evaluation and, for example, impact evaluation—entailing twice the work). It could be that the process evaluation can be easily embedded into an overarching data collection (e.g., by adding items to a survey, or broadening an observational data collection protocol).

Conversely, it might be that combining evaluation types, within a given project, could exponentially increase the overall research and evaluation effort. For example, data analysis can be enormously time-consuming, especially if the data are "messy" (e.g., not conforming to the mathematical assumptions underlying various analytic techniques, and/or replete with missing data on one or more variables). Therefore, answering even a few additional evaluation questions could mean large increases in the hours required for analyses devoted to answering those questions. The principle is merely this: one need not avoid combining multiple evaluation types into an overall research or evaluation project, but consider carefully the data collection and analyses that will be required to answer a given set of questions, and plan accordingly to ensure that the research or evaluation team has sufficient time and human capital to answer those questions.

Reporting unfavorable results: The saving grace of process evaluations

Disclosing unfavorable findings about one's P/CVE program does not necessarily mean that the program will lose monetary or political support. There is empirical evidence that disclosing unfavorable information about an entity makes the revelation of subsequent, favorable information of that entity even more believable than it otherwise might have been (so-called "two-sided communication;" see Schumpe et al., 2018). To disclose truths about unfavorable information takes integrity and honesty. Therefore, if a P/CVE program has any

null, or otherwise unfavorable results, recognize that disclosing them will tend to make the program's favorable results all the more believable. (The author has never been a party to an evaluation that has not had at least some favorable results to report.)

Conversely, studies also have demonstrated the intuitive finding that, after trust has been broken, it can be very difficult to rebuild it (Lewicki, 2006). Furthermore, research has shown that organizations tend to fail if they are unable to build the trust and commitment needed for maintaining cooperation with their professional partners (Ghoshal & Moran, 1996). For researchers, evaluators, and program managers alike: do not risk your long-term reputation, or a P/CVE program's long-term viability, by a short-sighted impulse to make all of a research or evaluation project's results appear favorable.

The saving grace of process evaluations. The need to disclose any/all null, or otherwise undesirable, programmatic outcomes underscores the importance—potentially, the saving grace—of process evaluations. If/when the results of a P/CVE program are not as favorable as had been hoped, how might program managers explain that? It would not be helpful to anyone, nor would it help one's professional reputation (as a researcher, evaluator, or manager of a P/CVE program) to shrug one's shoulders in response. Instead, a well-performed process evaluation might help to redeem unfavorable findings from a P/CVE program, if one is able to speak to what went awry *and why*. As mentioned, this is one of the pragmatic reasons for embedding process evaluation into every research and evaluation project. If a P/CVE program does not produce its intended outcomes or impacts (or does not produce them to the desired degree), a well-performed process evaluation can help program managers and other key stakeholders to understand where the lack of program fidelity, or underperformance, occurred.

In other words, if researchers or evaluators can discover which part(s) of a program were underperforming, or malfunctioning, so to speak, then program managers might have a chance to remedy the issue(s). If those problems are clearly identified and convincingly articulated to stakeholders, the integrity demonstrated by such insightful evaluation and reporting, *in combination with a feasible strategy to remedy the issue(s)*, is a footing upon which an organization may build a case for follow-on work and commensurate funding.

Additionally, to incentivize the discovery of insights that are endemic to process evaluations, funders should consider requiring process evaluations: either as a stand-alone

evaluation type for a given funded project, or as an additional sub-type of evaluation, embedded into another, overarching evaluation type (e.g., an outcome evaluation). Funders may consider such a requirement as an insurance policy on their P/CVE investments: that even if a given funded project produces null (or deleterious) results, the funder may still be able to learn—hopefully in fine detail—what weak link(s) in the program produced those results. Such crucial information can be considered due diligence, on behalf of the funder, with respect to responsible, evidence-based decision making about whether, or how, to correct any such issues in future programming.

Conclusion

The aforementioned potential "saving grace" of process evaluations underscores the importance of knowing not only what outcomes a given P/CVE program produces, but *how* those outcomes are produced: specifically, the extent to which they derive from the program's theory of change. As mentioned, this theoretical understanding of *why* a program is (or is not) working as well as expected is the backbone of evidence-based P/CVE program design and evaluation, and is essential to informing sound P/CVE program management decision-making. Consequently, without exception, resolve to integrate process evaluation into each of your P/CVE-related research or evaluation projects. Doing so, you will advance not only the scientific interests of the fields of P/CVE, but the interests of your key stakeholders who deserve to have a well-informed understanding of the underlying processes and mechanism of their programs: to guide their programmatic or policy-relevant decision-making toward preventing or countering violent extremism.

References

Brewer, M. B., & Crano, W. D. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 3–16). Cambridge University Press.

Colman, A. M. (2015a). Demand characteristics. In *Dictionary of psychology*. Oxford University Press.

Colman, A. M. (2015b). Experimenter expectancy effect. In *Dictionary of psychology*. Oxford University Press.

Ghoshal, S., & Moran, P. (1996). Bad for practice: A critique of the transaction cost theory. *Academy of Management Review*, *21*(1), 13–47. https://doi.org/10.5465/AMR.1996.9602161563

Hoewe, J. (2017). Manipulation check. *The International Encyclopedia of Communication Research Methods*, 1–5. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118901731.iecrm0135

Koehler, D., & Fiebig, V. (2019). Knowing what to do: Academic and practitioner understanding of how to counter violent radicalization. *Perspectives on Terrorism*, *13*(3), 44–62.

Koehler, D. (2017a). How and why we should take deradicalization seriously. In *Nature Human Behaviour* (Vol. 1, Issue 6). https://doi.org/10.1038/s41562-017-0095

Koehler, D. (2017b). *Structural quality standards for work to intervene with and counter violent extremism: A handbook for practitioners, state coordination units and civil society programme implementers in Germany*. Counter Extremism Network Coordination Unit (KPEBW). https://d-nb.info/112503100X/34

Lewicki, R. J. (2006). Trust, trust development, and trust repair. In M. Deutsch, P. Coleman, & E. Marcus (Eds.), *The handbook of conflict resolution: Theory and practice (2nd Ed).* (2nd ed., pp. 92–119). Jossey-Bass.

Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, *8*, 183–189. https://doi.org/10.1002/bs.3830080302

Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports*, *19*, 115–118. https://doi.org/10.2466/pr0.1966.19.1.115

Sagan, C. (2011). *The demon-haunted world: Science as a candle in the dark*. Ballantine Books.

Schumpe, B. M., Bélanger, J. J., Dugas, M., Erb, H. P., & Kruglanski, A. W. (2018).

Counterfinality: On the increased perceived instrumentality of means to a goal. *Frontiers in Psychology*, *9*(JUL), 1–15. https://doi.org/10.3389/fpsyg.2018.01052

Steele, C. (2010). *Whistling Vivaldi : how stereotypes affect us and what we can do.* New York : W.W. Norton & Company.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation*. Wiley.

Williams, K. D. (2009). Ostracism : A temporal need-threat model. *Advances in Experimental Social Psychology*, *41*, 275–314. https://doi.org/10.1016/S0065-2601(08)00406-1

JOURNAL FOR DERADICALIZATION

Spring 2022
No. 30
ISSN: 2363-9849

About the JD Journal for Deradicalization

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

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

Twitter: @JD_JournalDerad
Facebook: www.facebook.com/deradicalisation

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

ISSN: 2363-9849