Journal of MultiDisciplinary Evaluation
Volume 12, Issue 27, 2016

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Adaptation and Use of a Five-Task Model for Evaluability Assessment

Brad R. Watts
Evaluation Center, Western Michigan University

Heather M. Washington
West Virginia University

Background: Over four decades have passed since the concept of evaluability was introduced; however, the availability and accessibility of methodology, frameworks, checklists and other guidance on evaluability assessments remains limited (Smith, 2005). Evaluators who wish to conduct an evaluability assessment must adopt one of few existing models or operate without guidance. This case provides an example of one model and how it was utilized to conduct an evaluability assessment of an ongoing program intervention.

Purpose: This article provides a real-world example of how an evaluability assessment was conducted using a five-task model from the criminal justice field.

Setting: A criminal justice program operating with at-risk adolescent youth and police officers who patrol their neighborhoods.

Intervention: NA

Research Design: NA

Data Collection and Analysis: The original evaluability assessment collected data through interviews with program personnel, observations by the evaluability researchers, and analysis of survey data provided by the program (i.e., program participants' responses to survey questions about the program).

Findings: The evaluability assessment conducted in this case study example identified multiple changes necessary for the program to be evaluable. Technical assistance was provided in order to support the efforts of the program to prepare for evaluation, though, the evaluability assessment identified risks and costs for the program to consider prior to proceeding. The findings were developed through a five-task process, which evaluators may be able to adapt for use with other types of programs undergoing evaluability assessment.

Keywords: *evaluability assessment; program evaluation; criminal justice.*

The purpose of this article is to introduce to a broader audience of evaluators a model of evaluability assessment (EA) from the criminal justice program field and to provide a real life example of how this EA model was implemented. We document our direct experience of using the aforementioned model outline to conduct an EA for a multi-site program intervention involving at-risk youth and police officers (henceforth "the program"). The real-world example of implantation provides an opportunity to explain how we followed the EA model, along with highlighting potential pros and cons to the approach. Additionally, although the example EA was conducted with a program that fits within the field of criminal justice programs, we also discuss the potential for adoption of the approach for evaluators working in other fields where a general framework for conducting an EA is desired.

Background

Approximately four decades have passed since evaluability assessment (EA) was first formally conceptualized (Wholey, 1979); however, during the intervening years the availability of methodology, frameworks, checklists or other guidance for conducting EAs has been surprisingly limited (Smith, 2005). Additionally, many of the methods and guidelines published on evaluability assessment have been associated with a specific field (Smith, 2005) such as agriculture (e.g. Smith, 1989), government (e.g. Wholey, 1979), or criminal justice (e.g. Welsh, Harris, & Jenkins, 1996), which might limit the recognition and distribution amongst the wider field of evaluation. Other evaluability models have been developed for or by specific organizations (e.g. UN Women, 2015) and may not be seen as readily transferrable to other programs.

Although some additional publications on conducting evaluability assessments have been released in recent years (see for example Davies, 2013; Zandniapour, 2014; Peersman, Guijt, & Pasanen, 2015) these continue to face many of the limitations in distribution, recognition, and access that have hampered previous efforts. As such there remains no formal set of evaluability assessment models, guidelines, or checklists that is widely accepted across the field of evaluation. For example, one well-known source for evaluation materials is the listing of checklists on the website of The Evaluation Center (n.d.), which still has no documents specifically dedicated to conducting an evaluability assessment.

The current situation is that evaluators have only limited choices when deciding how to conduct an evaluability assessment of a program. Options include:

1. Adopt one of the few existing models and adapt it to the program that is being assessed.
2. Refer to a textbook or similar source for a general overview of evaluability.
3. Proceed without any outside assistance.

When faced with the prospect of conducting an evaluability assessment, the authors choose to adopt an existing model to serve as a guide for the process. The model that was selected fit with the theme and subject matter of the program being evaluated (i.e. a criminal justice field intervention); however, it was found that the structure and guidelines were not necessarily specific to the topic area and could potentially be adopted for use with a range of different programs. This article introduces this model to a broad evaluation audience and presents a case example of the specific steps and actions used to complete one evaluability study.

Project Overview

In 2012, we were approached by a program that was interested in undergoing an evaluation as part of an effort to improve practice, promote the program, and work toward a long-term goal of becoming an evidence-based practice. The program is a training-style intervention that targets "at-risk" teenage youth and the police officers from the neighborhoods where the youth live and attend school. Youth participants are offered training on how to communicate and interact with police officers and are then participate in a personal interaction with police officers that is intended to be positive. This interaction includes sharing personal histories, engaging in team-building exercises, and encouraging open questioning of each other about common behaviors or practices. The theory put forth by the program is that developing a positive and personal level of interaction between youth and officers will lead to fewer problems between both parties in the future.

Program activities typically take place over the course of a two-week period and consist of

six or seven sessions conducted during the after-school hours. The location can be any convenient public space, but is usually a school or community center located near the neighborhood where the youth participants live. Youth participants are selected from individuals ages 12-19 who have been identified as being high-risk because of residency in a high-crime or high-poverty area, or because of early minor interactions with the juvenile justice system (but who generally have not yet faced a serious arrest or charge, and have not spent time in juvenile detention). The police officers who participate in the program are predominantly front-line officers who patrol the neighborhoods where the youth participants reside.

The sessions are led by a professional facilitator and mediator who is employed by the program, although in some locations one or more individual community members have received training and certification to lead sessions as well. Financial support for operating the program, along with finding a host site location and gaining the involvement of the local police department occurs through the development of local community partnerships. Additionally, the local community partnership that brings in the program also helps to identify and recruit youth to the program from sources such as local schools, a neighborhood community center, or nonprofit agencies that provide at-risk youth with after-school programs and/or other interventions.

During the first two-to-three sessions of the program, only the youth participants attend and the focus is on discussing neighborhood issues, teaching communication skills, and preparing the youth to make an initial oral presentation to the police officers who will be attending the later sessions. The oral presentation consists of a brief talk given by the youth about who they are, their family, and why they are taking part in the program. These presentations from the youth are followed by brief presentations made by the officers, and are intended to be a way of breaking down barriers and highlighting some of the background similarities that may exist between the officers and youth.

During the fourth session, the police officers join the training with the youth and share the aforementioned personal life stories. The final training session consists of one-on-one activities between the youth and officers, team-building exercises, simulation training to demonstrate what happens during a traffic stop or group situation involving youth and police, and group discussion led by the program facilitator. Finally, at the end of the training there is a celebratory session where both the youth and police participants share a meal together (along with youths' family members and community stakeholders) and certificates of completion are awarded.

Why Evaluability Assessment?

The program originally expressed interest in undergoing a summative evaluation of its outcomes. A full, summative evaluation was thought by the program to offer several specific benefits, such as providing an independent measure of the program's impact, the creation of a report on outcomes that could be publicized to communities considering the program, and possible attainment of an evidence-based "model program" designation from The Office of Juvenile Justice and Delinquency Prevention's Model Programs Guide (Office of Justice Programs, n.d.). Additionally, a full evaluation seemed reasonable given the length of program operations; at the time the study was initiated the program had operated in 20 different communities, mostly in the northeastern U.S., over a period of more than 10 years. As an established intervention, the program's staff initially felt that the program should be ready to undergo evaluation; however, several early signs led us to suggest pursuit of an evaluability assessment before investing in a full, summative evaluation:

1. The original program logic model indicated numerous, broad goals that would likely be difficult to measure.
2. Prior research studies conducted on the program illustrated substantial methodological flaws. Two early, site-level studies failed to isolate program findings from significant external factors, including outside program effects and macro-level crime trends.
3. The program built and utilized its own approach and was not based a specific, empirically-tested model.

These early concerns regarding the readiness of the program for evaluation led us to suggest that an evaluability assessment be conducted in order to 1) determine whether or not the program was ready, and 2) identify any issues preventing a thorough evaluation and provide technical assistance so as to prepare the program for a full summative evaluation.

Program leaders agreed to the approach and grant funding was sought in order to support an evaluability assessment. Ultimately, funding for the evaluability assessment was provided by an award from the U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Selection of an Evaluability Assessment Model

In order to propose and successfully conduct our study, we first sought to identify and compare existing guidelines and models for evaluability assessment. As mentioned previously, the availability of accessible, generalized models for conducting evaluability assessments is somewhat limited, although it should be acknowledged that multiple tools have been developed since our study was originally conducted (e.g. Davies, 2013; Peersman, Guijt, & Pasanen, 2015; UN Women, 2015; Zandniapour, 2014). Our process began with a general search for evaluability assessment models, guidelines, checklists, and other documentation. The initial search identified several evaluability assessment guides for consideration, including textbooks from Wholey (2004) and Smith (1989), published guidelines and checklists (Trevisan & Huang, 2003; International Development Research Centre, 1996), and one model developed specifically for criminal justice programs (Kaufman-Levy & Poulin, 2003).

After reviewing the aforementioned options, the model developed for criminal justice programs (Kaufman-Levy & Poulin, 2003) was selected for our evaluability assessment project. Although the model was clearly created for use with criminal justice programs, the model was ultimately selected because it appeared to provide a concise set of tasks that could be used to guide our evaluability assessment project. These tasks include activities that were believed to offer potential value, specifically conducting observations to understand the intervention and examination of the capacity—both issues that seemed at the outset to be pertinent to the program undergoing assessment. The five tasks are fully explained in the next section.

Application of the Five-Task Evaluation Model

The model we selected for our evaluability assessment centers on five tasks that must be completed to determine whether or not a program is ready to undergo evaluation. These tasks are as follows:

1. Study the program history, design, and operation
2. Watch the program in action
3. Determine the program's capacity for data collection, management, and analysis
4. Assess the likelihood that the program will reach its goals and objectives
5. Show why an evaluation will or will not help the program and its stakeholders (Kaufman-Levy & Poulin, 2003 p.10)

The documentation of the five-task model provides some guidance on how to conduct each of these tasks, but it does not provide a checklist or any guidance on the synthesis of the findings from each task into an evaluative conclusion regarding the overall evaluability of the program being assessed. For each of the five tasks, we provide an example of how we applied the model to conduct a real world evaluability assessment. We also identify potential "pros" and "cons" to the approach and provide discussion on applying the five-task model as a general tool for evaluability assessment.

Task 1: Study the Program History, Design, and Operation

The first task focuses on developing an understanding of the program being assessed and building a relationship with program stakeholders. According to the guidelines in the model, this step should generate the answers to three key questions: 1) what is the program's history; 2) what is the program's design; and 3) how does the program actually operate? (Kaufman-Levy & Poulin, 2003 p.11). The task description does not prescribe specific activities or criteria for answering the questions.

To fulfil the first task, we conducted several activities. First, we conducted interviews with key program staff: the program founder and the director of the non-profit organization that administers the program. These interviews provided an oral history of the program and also identified the theoretical background (i.e. police legitimacy theory and conflict resolution theory) that were believed to support the program intervention that had been developed. Next, we reviewed records on the program; these materials included the official information

provided on the program's website, promotional materials, a program logic model, the program's training manual, site-level program activity records, and two previous studies of the program. From the interviews and document reviews, we compiled a matrix outlining the program's operations on a site-by-site basis over-time.

In addition to collecting information directly from the program, we also conducted a literature review in order to identify how the intervention approach described by program staff fits with actual criminal justice theory. The review initially focused on explaining the theoretical basis for the program through major works in the criminal justice research literature on police legitimacy (for example see Tyler, 2004; Tyler & Fagan 2008) and conflict resolution theory as applied to group dynamics (for example see McEvoy & Newburn, 2014; Staub, 2014); however, these theories did not wholly match with the program as described by official documents and program staff (or as discussed in the next section, actual program observations). For example, the program lacked the formal or legal approach associated with group-level conflict resolution theory (see Menkel-Meadow, 2014) and the program participants had generally not experienced direct conflict with one-another in the past. The program also did not appear to heavily focus on the main ideas behind police legitimacy theory, which views law adherence and police interactions as being a function of an individual's assessment of the correctness of the law and its enforcers and their own experiences with the procedural legal system (Tyler, 1990).

The literature review portion of the first task instead became an opportunity to search for and identify theory and research more directly aligned with the actual activities and goals of the program. Ultimately, we found that program appeared to be more aligned with existing approaches that are based on attitudinal research, which finds that attitudes are formed largely from direct experiences, as well as surrounding social and cultural influences (Leiber, Nalla, & Farnworth, 1998; Nihart, Lersch, Sellers & Mieczkowski, 2005). The attitudinal approach found in the literature provides a better alignment with actual program activities, which focus on communications and creating a positive, personal interaction between youth and police officers. Also, evaluation studies on similar programs were also identified, which utilized as outcomes measures of youth attitudes toward police or attitudes of police toward youth (for example, Hopkins et al., 1992; LaMotte et al., 2010). These studies provided evidence regarding the theory and outcomes associated with similar programmatic approaches, as well as providing insight into the attainability of goals.

Task 2: Observation of the Program in Action

The second task is to observe the program in operation, which provides an opportunity to confirm what the program does and whether the reality of the program aligns with the previously-expressed design and intent. To complete this task, we observed two separate program sessions. First, we attended all sessions of one complete class of the program as passive observers. The program operated on six evenings over a two-week period and was hosted by a non-profit school that provides programming and support services for at-risk youth and their families. The location for the program observations was a mid-sized city in the Northeastern region of the U.S. Second, because the program typically operates only single-gender classes (predominantly male) we also attended and observed several sessions of the program held with female youth and hosted at another non-profit agency.

In order to more systematically determine whether or not the actual program operates as expected, we developed a list of seven key program "elements" based on what key staff and official documents had indicated should be consistent traits of the program. These basic expected program traits included 1) who the program serves; 2) the nature of the partners that host and support the program; and 3) program facilitation, activities, and general topic areas. Table 1 provides an example of the expected program traits that were looked for during the observations and how adherence was assessed. In general, the observation stage served to confirm that the program operates as expected based on the description provided by official program materials and during interviews with key staff.

Table 1
Example of Criteria Examined During Observation Stage

Expected program trait	Observed?	Evidence in observation
Program serves "at-risk" youth	Yes	Youth were part of a reporting center diversionary program; participants indicated they lived in inner city neighborhoods or low-income rural communities
Police participants are front-line; from the neighborhoods where youth reside	Yes (mostly)	With exception of one lieutenant, all were patrol officers or sergeants who worked car or bike routes in city neighborhoods
Community partners engaged, provide support for program and follow-up	Yes	Local community organization hosted the sessions; volunteer trainees attended to become certified to support future sessions; police department and city offered funding
Process develops leadership, communication skills	Yes	Youth practiced creating and telling life stories; observed change in group behavior during course of program sessions
Develop bonds between police and youth	Yes	Visible shift in relations over course of three nights; started with life stories discussion, worked up to direct interactions and scenarios; youth and police dined together on final night

A key element of the observation task in our example was that it provided a hands-on way of examining implementation fidelity. While program documentation and conversations with program leaders can provide a depth of information about program operations, firsthand observation provides an opportunity to confirm the model fits the official description. Additionally, observation provides an additional opportunity to look for program traits that could be challenging to a future evaluation. In our example, this included seeing how the pre- and post-training surveys are distributed and collected by the program's facilitator, as well as getting a sense of what the participants are like and how they respond to the program during delivery. This contributed to our later recommendations to add a follow-up survey and work on how data forms are collected, organized, and stored.

Task 3: Program Data Capacity

The third task is to determine whether or not the program has the ability to successfully collect and maintain the data necessary to measure participant traits and outcomes for a future evaluation. To complete this task, three major aspects of current data capacity were examined: 1) the program's current system and practices for collecting data; 2) the actual data variables that the program collects; and 3) the consistency and use of the data by program staff. We began by working with the program to obtain all available information that had previously been collected on program participants, which in this case included pre- and post-training surveys, satisfaction surveys, and site-level analyses that the program had provided to communities that had hosted the program. Additionally, program staff were queried regarding data collection practices (i.e. who collected the data, where it was stored, who used the data and how), as well as the staffing and capabilities for data collection in the non-profit organization that operates the program.

Upon collecting the data, some capacity concerns were immediately identified. For one, nearly all of the program data had been not only collected but also maintained in hard-copy paper format. This included pre- and post-test instruments for police and youth participants, as well as some surveys, and participant demographic information. Although some data had been entered into spreadsheets that were used to conduct small site-level analyses, there was no way to readily examine and analyze the data that had been captured to look at any potential program-level effects. Additionally, a closer examination of the paper forms also revealed that the program had made changes to the survey forms over time, such as the removal

and addition of questions, as well as changes in question wording and the rating scales.

In response to the problems identified with the program's paper files, we created a database to assist the program with data collection and to provide an example of how data can be compiled and stored electronically. The new database was then populated with data already collected by the program, as well as with new data collected from sessions that concluded during the evaluability study. This database then served as both the basis for analyses conducted for the EA and as a deliverable for the program staff that can be used to support future data collection efforts.

The process of collecting and examining the data highlighted some issues with the program's past data collection practices (i.e., changes to the survey questions and rating scales), which ultimately hindered us from drawing strong conclusions about the program's effectiveness and capacity to reach its goals. Still, the analysis was able to draw some findings from a partial set of surveys that contained information that could be matched (i.e. the same questions using the same scales). Also, the data proved to be useful for other purposes such as tracking the number of youth and police officers who have participated in the program, noting where sessions were conducted, and calculating how many sessions were offered in each city. However, the situation also prompted us to recommend to the program that they discontinue some prior uses of the data, such as providing site-level analyses for some communities, since the data suffered from problems with consistency and attempted to draw conclusions across items and scales that were not always comparable.

With these findings in mind, we compiled a list of data collection strengths and weaknesses for review by the program's administrators. The focus was on identifying specific aspects of the data collection process that needed to be changed if the program were to successfully track their outputs and outcomes for a summative evaluation and for broader dissemination of program effectiveness to targeted communities. We also provided the program with basic technical assistance with using the newly constructed database and with instituting changes to the program's survey instruments. Table 2 illustrates the strengths and weaknesses that were identified and the collaborative responses we developed with the program to prepare them to be ready to collect data for a future evaluation.

Table 2
Data Collection Strengths and Weaknesses and Resulting Action or Recommendation

Trait	Change or Response
Strengths	
Program has interest in evaluation, has created forms and collected data in the past	Program to continue direct data collection
Larger nonprofit organization has some data capacity, intern assistance	Database developed by evaluator for use by admin staff at agency
Has developed strong working relationships with local police agencies in past	Opens possibility for collecting follow-up data from participants
Weaknesses	
Data collection was messy; no electronic system	Database developed by evaluator for use by admin staff at agency
Site-level analyses conducted by intern in past contained errors	Provided technical assistance to intern on data analysis; recommended stopping some reporting practices for sites based on limited numbers
Data forms and collection were allowed to vary across time	Creation of a new, standard forms for police and youth participants for pre/post/follow-up periods
Old pre/post survey items not always aligned with goals and outcomes	Identified instruments from literature; created new pre/post/follow-up data collection forms
Data collection does not capture medium or long-term outcomes	Recommended follow-up to occur 3-6 months following program participation

During our evaluability study, the general data capacity of the organization was strong overall, but the specific weaknesses that were identified would have severely hampered efforts to conduct a summative evaluation. While some organizations may face building a data collection process from scratch, the program in our study believed that a sufficient system was in place, only to find that significant improvements were necessary to ensure that the date would be useful for its likely purpose in an evaluation. Resulting recommendations included adding a follow-up survey, changing the questionnaire items to better fit with desired outcome measures, and improving the general organizational structure of the organization's data collection efforts.

Task 4: Likelihood of Program Attaining Goals

The fourth evaluability assessment task is to determine whether or not the program has set reasonable goals that can be attained by the intervention. If the program's goals are not theoretically tied to the activities of the intervention, or if the magnitude of the program goals and activities is not aligned, the results of an evaluation are sure to be disappointing. Additionally, there may be early signs of success or failure—such as anecdotal success cases or obvious implementation problems—that could also provide early evidence of whether or not a program is on its way to successfully attaining its goals.

For our study, we took two approaches to assessing the potential for goal attainment. First, a literature review was conducted to identify existing research and evaluation studies that had been completed for similar types of program interventions. The goals from these studies were then compared with the program in order to identify overlapping goals and to find examples of how outcomes related to those goals were measured. This step also resulted in the identification of goals that had been successfully

attained for other programs, based on the presentation of evidence that outcomes associated with the goals were positively impacted by the similar intervention. The result was evidence regarding attainability, wherein goals that were successfully measured and that showed outcomes for similar programs were considered to have a high likelihood of attainability, while goals that other studies dismissed , found to be unmeasurable or for which they could not find significant results were considered to have a lower likelihood of attainability.

Second, the program's existing data were analyzed to determine whether there was preliminary evidence of success in attaining outcomes associated with major program goals. Although those data had significant limitations, as was identified during the assessment of program data capacity step, the pre-existing data also provided a source of unique data from actual program participants. Using the stated goals of the program as a guideline, we identified a subset of survey questions that were intended to measure outcomes related to the behaviors and attitudes of participants. These were subsequently narrowed down to those questions that had been asked consistently (in terms of wording and format) and collected over multiple training sessions.

To test for preliminary evidence of change in attitudes and behaviors, we analyzed the change in ratings between pre-training and post-training surveys that were completed by the youth and the police participants. Because the individual survey items and rating scales had changed somewhat over the course of the program's implementation, it was not possible to analyze every question. Still, the analysis provided some indication of where the program might be likely to show evidence of measurable change in a future evaluation. For example, on 12 items matched across the youth pre- and post-training surveys there was evidence of a statistically significant and positive change in the ratings immediately following the program.[1] However, the analysis of the pre- and post-training surveys of police officers who had participated in the program found a significant and positive change pattern on only three of 14 items.[2]

The results of the analysis of previously-collected data was informative not as a source of evidence of program impact, but because it helped to identify what outcomes are more or less likely to have measureable results in a future evaluation. For example:

- Youth ratings showed significant and positive change between pre and post surveys. This suggests a reasonable likelihood that the program could be having at least a short-term impact on youth outcomes.
- The weak and limited findings from the analysis of the pre/post surveys completed by the police participants suggests that it is less likely that the program is having a significant short-term impact on this group.
- The variation in the number of valid items and surveys that could be matched for both youth and police participants in the past supports a low likelihood of obtaining the data necessary for a rigorous evaluation without improvements to data collection and storage capabilities.

The findings from the analysis of existing data would likely not be strong enough to provide sufficient evidence of the program's impact on the major desired outcomes; however, in the context of an evaluability study, the data provide some support for the likelihood of goal attainment. As a result, it was suggested to the program that its intermediate goals of shifting youth attitudes appeared to be attainable, while influencing police attitudes may be more challenging. Changes were recommended for strengthening data collection on both youth and police, including improvements to the items on the pre- and post-training surveys, as well as the addition of a follow-up survey.

Task 5: Why an Evaluation will or will not Help Stakeholders

The final task of evaluability is to make a determination of whether or not an evaluation will provide value to the wide range of stakeholders of the program being evaluated. Ultimately, the results should be useful to multiple program stakeholders regardless of the ultimate findings. To assess the value or "helpfulness" of a potential evaluation of this program, we looked at how the findings could be used by three major stakeholder groups: the program itself, the communities that host the program, and the larger field of juvenile justice. In addition to making an overall determination of evaluability and providing technical support for the program, our evaluability

[1] Based on the results of a Wilcoxon signed-rank test for differences in distribution, $p < 0.05$ and individual item *N*s ranging from 43 to 140.

[2] Based on the results of a Wilcoxon signed-rank test for differences in distribution, $p < 0.05$ and individual item *N*s ranging from 12 to 42.

study also noted the risks and rewards facing each of the stakeholder groups. For example,

- The program can potentially benefit from an evaluation that provides evidence to help market its activities to a wider range of communities. The downside is that preparing for an evaluation will require increased costs and staff effort, as well as the risk of negative or inconclusive findings.
- The communities that use and support the program benefit from general evidence about the efficacy of the program, but may also desire more localized information, which a summative evaluation will not provide on its own.
- Those who are interested in juvenile justice will be interested in the results of an evaluation if it provides evidence supporting the impact of a model that is replicable to a broad range of youth and police officers in communities across the country. However, although the program can be evaluable with some reasonable changes, it may still be difficult to provide evidence meeting the highest standards of the National Institute of Justice's evidence-based rating system. [3] An evaluation that satisfies programmatic and community-level stakeholders may fall short of the standards of academic and governmental stakeholders without additional improvements.

The focus of the fifth task is on delivering the results and proving that an evaluation is worthwhile, which is a process that could be handled in many different ways depending on the nature of the evaluability study and the stakeholders. In our example, the venue for delivering results was a formal report document, which was required by the grant funding which supported the evaluability study. However, the potential value of an evaluation, along with how to become more prepared for an evaluation was regularly communicated prior to this final task. Related activities included sharing early findings, providing technical assistance and support (such as helping to revise the logic model and design new forms), and discussing what kinds of results might and might not be reasonable to expect from a summative evaluation.

[3] See https://www.crimesolutions.gov/

Discussion

At the time the example study was conducted in 2013, the "five-task" approach (i.e. Kaufman-Levy & Poulin, 2003) was selected because of its accessibility (the article describing the approach is freely available on-line for anyone to access), its relevance to the client (the development of the five-task approach was part of a study funded by the U.S. Department of Justice), and the simplicity of the guidelines. Additionally, the five-tasks focused on specific activities, while still providing flexibility on how each should be conducted. For the authors, who had not previously conducted a formal evaluability assessment, the five-tasks provided a concrete set of actions to take, which provided a starting point for the process and ensured that some major aspects of evaluability were examined.

Although the five-task approach successfully provided a framework for our own evaluability study, it also serves to highlight the gaps in support for evaluability assessment. The choices for formal guidance on conducting an evaluability assessment remain limited (despite some recent introductions) and there remains no clear, preferred source for evaluability resources. While the five-task approach is an option worth considering, there remains room in the field for developing and improving a model or models of evaluability, as well as checklists and other hands-on tools, that would be both widely accessible and applicable to a wide range of program types and settings.

We hope that our work will serve as a real life example of an evaluability assessment that is helpful to researchers seeking to conduct their own evaluability study. While useful, the tasks outlined by this framework did not prepare us for some of the challenges we met along the way. As such, it is important to note some these challenges here. First, when gathering information about the program history, design, and operation (Task 1), we were not prepared for the fact that much of the program history consisted of oral narratives. With little written down, it was difficult to construct a complete history of the program. Second, when completing Task 2, we were surprised to find that, while the curriculum of the program remained consistent, the way in which the information was gathered and the topics for discussion arose organically over the course of each two week session. As such, it was difficult to compare one rendition of the program to another. Third, we found that the programs goals needed to be

revised in such a way that would allow for data collection and measurement (Tasks 3 and 4).

Despite these challenges, we believe we were successful in applying this framework to a criminal justice program. This framework could be also be used for evaluability assessments in other fields such as psychology, social work, and medicine. As options for evaluability slowly expand, those who conduct evaluability assessments should not have to find themselves recreating the wheel, but should look to this five-task model, along with other options, as a source of accessible guidance.

Acknowledgements

The evaluability study discussed in this article was financially supported by a grant from the U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention (2013-PB-FX-0004). The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

References

Davies, R. (2013). Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations. Report of a Study Commissioned by the Department for International Development. Retrieved on February 22, 2016 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf

Hopkins, N. (1994). School pupils' perceptions of the police that visit schools: Not all police are pigs. *Journal of Community and Applied Psychology*, *4*, 189-207.

International Development Research Centre. (1996). Guidelines for Preparing an Evaluation Assessment. Retrieved February 22, 2016 from http://web.idrc.ca/en/ev-29113-201-1-DO_TOPIC.html

Kaufman-Levy, D. & Poulin, M. (2003). *Evaluability Assessment: Examining the Readiness of a Program for Evaluation*. Washington, DC: Justice Research and Statistics Association. Retrieved from http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf

LaMotte, V., Ouellette, K., Sanderson, J., Anderson, S.A., Kosutic, I., Griggs, J., et al. (2010). Effective police interactions with youth: A program evaluation. *Police Quarterly*, *13*, 161-179.

Leiber, M., Nalla, M. & Farnworth, M. (1998). Explaining juveniles' attitudes toward the police. *Justice Quarterly*, *15*, 151-174.

McEvoy, K. & Newburn, T. (Eds.) (2014). *Criminology, Conflict Resolution and Restorative Justice*. London: Palgrave Macmillan.

Menkel-Meadow, C. (2014). Alternative and appropriate dispute resolution in context. In Coleman, P.T., Deutsch, M., & Marcus, E.C. (Eds.) *Handbook of Conflict Resolution (3rd Ed.)* (pp.971-997). Hoboken, NJ: John Wiley & Sons, Incorporated.

Nihart, T., Lersch, K., Sellers, C. & Mieczkowski, T. (2005). Kids, cops, parents, and teachers: Exploring juvenile attitudes toward authority figures. *Western Criminology Review*, 6, 79-88.

Office of Justice Programs. (n.d.). OJJDP Model Programs Guide [website]. Available at http://www.ojjdp.gov/mpg/

Peersman, G., Guijt, I., and Pasanen, T. (2015). *Evaluability Assessment for Impact Evaluation*. A Methods Lab publication. London: Overseas Development Institute.

Smith, M.F. (1989). *Evaluability Assessment: A practical approach*. New York: Springer.

Smith, M. F. (2005). Evaluability assessment. In Mathison, S. (Ed.), *Encyclopedia of Evaluation* (pp. 136-139). Thousand Oaks, CA: Sage.

Staub, E. (2014). Reconciliation between Groups In Coleman, P.T., Deutsch, M., & Marcus, E.C. (Eds.) *Handbook of Conflict Resolution (3rd Ed.)* (pp.971-997). Hoboken, NJ: John Wiley & Sons, Incorporated.

The Evaluation Center. (n.d.). Evaluation Checklists [website]. Available at http://www.wmich.edu/evaluation/checklists

Trevisan, M.S. & Huang, Y.M. (2003). Evaluability assessment: a primer. Practical Assessment, Research & Evaluation, 8(20). Available from http://PAREonline.net/getvn.asp?v=8&n=20

Tyler, T.R. (1990). *Why People Obey the Law*. New Haven: Yale University Press.

Tyler, T.R. (2004). Enhancing police legitimacy. *The Annals of the American Academy of Political and Social Science*, *593*, 84-99.

Tyler, T.R. & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities? *Ohio State Journal of Crime*, 231-275.

UN Women. (2015). *How to Manage Gender-Responsive Evaluation: Evaluation*

Handbook. New York: United Nations Women Independent Evaluation Office.

Welsh, W.N., Harris, P.W., & Jenkins, P.H. (1996). Reducing overrepresentation of minorities in juvenile justice: Development of community-based programs in Pennsylvania. *Crime and Delinquency*, *42*(1), 76-98.

Wholey, J.S. (1979). *Evaluation: Promise and performance*. Washington, DC: The Urban Institute.

Wholey, J.S. (2004). Evaluability assessment. In Wholey, J.S., Hatry, H.P., Newcomer, K.E. (Eds.), Handbook of Practical Program Evaluation (2nd ed.) (pp.33-62). San Francisco: Jossey Bass.

Zandniapour, L. (2014). *Impact Evaluability Assessment Tool*. Washington, DC: Corporation for National and Community Service. Retrieved February 22, 2016 from http://www.nationalservice.gov/sites/default/files/resource/SIF_Impact_Evaluability_Assessment_Tool_Final_Draft_for_Distribution.pdf