4.2 PRAGMATIC STRATEGIES & CONSIDERATIONS FOR EVALUATING MENTAL HEALTH PROGRAMS

Tyler Frederick

INTRODUCTION

Evaluation research involves gathering a wide variety of indicators in order to better understand how a program is operating, its impact, and what can be done to improve it. The performance indicators that many organizations collect routinely as part of their daily operations can provide important information for a program evaluation (e.g., attendance, client demographics), but are not themselves considered evaluation research. Evaluation involves systematically and intentionally collecting and reviewing information in order to understand and strengthen a program.

This chapter offers service providers guidance around evaluating programming within their organizations, with a particular focus on mental health initiatives. It discusses developing evaluation questions and choosing sources and methods for obtaining information. It also examines ethical considerations in conducting evaluation research.

COLLABORATION VERSUS GOING IT ALONE

Organizations can and should conduct their own evaluation research when possible. However, collaborating with external evaluation professionals can be extremely helpful, particularly if the aim is to conduct a complex evaluation or if there are no internal personnel to take on the task. Additionally, a third-party evaluator can bring a valuable outside perspective to the evaluation and help address ethical issues that can arise when a program asks its own staff and clients to participate in research. The two main options for getting expert guidance are to hire a consulting firm or solicit the help of a local university.

The main advantage to hiring a consulting firm is the quick turnaround time; the main drawback is the cost. Universities, in comparison, are a good low-cost option, but turnaround time can be slower because researchers at a university will have other things competing for their time. For organizations that want to approach a university, a good

first stop is a research partnerships office. Most large universities have a department that facilitates collaborations between the university and outside organizations and companies. If there is no such office, then the best approach is to look at faculty profiles within departments such as social work, psychology, sociology, public health, and medicine. Organizations should look for researchers whose research interests overlap with the population and focus of the initiative.

Simple projects can likely be carried out on a small budget, with student volunteers and the faculty member providing time in-kind. Projects that are more complex might require writing a grant to solicit funds. Organizations working with a university should take the time to draft an agreement to clarify roles and responsibilities and to determine ownership of the data. University researchers are usually motivated to participate in evaluation research so they can use the data for their own research and publishing opportunities. University partnerships offices can be helpful in drafting these types of agreements.

This chapter can inform discussions with a consultant or university researcher, and will also be useful to organizations that want to conduct their own research.

TYPES OF EVALUATION

There are six main types of evaluation research:

- Needs assessment: to understand the characteristics and needs of a client population and to identify current gaps in service for the purposes of designing new programs;
- Monitoring review / compliance with standards: to ensure that a program complies with the requirements of governments or funders and to ensure that a program is showing fidelity with a chosen implementation model;
- Implementation/process evaluation: to understand how a program is operating and to identify issues and challenges with its structure and operation;
- Impact/outcome evaluation: to understand the short-, medium-, and long-term impacts the program is having on those involved and to ensure that it is fulfilling its intended mandate;
- Program review: to update a program and make minor adjustments to ensure it is responsive to shifts in client needs, staffing, and budget priorities; and
- Efficiency assessment: to improve operational efficiency and reduce program costs.

Determining the appropriate type of evaluation is a key part of the process because it shapes the data that will be collected and analyzed. This chapter focuses on needs assessments, process evaluations, and impact evaluations. These are the most common types of evaluation and the steps for conducting them can be easily adapted to other types of evaluation.

LOGIC MODELS & DEVELOPING RESEARCH QUESTIONS

Once the type of evaluation to use has been decided, the next step is to figure out what questions the evaluation needs to answer. Questions developed at this stage are meant to be broad and guiding. Organizations should aim for two or three central questions and make sure they match the type of evaluation.

Logic models can be valuable in identifying these guiding research questions. They involve a process of determining, documenting, reviewing, and modifying the intended structure of a program (W. K. Kellogg Foundation, 2006). The main components of a logic model are:

- Resources/inputs: resources the program will use in fulfilling its goals;
- Activities: the main tools, structures, and processes of the program;
- Outputs: the direct product of the program's activities;
- Outcomes: specific changes the program aims to achieve among program participants. These outcomes are usefully divided into short-term, medium-term, and long-term outcomes; and
- Impact: broader and more fundamental changes that will occur within the organization, community, or relevant system due to the operation of the program.

Combining these components, the typical structure of a logic model looks like this:
Resources/inputs → activities → outputs → outcomes → impacts

Logic models are valuable for research because the different types of evaluation address questions relevant to specific components of the logic model.

NEEDS ASSESSMENTS

Needs assessments address questions at the beginning of a logic model. The goal is to understand the characteristics and needs of the population of interest. This knowledge can then be used to determine intended outcomes and design activities.

Example questions for a needs assessment:

- What are the demographic, social, and background characteristics of the population we serve or want to serve?
- What are the experiences (successes, challenges, barriers) of this population?
- What needs have experts identified among this population?
- What needs does this group identify for itself?

PROCESS EVALUATIONS

Process evaluations review a program's activities and are used to ensure they are operating as intended. They are also useful for identifying gaps and barriers in the processes through which program activities are being delivered.

Example questions for a process evaluation:

- How are participants experiencing the program activities and what is their level of satisfaction?
- How are staff members experiencing the program activities and what is their level of satisfaction?
- How easily are participants navigating their way through the intended structure of the program?
- What are the main barriers participants are experiencing to their full and successful participation in the program activities?
- Are participants progressing through the program in line with the intended time frame?

OUTCOME EVALUATIONS

Outcome evaluations are primarily concerned with determining whether participants are experiencing the intended outcomes and impacts of the program. It is important to decide what level of outcome the evaluation will target (i.e., short-, medium-, or long-term). In deciding this, it is necessary to ensure enough time has lapsed to gauge adequately whether the intended outcome has been achieved. For example, an initiative may need to run for one year or more before any type of long-term outcome can be effectively evaluated.

Example questions for an outcome evaluation:

- Is this program fulfilling its intended short-, medium-, and long-term outcomes?
- What impact is this program having on participants' mental health (short-, medium-, and long-term)?
- Are participants using the skills they learned in their daily lives?
- Do participants feel the program is adequately meeting their needs?

IDENTIFYING DATA SOURCES

Once the guiding research questions have been identified, the next step is to decide from whom and where the information for the evaluation will come. The following section identifies key sources of information and suggests possible methods to use for obtaining information from each type of source.

ACADEMIC EXPERTS

Getting access to the academic research on a particular population can be a valuable first step in understanding the social and demographic characteristics of a population and its needs, and for understanding common problem areas. A recommended way to gain this information is to reach out to a scholar who does research in the particular area of interest. Looking at the profiles of researchers at a local university can be a helpful start. Google Scholar (scholar.google.ca) is a useful tool for identifying people who do research in the area of interest. Email them to ask for help locating key pieces of information on the population of interest.

Organizations can also do their own searching through the Internet, but this can be unreliable because so much information is of unclear quality and providence. If you are conducting your own search, use Google Scholar or a trusted organizational website or clearinghouse (e.g., www.homelesshub.ca). On Google Scholar, include keywords about the population or topic, and use terms like "systematic review," "scoping review," or "meta-analysis," which will identify studies that review a large quantity of information on the subject. The online Cochrane Library can also be a useful place to find systematic reviews on various mental health topics (www.cochranelibrary.com/cochrane-database-of-systematic-reviews/index.html).
Possible methods: literature reviews, expert interviews

STAFF, PARTNERS, & STAKEHOLDERS

Staff can be a valuable source of information for process and outcome evaluations because they can indicate needs they are seeing in the community or challenges they are experiencing with a particular program. The method chosen for obtaining information is important because ethical issues can arise when staff members are asked to provide information and feedback (these considerations are discussed in a later section). Stakeholders, organizational partners, and funders can also be valuable sources of information because they can describe how a particular organization or program is perceived outside the organization. They can also discuss gaps in service, funding priorities, and best practices within a particular service sector.
Possible methods: interviews, focus groups, anonymous surveys, anonymous comment boxes

PERFORMANCE INDICATORS & INTERNAL STATISTICS

The statistics that organizations collect routinely as part of their daily operations can be useful for program evaluations. For example, attendance statistics can provide information on who is being served (and therefore who is not being served), how long clients are accessing services, and repeat clients. In a process evaluation, internal statistics are most informative when they are paired with other methods of obtaining information. For example, internal statistics could help an organization see that younger clients seem to be discharged more frequently than older clients. This finding can then be explored in more detail through surveys or interviews.
Possible methods: data analysis

CLIENTS & RESEARCH PARTICIPANTS

Clients are a key source of information for program evaluations because they are best positioned to understand how well a program or service is meeting their needs, and to identify program strengths and weaknesses. Ethical considerations around asking clients to participate in a program evaluation need to be addressed and are discussed in a later section. *Possible methods*: interviews, focus groups, case studies, anonymous surveys, anonymous comment boxes

CHOOSING AN EVALUATION METHOD

COMMENT BOXES

Comment boxes can be a simple, low-cost way to gather information on how a program is functioning. Furthermore, they are anonymous, which is helpful for soliciting honest feedback. However, where comment boxes are installed is important. For example, having a box at the front reception desk may discourage people from submitting feedback because they do not want to be seen filling out a comment card. Putting the box in a less busy area will address this issue. Another option is to create opportunities where all program participants submit a comment form, whether they complete it or not, so individual responses cannot be identified. It is important that the people being asked for feedback know that someone is reading the comments and taking steps to address them. For example, a newsletter or poster can summarize feedback and action, or anonymous feedback (retyped to protect anonymity) can be posted on a board with an accompanying response.

SURVEYS

Surveys are useful for collecting information on a set of specific questions. The questions are typically close-ended (yes/no or multiple choice). Surveys can include a few short open-ended questions, but they should be used sparingly because they can create survey fatigue.

Surveys can be used in needs assessments to gain a better understanding of the characteristics and needs of a client population (e.g., demographic information, top needs, service use). In process evaluations, surveys are useful for asking about specific qualities or characteristics of a program and for gauging satisfaction with particular components. For outcome evaluations, surveys can elicit information about how clients have been impacted by a particular initiative and can track change over time for key indicators (e.g., mental health, life satisfaction, hope, symptomology).

Constructing survey questions

Surveys should be as short as possible and should include no more than 50 questions. Most surveys include at least a few basic demographic questions (e.g., age, gender, sexuality). These questions make it possible to identify differences across groups. For small surveys (50 or fewer people), keep demographic questions to a minimum to protect anonymity.

For help developing background and demographic questions, consult the questionnaires used by Statistics Canada that are listed in the "Definitions, Data, Sources and Methods" section of its website (www.statcan.gc.ca/eng/concepts/index?HPA=1). It is possible to search by subject and to focus on surveys with youth. The link for each survey contains a PDF file of the survey questionnaire. The questions can be adapted for review purposes.

For outcome surveys, organizations may want to include scales or assessment tools that asses various components of mental health and well-being. Beidas et al. (2015) have put together an excellent list of brief, free, and validated assessment tools. Most of the tools they list are intended for screening purposes only, which means they can identify a potential mental health problem, but not provide a diagnosis. Other free, well-validated scales include the World Health Organization's (1997) Quality of Life scale, which is available through its website (look for the BREF version), and the GAIN set of appraisal tools that assesses mental health and addiction domains (GAIN Coordinating Center, n.d.).

If you are constructing questions from scratch, consider using Likert-style, multiple-choice questions because they provide a continuum of responses and therefore can be more informative than a simple yes/no question. A question can look like this:

How satisfied are you with the mental health group you have been attending?

1. Very satisfied
2. Somewhat satisfied
3. Neutral
4. Somewhat dissatisfied
5. Very dissatisfied
6. Don't know
7. Choose not to answer

When developing questions from scratch, watch for the following common mistakes:

- Double-barrelled questions: questions that actually contain two questions; for example, "How satisfied were you with your caseworker and the amount of time the caseworker spent with you?";
- Response options that are not mutually exclusive: the question might contain more than one response option that might be true; for example, "How old were you when you first felt like you experienced symptoms of mental health problems: 10–12, 12–18, or 18–25?"; and
- Response options that are not exhaustive: the person might have a response that does not match any of the listed response options. A good way to avoid this problem is to include a catchall option like "None of these choices applies to me" (this wording is preferred to "Other" because it carries less negative connotations).

Establishing time points for surveys: Cross-sectional versus longitudinal surveys

Needs assessments and process evaluations usually involve a survey at only one point in time. For a needs assessment, the survey is usually conducted before the program or initiative begins. The survey for a process evaluation usually happens after the program has been running for a sufficient enough amount of time that participants are able to comment on its components.

Outcome evaluations, on the other hand, benefit from a pre-test/post-test design so it is possible to measure change over time. Conducting surveys before and after a program is essential to drawing valid conclusions. For example, in a survey conducted at the end of a six-month mental health group, all participants report good or excellent mental health. Based on these findings, the program developers conclude that the program was a success. The problem, however, is that without knowing each participant's self-reported mental health before the program begins, there is no way to determine whether participants in

fact experienced any change in their mental health as a result of the program. A problem with pre-test/post-test surveys is that participants cannot remain anonymous. Ethical considerations around anonymity are discussed in a later section.

CONTROL GROUPS

For outcome evaluations, the gold standard for assessing impact is to include a control group in the evaluation. Control groups are a sample of people who are similar to the people whose outcomes are being assessed, but who have not completed the program. Comparing the two groups allows evaluators to make sure any changes that were observed between the pre-test and the post-test can be attributed to involvement in the program. If the program group and the control group both change, it suggests that something other than the program caused the change. For example, sometimes people improve on their own over time without an intervention. Adding a control group obviously introduces a level of complexity, but without one, findings about the impact of a program must be interpreted with caution. One good option for establishing a control group is to use individuals on the program's wait list if one exists.

FOCUS GROUPS

Focus groups involve a small group of people (ideally five to eight) who are guided through discussion by a facilitator. A strength of this format is that participants can elaborate on the responses of other participants, which can be useful for establishing a broad understanding of a program or gaining insight into a shared group experience. Focus groups are also cost-effective and can be quicker to conduct than individual interviews. However, they require a moderator who has experience facilitating group discussions, and they are not anonymous. Lack of anonymity means that participants might be reluctant to speak up, particularly if they do not agree with the developing group consensus. Furthermore, sensitive topics should never be discussed in a focus group because confidentiality outside the group cannot be guaranteed and because the discussions might be triggering for some individuals. To get the most of the focus group, facilitators should be experienced in managing group conversations and should be prepared with five or six open-ended questions. Designating someone to take notes is a good way to capture main themes that emerge during the discussion. Audio-recording the session and transcribing the discussion is possible, but the process can be time-consuming.

QUALITATIVE INTERVIEWS

Qualitative interviews involve one-on-one conversations between an evaluator and a participant. They are particularly useful for gaining in-depth information about a person's unique needs or about the person's experiences with a program. They are also an appropriate format for discussing sensitive topics. The trade-off for the depth of the information is that findings may not be generalizable to a broader population. The ability to generalize is particularly limited in interviews (and other methods with small sample sizes) because certain types of clients might be more likely to participate. For example, people who had a bad experience with a program may be more likely to volunteer for an interview than people who had a good experience. The feedback might be useful in highlighting challenges with the program, but it might not provide an accurate overall picture of participants' experience with the program. Conducting 10 to 20 interviews captures a broad range of experiences and opinions and increases generalizability (the more interviews the better, but they are time intensive). An alternative is to conduct a survey to assess the group as a whole and then follow up with a few interviews to explore pertinent themes in more detail. If participants are all very similar to one another and the list of questions is short, then fewer interviews will be needed for common themes to emerge. As the participant sample becomes more diverse and the range of questions becomes broader, more interviews need to be added to compensate for the additional variation and complexity. To get the fullest picture, interviews should continue until the evaluators begin to hear repetition of themes, feelings, and experiences—this point in the data collection process is called saturation.

In-depth interviews can be thought of as guided conversations. Using open-ended questions (5–10 is ideal), these interviews are less structured and tend to follow participants where they want to go (within limits). The list of questions is not intended to be rigidly followed, but reminds the interviewer of key topics as the conversation progresses and flows naturally. The interviewer should probe for details and ask follow-up questions. Interviewers can use various strategies to ensure a successful interview. It is a good practice to ask participants to give examples to help illustrate their points. Beginning the interview with easier and less intrusive questions helps build rapport. Being open, genuine, and empathic, and using active listening skills are important characteristics of a good interviewer. A useful technique that captures a number of these skills is to cast the interviewee as the expert and to conduct the interview with that as the underlying principle.

CASE STUDIES

Case studies are another source of detailed information. They usually involve an in-depth review of three or four clients. Like interviews, they provide rich information but the findings are not easily generalized to other clients. A case study usually involves an interview with a case manager or other staff member, an interview with the client, and a review of internal case files. The idea is to gain as complete a picture of the client's experience as possible, including how the client came into the program and progressed through it, and what outcomes they experienced. It is useful to select case studies along a key dimension or central question. For example, an evaluation can review the case of a person who excelled in the program, another who struggled, and another somewhere in the middle. Case studies are particularly valuable for process evaluations because they provide detailed information on how a person experienced the program, including key points of friction between the person's circumstances and the structure of the program.

ETHICAL CONSIDERATIONS

Research ethics are a central component of high-quality evaluation research. The key components of ethical research are informed consent, voluntary participation, and confidentiality. A key principal cutting across these components is the need to acknowledge power imbalances and to ensure that participants are given an honest opportunity to choose whether and how they participate, that the research respects their safety and comfort, and that participants have an opportunity to share their stories as honestly as possible without being mediated or filtered. It is also important to acknowledge and thank participants for their involvement. This can be a simple thank you or a written note, but we strongly encourage organizations to consider a small honorarium as a token of appreciation and to acknowledge the person's time. Gift cards, food vouchers, and personal products are all good options. An honorarium does not need to be expensive; in fact, organizations should be careful not to use honoraria to convince people to participate because this practice does not reflect the value of voluntary participation (discussed below).

It is also useful to consider having any proposed evaluation project reviewed by an internal ethics committee. Committee members are familiar with the federal Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2; Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of

Canada, & Social Sciences and Humanities Research Council of Canada, 2014). An ethics review is not technically required for evaluation research under TCPS2 rules, but it is recommended to help identify unintended ethical issues (we all miss things), as well as to give outside people with research experience a chance to consider the research design and provide input.

INFORMED CONSENT

Research participants have the right to understand exactly what the research involves before agreeing to participate. The process of obtaining informed consent may include giving participants the opportunity to ask questions and get answers. Best practice around informed consent involves giving participants an information sheet that details the research. It should describe the method (interview, focus group, etc.), estimated time required, and risks associated with the research, and explain confidentiality and how personal information will be stored and shared. The evaluator should go through the sheet with the person to address potential literacy issues. Typically, both sign the sheet to indicate the content has been discussed and the participant freely agrees to participate.

VOLUNTARY PARTICIPATION

Participation in research should be voluntary: participants join of their own volition, without any feeling of pressure or coercion. This is a particular concern in evaluation research because participants may feel pressured to participate as a condition of receiving services or as a condition of their employment. Key to the informed consent process is clearly notifying participants that they are under no pressure to participate and that there will be no consequences for not participating. Voluntary involvement also means participants can choose not to answer questions and can withdraw their information from the evaluation even after the information has been collected (usually up until the point when analysis has started).

CONFIDENTIALITY

Participant information should be kept confidential to the fullest extent possible. This means striving to make participant involvement and information anonymous. Anonymous data means that even the researcher does not know who provided a particular response and that no identifying information (like names or birthdates) is collected at any point in the data collection process. This type of research is really only possible through anonymous, self-completed surveys and comment boxes. When the research cannot be anonymous, every effort should be made to keep the information as confidential as possible. In the case of pre-test/post-test surveys, this involves collecting the least amount of information needed to match the two surveys (e.g., initials and day of birth) and storing that information separate from the surveys. This can be accomplished by recording the ID number and the identifying information in a separate password-protected Excel document. A person's survey booklet is then identified with only their ID number. This same ID number is used for the second survey, allowing the evaluators to match the survey booklets.

DATA ANALYSIS

The final step after collecting evaluation data is analyzing the data. When conducting interviews and focus groups, transcripts are best, but detailed notes can also work. The goal of analyzing qualitative information like focus groups, interviews, case studies, and comment boxes is to carefully read through the information and identify themes and trends. It is useful at this stage to reflect on the broad research questions that were established at the beginning of the process and to organize the themes according to those questions. During data analysis, it is important for researchers to be open and self-reflexive to their biases to ensure they are not simply picking out themes that confirm their own understanding and interpretation of things. The goal of qualitative analysis is to really respect and honour the perspective of the interviewees. A helpful strategy for remaining open and self-aware is to pay particular attention to quotes from participants that contradict expectations.

Analyzing the results of quantitative surveys has its own challenges. It may require specialized statistical knowledge about how to assess the relationship between questions or variables. For example, the average score on a set of questions measuring life satisfaction increases by five points from the pre-test to the post-test. Without conducting additional

analyses, it is impossible to know whether this is a real statistical difference or whether a five-point difference is something that could easily happen by chance. It is best to get the help of someone with data analysis experience or to learn how to conduct basic statistical tests before drawing conclusions from the results of a survey. Common techniques to consider include descriptive statistics, which look at characteristics such as the distribution of data (histograms, frequency tables), the central tendency (mean, mode, median), and the dispersion of the data (standard deviation). Analysis can also involve using inferential statistics by analyzing cross-tabulation tables using chi-square tests, correlations using Pearson's *r*, and paired sample *t*-tests.

RESOURCES

Canadian Evaluation Society
evaluationcanada.ca

European Monitoring Centre for Drugs and Drug Addiction best practice tools
www.emcdda.europa.eu/themes/best-practice/tools

European Monitoring Centre for Drugs and Drug Addiction evaluation instruments bank
www.emcdda.europa.eu/eib?LanguageISO=EN

Evaluation handbook (W. K. Kellogg Foundation, 2010)
www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook

Program evaluation reference and resource guide (Ontario Treasury Board, 2007)
otf.ca/sites/default/files/274278.pdf

REFERENCES

Beidas, R. S., Stewart, R. E., Walsh, L., Lucas, S., Downey, M. M., Jackson, K., . . . Mandell, D. S. (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. *Cognitive and Behavioral Practice, 22*, 5–19.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada. (2014). *Tri-council policy statement: Ethical conduct for research involving humans*. Retrieved from www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf

GAIN Coordinating Center. (n.d.). GAIN instruments. Retrieved from http://gaincc.org/instruments/

W. K. Kellogg Foundation. (2006). *Logic model development guide*. Retrieved from www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide

World Health Organization. (1997). Quality of Life instrument—brief version (WHOQOL-BREF). Retrieved from www.who.int/mental_health/publications/whoqol/en

ABOUT THE AUTHOR

Tyler Frederick, PhD, is a sociologist and an assistant professor at the University of Ontario Institute of Technology. He is a community-based researcher with a focus on marginalized young people. His research focuses on how young people navigate homelessness and how this process shapes their mental health, identity, and well-being.