A Guide to Evaluating Marine and Family Division Programs, Services, and Capabilities

An Introduction to Program Evaluation

Stephanie Brooks Holliday, Stephani L. Wrabel, Thomas E. Trail, Margaret Tankard, Dana Schultz, Laura Werber, Sarah O. Meadows

RAND CORPORATION

For more information on this publication, visit **www.rand.org/t/TL264z1**

Published by the RAND Corporation, Santa Monica, Calif.

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-1-9774-0009-3

Cover: Photo by U.S. Marine Corps Cpl. Alexander Mitchell.

Preface

The U.S. Marine Corps (USMC) Marine and Family Division (MF) portfolio includes a variety of programs and services that cut across a number of domains, including behavioral health, personal and professional development, and family readiness. Given increased emphasis on budgetary constraints and efficient spending, assessing the effectiveness of programs and services has gained significance. For MF programs and services to remain meaningful, credible, and practical, their effectiveness must be able to be demonstrated. MF has identified program evaluation as a critical tool to determine whether its programs and services are being implemented as planned and achieving their intended goals.

The goal of this document is to provide an introduction to program evaluation for MF program managers. Although some program managers may have experience with aspects of program evaluation, the process may be less familiar to others. Therefore, this document provides a broad overview of the process of conducting a program evaluation, including types of evaluation, the ways that data are collected and analyzed, and how the results of an evaluation are used. For program staff interested in more detail about the process, this review provides a reference list of resources, as well as a series of worksheets that can be used to plan and implement a program evaluation.

This research was sponsored by the USMC MF and conducted within the Forces and Resources Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community.

For more information on the RAND Forces and Resources Policy Center, see http://www.rand.org/nsrd/ndri/centers/frp.html or contact the director (contact information is provided on the web page).

Contents

Figures

Tables

Worksheets

Introduction

In an effort to help service members and their families navigate the stresses associated with deployments and other aspects of military life, the U.S. Marine Corps (USMC) has developed a set of programs, services, and policies that provide support to Marines and their families. The Marine and Family Division (MF) portfolio includes a variety of programs and services that cut across a number of domains, including behavioral health, personal and professional development, and family readiness. These resources are associated with Marine personal readiness, so the portfolio of programs and services is indirectly, and critically, tied to the USMC mission and its future success.

Given increased emphasis on budgetary constraints and efficient spending, assessing the implementation and effectiveness of programs and services has gained significance. Although there have been evaluation efforts within the MF to date, there is no overarching MF evaluation plan that can be applied across programs and services. For MF programs and services to remain meaningful, credible, and practical, their effectiveness must be able to be demonstrated.

Program evaluation is the application of systematic processes for collecting and analyzing information about a program to determine the program's quality of operation and ability to meet intended purposes (i.e., effectiveness).[1] At its core, program evaluation is used to guide decisionmaking about the right set of activities to offer, inform efforts to improve current activities, identify areas of practice that are underperforming or are performing above expectation, and demonstrate the value of a particular program in effectively contributing to high-priority goals within an organization.

Program evaluation is valuable as a tool regardless of the current performance of a program. It is useful for programs that appear to be struggling to meet intended purposes, as the evaluation can identify program strengths and weaknesses and inform an action plan for improvement. But well-implemented and seemingly effective programs can also use the results of an evaluation to improve the efficient allocation of resources, streamline the delivery of program components, and turn good programs into great programs for members of our military services.

In this document, we provide a brief introduction to program evaluation by focusing on five steps:

1. identifying program goals and core components with a logic model

[1] P. H. Rossi, M. W. Lipsey, and H. E. Freeman, *Evaluation: A Systematic Approach*, 7th ed., Thousand Oaks, Calif.: Sage Publications, 2004; J. S. Wholey, H. P. Hatry, and K. E. Newcomer, eds., *Handbook of Practical Program Evaluation* (3rd ed.), San Francisco, Calif.: Jossey-Bass, 2010.

2. planning for evaluation
3. operationalizing evaluation measures
4. implementing an evaluation and interpreting results
5. using evaluation results for continuous quality improvement.

Although we refer to "programs" throughout much of this document, these steps apply to the evaluation of both programs and services.

This document also contains two appendixes:

- Appendix A provides a list of additional resources for readers interested in reading more about program evaluation.
- Appendix B includes a series of worksheets designed to help users develop a program evaluation plan.

This document will be a useful resource for MF Division program managers, particularly those whose programs have been selected for evaluation. For staff familiar with program evaluation, this information will serve as brief review. For those less familiar with program evaluation, this document will provide a context for the evaluation process. A more detailed version of this guide, which includes a set of Excel worksheets that program evaluators may use to guide their own evaluation efforts, can be found in *A Guide to Evaluating Marine and Family Division Programs, Services, and Capabilities*.[2]

[2] S. Holliday, S. L. Wrabel, T. Trail, M. Tankard, D. Schultz, L. Werber, and S. O. Meadows, *A Guide to Evaluating Marine and Family Division Programs, Services, and Capabilities*, Santa Monica, Calif.: RAND Corporation, TL-264-USMC, 2021.

Five Steps of Program Evaluation

Step 1. Identifying Program Goals and Core Components with a Logic Model

An important first step in program evaluation is to clearly articulate the program goals and link them to the core components of a program or service. The core components can be organized into a logic model that shows how a program or service works and what outcomes are expected to achieve program goals. Logic models can be the basis for an evaluation plan to assess the implementation process and examine effectiveness. Here, we discuss the process of identifying the core components of a program or service.

SMART Goals

Before initiating a program evaluation, it is helpful to first clearly define the goals of the program or service. Well-defined goals follow the SMART principles.[1] That is, they should be Specific, Measurable, Achievable, Relevant, and Time-bound. This is what is meant in the context of identifying program goals:[2]

- **Specific**: The program goal(s) should describe exactly what is expected to change, and for whom.
- **Measurable**: There must be a way to determine whether a goal has been met, the extent of progress that has occurred toward this goal, or change on the expected outcomes as a result of a program.
- **Achievable**: The goals of a program must be feasible for the target population (e.g., based on prior empirical expectations for change) and with the available resources.
- **Relevant**: The goals should be related to the program purpose and mission.
- **Time-bound**: The time frame in which the change is expected to occur should be specified.

Logic Models

Once program or service goals have been identified, the next step is to identify the core components of the program or service. These elements can be organized using a *logic model*. A logic model is a widely used way to visually present the relationships between a program or ser-

[1] C. A. Lesesne, K. M. Lewis, C. Moore, D. Fisher, D. Green, and A. Wandersman, *Promoting Science-Based Approaches to Teen Pregnancy Prevention Using Getting to Outcomes: Draft 2011*, unpublished manual, 2011.

[2] J. D. Acosta, R. N. Ramchand, A. Becker, A. Felton, and A. Kofner, *RAND Suicide Prevention Program Evaluation Toolkit*, Santa Monica, Calif.: RAND Corporation, TL-111-OSD, 2013.

vice's core components and summarize the way that a program or service works. Logic models include the following elements:

Resources/inputs. This refers to the assets needed and available to support the operation of the program. The performance of activities and the fulfillment of goals depend on these resources. A successful program or service needs adequate resources to fulfill needs, as well as prudent allocation of these resources to avoid wasting time and money. Resources can include staff, including the number of staff and types of credentials needed; facilities and infrastructure (e.g., technology); funding; and any other resources needed for operational purposes, such as training materials, testing equipment, or external contractors.

Activities. Activities are the specific actions or strategies that make up the program or service and are employed to reach goals. This includes the tools, services, and products that the program or service provides to the target population. When developing the logic model, it can be helpful to consider not only what activities are part of the program or service, but the frequency and duration of each activity, who delivers or supports its implementation, and the recipient of each activity (e.g., active-duty Marine, spouse, family). Although not a formal component of a logic model, it is also worthwhile to consider the **target population**—that is, the individuals that the program or service intends to serve with its activities.

Outputs. Outputs are the amount, quality, and/or volume of each activity within a program or service. These can include the number of individuals trained, number of individuals tested, the quantities of materials developed or distributed, or participant satisfaction.

Outcomes. Outcomes are changes in the target population expected as a result of engaging in the activities and often include changes in knowledge, attitudes, skills, or behaviors. Outcomes are observed over the short and intermediate terms. Identifying outcomes is important because, ultimately, a goal of program evaluation is to understand and demonstrate how a program or service has resulted in changes in the outcomes of interest. Outcomes are typically measured at the individual or unit level.

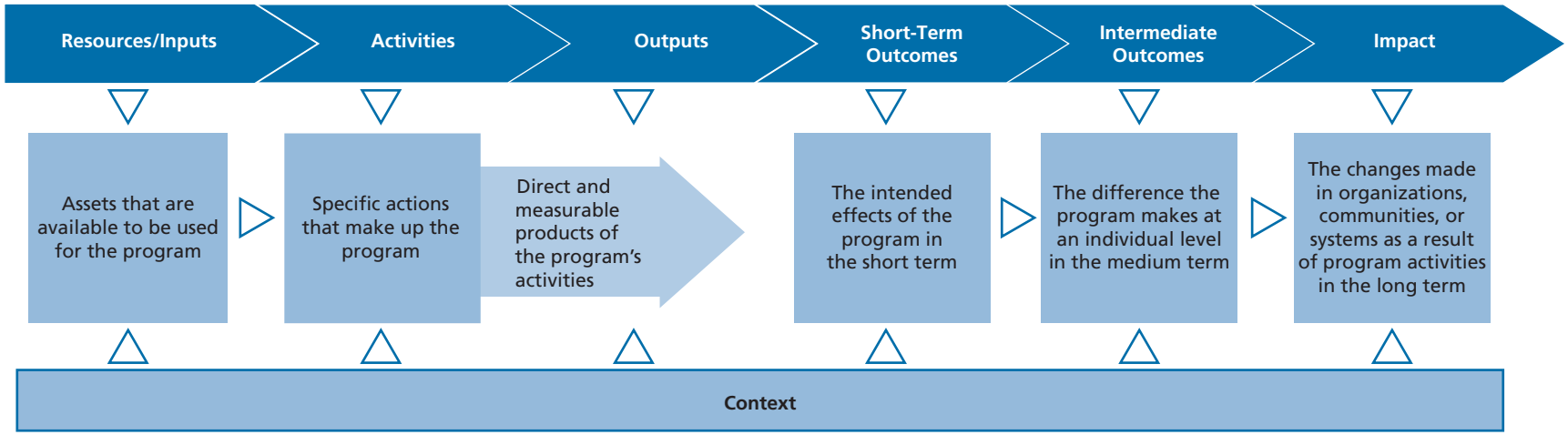
Impact. These impacts occur at the aggregate level—that is, at the unit, installation, or total-force level. Across the MF portfolio, programs and services vary widely in their focus and activities. However, they are largely designed to increase readiness, improve resilience, increase retention in the service, and improve quality of life for service members and their families.

Context. It is also important to consider the local and external context in which a program or service operates. These could be political, social, cultural, or economic situations within the USMC or at specific installations that can facilitate or hinder the utilization of MF programs, services, or capabilities. Since a program or service may work differently in different settings, logic models can identify contextual factors that may influence both implementation and the subsequent outcomes and explain why those outcomes differ across settings (or contexts).

A basic outline of a logic model appears in Figure 1.

Logic models are useful tools for program evaluation because they facilitate the visualization of relationships between the core components of a program or service and, therefore, guide the development of an evaluation plan. In addition, logic models can depict the interrelationship between elements of the major components. This enables programs and services to more clearly consider the links between outputs and outcomes—both actual and desired—and helps in the identification of potential disconnects (e.g., intended outcomes that are not supported by any program outputs).

Figure 1
Logic Model Overview



RAND *TL264/1-1*

Step 2. Planning for an Evaluation

After outlining the core components of a program or service in a logic model, the next step is developing an evaluation plan that enables an assessment of program performance and effectiveness. Evaluations generally can be categorized as process evaluations or outcome evaluations. Process evaluations focus on how a program or service was implemented, and whether it was delivered as intended. Outcome evaluations focus on the outcomes and impacts that a program or service is designed to achieve to meet its goals. Figure 2 shows how process and outcome evaluation map onto the core components of the logic model.

Process Evaluation

Process evaluations generally aim to "[assess] what activities were implemented, the quality of the implementation, and the strengths and weaknesses of the implementation" (Chinman, Imm, and Wandersman, 2003).[3] As shown in Figure 2, process evaluations correspond to the resources/inputs, activities, and output sections of the logic model and use *measures of performance* (MOPs) to assess program usage and implementation (see Step 3 for more details on defining MOPs).

There are a number of factors that may be considered as part of a process evaluation:

- service utilization patterns, including participation, attendance, dosage (i.e., how much of a service was used), and/or attrition[4]
- staff qualification and available resources[5]
- fidelity/adherence and quality of service implementation[6]
- provider or client perceptions of the usefulness or relevance of the program[7]
- satisfaction with the program, including satisfaction of program staff and of individuals served by the program.[8]

This component of an evaluation helps an evaluator to understand how the program or service was delivered to meet goals and objectives. When they are well designed, process evaluations can produce critical findings on facilitators and barriers to implementation that offer important insights for future replication of a program or service. In addition, understanding how a program was implemented provides a context for the results of an outcome evaluation. If a program is not achieving the intended outcomes, it may be because it is not reaching the

---

[3] M. Chinman, P. S. Imm, and A. H. Wandersman, *Getting to Outcomes™ 2004: Promoting Accountability Through Methods and Tools for Planning, Implementation, and Evaluation*, Santa Monica, Calif.: RAND Corporation, TR-101-CDC, 2003.

[4] Acosta et al., 2013; Centers for Disease Control and Prevention, *Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide*, U.S. Department of Health and Human Services, 2011; M. Lalayants, "Differential Program Evaluation Model in Child Protection," *Child Welfare*, Vol. 91, No. 4, 2012, pp. 9–40..

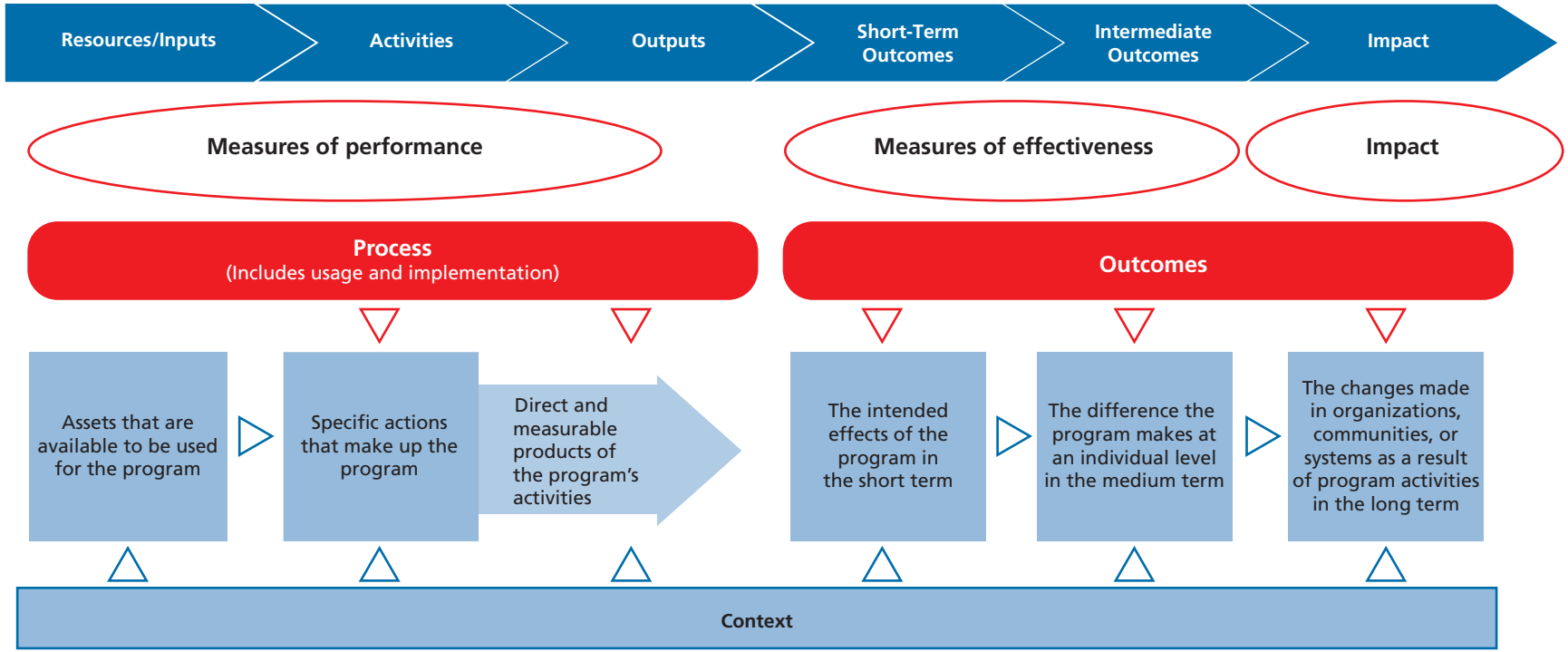
[5] Centers for Disease Control and Prevention, 2011; W. K. Kellogg Foundation, *W.K. Kellogg Foundation Evaluation Handbook*, Battle Creek, Mich., 2004.

[6] Acosta et al., 2013; Chinman, Imm, and Wandersman, 2003.

[7] Acosta et al., 2013; Chinman, Imm, and Wandersman, 2003; Lalayants, 2012; R. C. McNeil, "A Program Evaluation Model: Using Bloom's Taxonomy to Identify Outcome Indicators in Outcomes-Based Program Evaluations," *Journal of Adult Education*, Vol. 40, No. 2, 2011, pp. 24–29.

[8] Acosta et al., 2013; Chinman, Imm, and Wandersman, 2003; Lalayants, 2012.

Figure 2
Framework for Evaluating a Program or Service



RAND *TL264/1-2*

target population, it is not being implemented with fidelity, and/or program participants are dissatisfied with certain aspects of the program.

Outcome Evaluation

Outcome evaluations focus on the short-term and intermediate outcomes expected to result from a program or service. Rather than evaluating the implementation of a program, outcome evaluations focus on program results or outcomes—that is, whether changes that are expected as a result of the program actually occurred.[9] Outcome evaluations use *measures of effectiveness* (MOEs) to assess these outcomes, which are often measured at the client level or program level (see Step 3 for more details on defining MOEs).[10]

There are several design options that can be used for an outcome evaluation. These options vary in their level of methodological rigor. In general, the more rigorous the design, the more confident an evaluator can be that any changes observed are the result of a program or service and not other factors (e.g., other programs or services, outside contextual factors such as social support or the passage of time). Here, we introduce three design options. The advantages and disadvantages of each are summarized in Table 1.

1. **Experimental evaluations**, also known as randomized controlled trials (RCTs), compare people randomly assigned to participate in a program or service to individuals assigned to a control group. Members of the control group forgo participation in the program or service during the evaluation, although they may be able to access it after the evaluation is complete. This is the most methodologically rigorous evaluation option,[11] but RCTs can be challenging to implement in real-world settings because they are resource-intensive and require a group willing to delay their participation in the program or service.
2. **Quasi-experimental evaluations** compare participants in a program or service to individuals in a comparison group. Unlike experimental designs, however, the comparison group is not randomly assigned.[12] Instead, a comparison group is selected to be as similar to the group participating in the program or service as possible. Although somewhat less logistically challenging to implement, the comparison group in a quasi-experimental evaluation cannot be assumed to be equivalent,[13] making this a somewhat less rigorous option.
3. **Observational evaluations** focus on participants in a program or service alone without a control or comparison group. In some observational evaluations, participants are administered surveys before and after participation in the program to examine changes

---

[9] McNeil, 2011; Acosta et al., 2013; Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury, *Program Evaluation Guide*, Arlington, Va.: Defense Health Agency, 2012.

[10] W. K. Kellogg Foundation, 2004.

[11] Acosta et al., 2013; National Center for Education Evaluation and Regional Assistance, *Identifying and Implementing Educational Practices Supported by Rigorous Evaluation: A User Friendly Guide*, Washington, D.C.: Institute of Education Sciences, 2003.

[12] Defense Centers of Excellence, 2012.

[13] L. T. Martin, C. Farris, D. M. Adamson, and R. M. Weinick, *A Systematic Process to Facilitate Evidence-Informed Decisionmaking Regarding Program Expansion: The RAND Toolkit, Volume 3*, Santa Monica, Calif.: RAND Corporation, RR-487/3-OSD, 2014.

in knowledge, attitudes, skills, or behaviors. Other observational evaluations administer only post-participation measures but ask participants to reflect on changes that occurred during participation.[14] Observational evaluations are less resource-intensive but make it challenging to rule out other explanations for outcomes that are observed.[15]

Table 1
Advantages and Disadvantages of Outcome Evaluation Design Options

Evaluation Design	Advantages	Disadvantages
Experimental • Compares program participants to randomly assigned control group	• "Gold standard" evaluation design for program effectiveness evaluation • Strong control groups that are equivalent to treatment groups making results clearer • Builds program or service research capacity	• Difficult to implement in real-world settings • High level of resources and expertise needed for recruitment, retention, and data collection • Reluctance of some individuals to participate in RCTs • Reduces the number of individuals served if recruitment is low
Quasi-experimental • Compares program participants to comparison group	• Comparison group drawn from the same or similar target population, so similar to treatment group • Able to serve more individuals since not losing half to randomization	• Difficulty recruiting and retaining individuals in a comparison group • Need for sophisticated statistical adjustments that may or may not adequately control for bias because of differences between groups • Funding and staffing issues of concern for feasibility
Observational—pre-post design • Examines program participants alone, comparing status before and after participation	• Less resource intensive, as no data collection from control or comparison group required • Straightforward way to measure change	• More challenging to rule out alternative explanations for effects that are observed
Observational—post-test only • Examines program participants alone, using post-participation assessment only	• Less resource intensive • Only one assessment point needed	• More challenging to rule out alternative explanations for effects that are observed • Recall bias or inaccuracies when asked to report on pre-activity status

SOURCE: Acosta et al. (2013); RAND National Defense Research Institute review of relevant literature.

Types of Evaluation Data

Objective and Subjective Data

Both objective and subjective data may be collected as part of a program evaluation. *Objective data* refers to information that is observable and not subject to influence by an individual's perceptions or beliefs.[16] By contrast, *subjective data* capture personal opinions, beliefs, and experiences. Both types of data can be used for process and outcome evaluation. For process

[14] Acosta et al., 2013.

[15] Defense Centers of Excellence, 2012.

[16] J. Hourcade, P. Parette, and H. Anderson, "Accountability in Collaboration: A Framework for Evaluation," *Education and Training in Developmental Disabilities*, Vol. 38, No. 4, 2003, pp. 398–404.

evaluation, objective data might include the number of people participating in a workshop, and subjective data could include ratings of satisfaction. For outcome evaluations, objective data could include scores on a knowledge test, while subjective data could be perceptions of social support or community connectedness. The selection of the most appropriate type of data will depend in part on the measures of performance and effectiveness identified for the evaluation (see Step 3 for more details on defining these).

Qualitative and Quantitative Data

Program evaluations may rely on qualitative data, quantitative data, or some combination. *Qualitative data* include descriptive information about the program or service and the implementation process and are typically associated with process evaluation. Qualitative data collection methods include interviews with key informants (e.g., program staff and administrators, program participants), focus groups, and observation of activities.[17] Qualitative data can provide rich information or context but can also be time-consuming and resource-intensive to collect and analyze.

Quantitative data are data that are in a numerical format and can contribute to both the process and outcome evaluation components of the evaluation.[18] Data are often collected through surveys or questionnaires or may be obtained from existing clinical or administrative data.[19] Data can be collected across multiple levels, including the individual, group, and program levels.[20]

Primary and Secondary Data

Another consideration is the use of primary data, secondary data, or some combination.[21] *Primary data* refers to information collected specifically for the evaluation. This might include attendance lists or satisfaction surveys administered to participants in a program or service (process evaluation), or behavioral health measures or knowledge tests (outcome evaluations). Benefits of primary data are that the program and/or evaluator determines what is collected and when and is aware of the limitations of the data. However, the burden of data collection falls on the program or service.

Secondary data are compiled for purposes outside the administration of the program or service but are relevant to the outcomes of the program or service. This could include large survey efforts, administrative data, or data collected from other evaluations. Although the burden for data collection falls outside the program or service, programs have less control over

---

[17] Chinman, Imm, and Wandersman, 2003; P. LeBlanc and C. Lacey, *Peace Works: A Program Evaluation Model*, paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, La., 2002; W. K. Kellogg Foundation, 2004; World Health Organization, *WHO Evaluation Practice Handbook*, Geneva, Switzerland, 2013.

[18] World Health Organization, 2013.

[19] Chinman, Imm, and Wandersman, 2003; Lalayants, 2012; World Health Organization, 2013.

[20] LeBlanc and Lacey, 2002; D. G. Unger, E. A. Park, P. Antal, P. A. Tressell, K. Rigney, K. DeRasmo, and J. Kassess, "Serving Children with Special Social and Emotional Needs: A Practical Approach to Evaluating Prevention Programs in Schools and Community Settings," *Journal of Educational & Psychological Consultation*, Vol. 11, No. 2, 2000, pp. 273–296.

[21] World Bank, *Doing Impact Evaluation Number 6: Data for Impact Evaluation*, Washington, D.C.: Poverty Reduction and Economic Management, Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation, World Bank, 2009.

the content of secondary data, and it can be challenging to fully know the limitations to the data (e.g., mistakes or errors in the data).[22]

Evaluation Considerations

When designing a program evaluation, there are several factors to consider. Here, we briefly identify some of the factors that shape program evaluation design.

Number of Participants

Program evaluations can include the entire population of individuals who use a program or service or may focus on a sample or subset of the population. There are many benefits to surveying the population,[23] as an evaluator can be sure that the evaluation represents all participants. However, depending on the number of people using the program or service, this can be time- and resource-consuming. Therefore, it may be more realistic and less burdensome to recruit a sample. In this case, the factors that influence the size of a sample include the flow of participants through the program and the methods being used, as qualitative methods are time-consuming with large samples.

In general, a larger sample tends to be more representative as long as it was recruited in a way that reflects the overall population using the program or service. However, a large sample will not always result in an unbiased sample. For example, a sample of 1,000 spouses of junior enlisted Marines would provide very precise estimates on any given outcome measure. Those estimates, however, would tell you something about only that particular group of spouses, and the results may not hold for spouses of senior enlisted officers or junior officers. There are ways to increase the representativeness of a sample if it is not possible to collect data from all participants in a program or service.[24] For example, evaluators might use random sampling, which means that each person who uses a program during the evaluation period has an equal chance to be enrolled in the evaluation. Another option is stratified random sampling, which helps select a sample similar to the overall population that uses a program or service with respect to certain characteristics (e.g., gender, age).

Timing of Evaluation

Ideally, an evaluation is planned and implemented prior to the start of the program or service. This has several advantages for process and outcome evaluations. For process evaluation, this allows for usage data to be collected in real time; for outcome evaluation, it enables a program to collect pretest data from participants. Another consideration related to outcome evaluation is the timing of outcomes. Some effects are observed in the short term (e.g., immediately after the program ends), whereas others will not be observed until sometime after the program has ended (e.g., six months later). It is important for programs to consider the time frame over which effects are expected to be observed.

---

[22] J. J. Hox and H. R. Boeije, "Data Collection, Primary Versus Secondary," in K. K. Leonard, ed., *Encyclopedia of Social Measurement*, Amsterdam, Netherlands: Elsevier, 2005, pp. 593–599; World Bank, 2009.

[23] L. Daugherty, L. Werber, K. Kamarck, L. Harrington, and J. Gazis, *Officer Accession Planning: A Manual for Estimating Air Force Officer Degree Requirements*, Santa Monica, Calif.: RAND Corporation, TL-196-AF, 2016.

[24] C. Teddlie and F. Yu, "Mixed Methods Sampling: A Typology with Examples," *Journal of Mixed Methods Research*, Vol. 1, No. 1, 2007, pp. 77–100.

Available Evaluation Resources

Program evaluations require time, infrastructure, and other resources, and certain types of outcome evaluations (e.g., RCTs) can be more resource-intensive than others. Key considerations include the existing infrastructure for collecting, storing, and analyzing data, and financial demands, such as staffing and software needs.

Another important consideration is human capital needs, including the availability of staff with evaluation expertise and staff to carry out evaluation activities. Regarding evaluation expertise, it is important to have evaluation personnel with expertise in the following domains:[25]

- **Program or service content**: This person is familiar with the program's objectives and activities as well as the day-to-day operation. This person will be an asset when developing the logic model, determining what type of evaluation plan will be feasible to implement, and in the ultimate interpretation of results of the evaluation.
- **Evaluation design and methods**: This person will design the evaluation plan and oversee the implementation of the evaluation. This person may be internal or external to the program being evaluated, depending on the expertise of program staff. This type of expertise is especially helpful if there is a need to design data-tracking forms or develop measures for the evaluation, and when implementing more technical evaluation designs, such as RCTs.
- **Data analysis**: It is important to have personnel responsible for analyzing and interpreting the data yielded by the evaluation. Multiple data analysis experts may be needed, depending on the type of data collected (e.g., qualitative versus quantitative data).

In addition, it is important to consider the availability of personnel for key evaluation tasks, such as developing and preparing data collection instruments; collecting data related to implementation and usage; collecting data on short-term outcomes; and following up with participants as needed to assess long-term outcomes. The evaluators may complete these tasks, but it is often useful to have the assistance of program or service staff in completing these tasks, since they are interacting with program participants on a more regular basis.

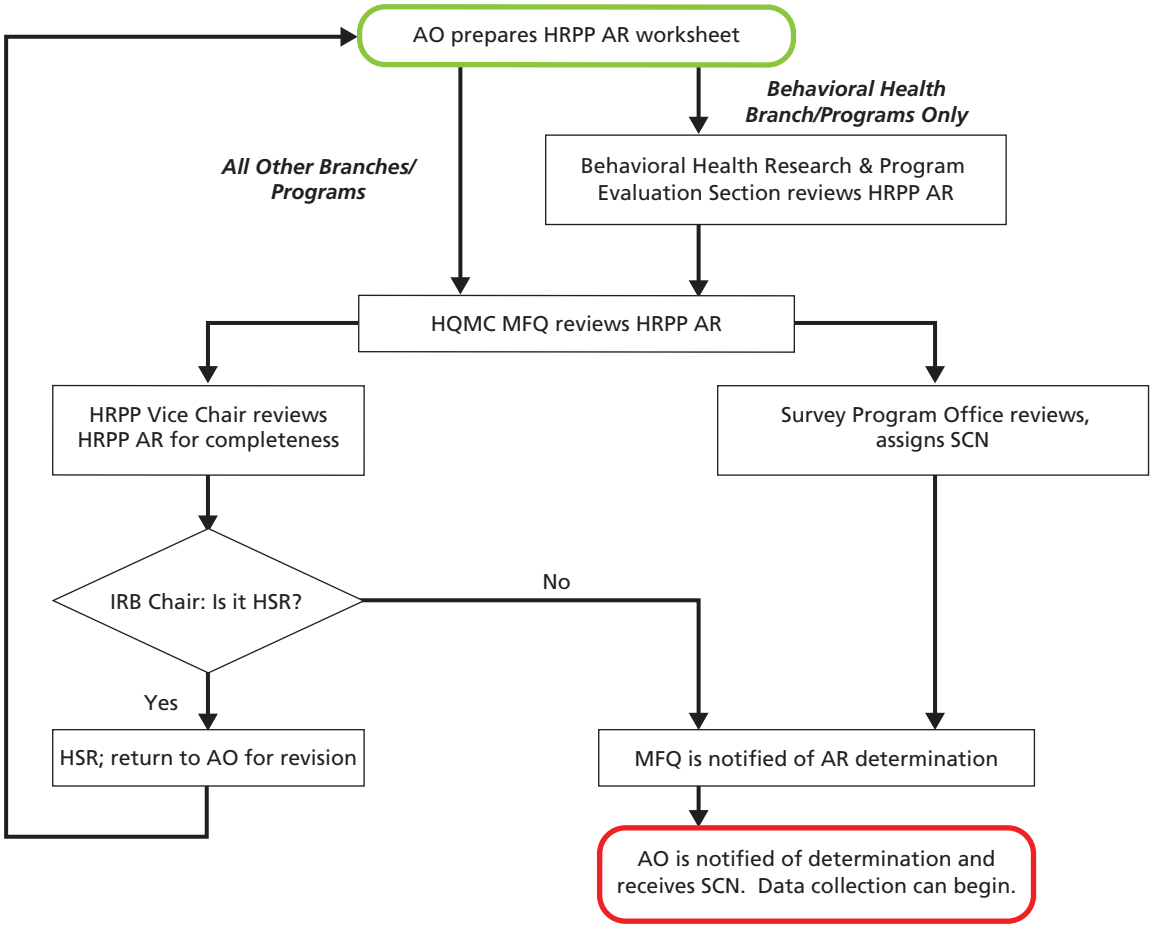
Collecting Sensitive Data and Protection of Participants

Some of the information collected about participants could be sensitive (e.g., mental health symptoms, financial standing, marital satisfaction). Therefore, it is critical to have a plan to safeguard the identity of participants and any information collected about them. This is especially important when identifiable information is needed as part of data collection—for instance, if the program is collecting data at multiple time points and needs to link it to a specific person over time, or is using identifiable information (e.g., home address, email address) to send out evaluation materials, such as surveys.

Before undertaking an evaluation effort, it is also important to be familiar with sources of regulatory and ethical guidance related to the protection of human subjects. There are certain distinctions related to "quality improvement" activities and human subjects research that may apply to a given evaluation. However, this is not a determination that should be made by an evaluator, program, or service in isolation. You should follow the MF-designated USMC Human Subjects Research Protection Program (HRPP) process (see Figure 3). All formal

[25] W. K. Kellogg Foundation, 2004.

Figure 3
HRPP Applicability Review Process



Key

AO = Action Officer	IRB = Institutional Review Board
AR = Applicability Review	MFQ = Marine & Family Programs Future Operations Branch
HRPP = Human Research Protection Program	SCN = Survey Control Number
HSR = Human Subjects Research	SME = Subject Matter Expert
HQMC = Headquarters Marine Corps	

NOTES:
1. The AO (can be at the HQMC or installation level) submits appropriate HRPP AR worksheet to HQMC MFQ (National Research Section).
2. HQMC MFQ conducts a preliminary review of the HRPP AR worksheet and may involve SME input prior to submitting HRPP AR worksheet to the HRPP Vice Chair and any surveys to the Survey Program Office.
3. The HRPP Vice Chair and the Survey Program Office conduct concurrent reviews of the submitted HRPP AR worksheet.
   a. HRPP Vice Chair makes preliminary determination regarding HSR, ensures all necessary documentation is present, then forwards their determination to the IRB Chair.
   b. Survey Program Office reviews survey, ensures scientific merit, and assigns SCN.
4. IRB Chair makes the final HSR determination based on submitted documentation from the HRPP Vice Chair
   a. If determined to NOT be HSR, the original submitting AO is good to go, but must have the SCN for any surveys prior to initiation.
   b. If determined to be HSR, the original submitting AO will need to submit additional paperwork per policy.

RAND *TL264/1-3*

mechanisms to solicit feedback from program participants, staff, commands, and other stakeholders (e.g., research, surveys, questionnaires, course evaluations, suggestion box materials) are vetted through the HRPP Applicability Review Process *prior* to use. HRPP provides study determinations concerning whether research, surveys, questionnaires, course evaluations, and suggestion box materials are in accordance with ethical principles, laws, regulations, and policies to ensure respect for persons, beneficence, and justice for human subjects. Program evaluations are considered aspects of research and *must* be vetted through the HRPP *before* beginning. Further information concerning the HRPP Applicability Review Process is available from the HQMC MF National Research Section: https://ehqmc.usmc.mil/sites/family/mfq/national_research/default.aspx. Other Department of Defense and Marine Corps guidance related to human subjects protection and survey administration includes

- Department of Defense Instruction 3216.02, *Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research*
- Marine Corps Order 3900.18, *Human Research Protection Program (HRPP)*
- Secretary of Navy Instruction 3900.39D, *Human Research Protection Program*
- HRPP and Institutional Review Board website (http://www.tecom.marines.mil/Resources/HRPP/)
- Marine Corps Order 5300.18, *Marine Corps Survey Program*
- 45 Code of Federal Regulations 46.102.

Step 3. Evaluation Measures

An important aspect of designing an evaluation is the selection of evaluation measures. As noted in Step 2, there are two types of evaluation measures: MOPs and MOEs.

MOPs are a component of process evaluations and are used to evaluate the usage and implementation of a program. On the logic model, this corresponds to the program inputs, activities, and outputs (see Figure 2). For example, MOPs might address questions related to what program activities are implemented, how many people participated in the activities, and how satisfied program participants are. MOPs have a more immediate time frame.

MOEs are a component of outcome evaluations and are used to measure whether a program or service has achieved the intended outcomes. MOEs allow organizations to determine what participant-level changes resulted from program or service activities. These outcomes may be observed in the short term and intermediate term. A summary of the characteristics of MOPs and MOEs is presented in Table 2.

Defining Measures of Performance

A program's logic model is a useful tool for identifying MOPs, as it clearly outlines the key elements of implementation, including resources, target population, and activities. In turn, these dictate the MOPs. For example, program staff might be interested in knowing how many people the program served, how many workshops were held, or how satisfied program participants were.

Table 3 provides sample MOPs and definitions for each component of the logic model. There are certain considerations that can guide the process of selecting MOPs and determining

Table 2
Characteristics of Measures of Performance and Measures of Effectiveness

	Measures of Performance (MOPs)	Measures of Effectiveness (MOEs)
Type of evaluation	Process/implementation evaluation	Outcome/impact evaluation
Questions addressed	What activities are implemented? How many people participate in activities? How satisfied are participants?	What participant-level changes resulted from program activities?
Level of analysis	Individual or family level	Individual or family level
Time frame	Immediate	Short term and intermediate

how to operationalize them. First, the more specific the MOP, the more straightforward it will be to decide how it should be operationalized. Another consideration is the source of data that will be used for the MOP—are data needed for the MOP already being collected? If not, how can data be collected effectively while minimizing burden?

Table 3
Sample Measures of Performance

Measure of Performance	Definition
Inputs/Resources	
Staff are adequately trained	Number of staff with required level of training/number of staff
Equipment has required registration	Proportion of available pieces of equipment that have required registration
Activities/Outputs	
Amount of program activates	Number of workshops taking place each year, by topic area
Average attendance, by type of session	Number of families who attend a session/Number of sessions
Participant satisfaction	Proportion of participants who report being satisfied or very satisfied with activity

Defining Measures of Effectiveness

MOEs measure short-term and intermediate outcomes and help to determine if a program or service is achieving its intended goals. For the MF branch programs and services, the desired outcomes generally fall into one of four broad categories: knowledge, skills, attitudes, and behaviors. For example, MOEs related to the acquisition of new *knowledge* may include a better understanding of the USMC mission, learning best practices in nutrition and exercise, or knowledge of employment options. *Skills* might include communication skills, positive coping, financial literacy, or job skills. MOEs related to *attitudes* might include marital satisfaction, self-efficacy, or social cohesion. Finally, *behaviors* might include reducing negative behaviors (e.g., substance misuse) or building a social network. As with MOPs, the logic model can be used as a tool to identify MOEs.

After identifying MOEs, the next step is determining how to operationalize these MOEs. Some MOEs are specific to the content of a given program. For example, when a program or service offers workshops, courses, or sessions on a specific topic, the primary short-term outcome is often increased knowledge specific to that topic. In these cases, the best way to measure changes in knowledge might be to develop and administer a measure specific to the content of that session. By administering this measure before and after the workshop, an evaluator can determine whether there is any improvement in knowledge. Other MOEs might include the proportion of participants who accomplished a particular goal—for example, the proportion of participants who successfully obtained a job, or who met physical fitness standards.

Another option for operationalizing MOEs might be to identify an instrument developed and validated through previous research or evaluation projects. For example, an outcome associated with MF branch programs might be improved perceptions of social support from other military families. There are a number of evidence-based instruments (that is, instruments developed and validated by scientific research) that measure social support, including the Reduced Sound Relationship House Questionnaire.[26] Rather than developing a new measure of social support specific to the workshop, an evaluator could administer this instrument. Even if an evaluator cannot find an instrument perfectly suited to a given program, it may be possible to identify and adapt an existing scale for the evaluation.

Table 4 provides examples of MOEs that may be used to measure short-term and intermediate outcomes of a program. It also provides examples of ways these MOEs could be operationalized.

Step 4. Implementing an Evaluation and Interpreting Results

Implementing an Evaluation Plan

There are several key decisions to make for translating an evaluation plan to implementation of an evaluation, including determining who will conduct the evaluation and over what time frame. Evaluations often require the coordinated efforts of multiple people. For example, MF branch headquarters may design an evaluation plan. To ensure consistency across installations, it could identify MOPs and MOEs and determine how and when they will be measured. When it comes to collecting data, program or service managers and staff at the installation level are likely to play key roles by assisting in such tasks as documenting attendance at workshops, directing USMC family members to a feedback survey to complete, recording referrals that they make, or providing information about their credentials. Ideally, the specific roles and responsibilities of all individuals involved in the evaluation effort are defined in advance.

Another important decision is the timeline of an evaluation. Data collection and analysis can take place at specific time intervals (e.g., quarterly, yearly) or in relation to specific sessions or events (e.g., a workshop). Time intervals may be selected based on when and across what period of time specific changes are expected to take place, such as changes in program content or changes in the lives of participants. Other considerations might be the flow of participants. For example, if a program serves large numbers of individuals, it may be possible to examine the implementation and outcomes over a shorter time frame. Finally, some programs and services may have specific requirements related to their implementation. For example, a Marine

[26] Gottman, 1994; Gottman, 1999.

Table 4
Sample Measures of Effectiveness

Measure of Effectiveness	Definition/Data Source
Short-Term Outcomes	
Completed resume	Proportion of family members receiving career coaching who developed or updated a resume
Skills for using outdoor equipment	Improved knowledge, skills, or abilities related to outdoor equipment. Quiz or measure should be developed based on specific content of skills training.
Ability to understand a Leave and Earnings Statement	Quiz should be developed regarding specific program learning goals related to understanding military pay.
Intermediate Outcomes	
Improved unit cohesion	Ratings of unit social cohesion. Sample measures include Group Cohesion–Social Scale[a] or Cohesion Questionnaire–Company Survey Social Scale.[b]
Increased social support from other military families	Multidimensional Scale of Perceived Social Support[c]
Improved relationship skills	Reduced Sound Relationship House Questionnaire[d]

[a] A. Chang and P. Bordia, "A Multidimensional Approach to the Group Cohesion–Group Performance Relationship," *Small Group Research*, Vol. 32, 2001, pp. 379-405.

[b] M. D. Smith and J. D. Hagman, *Year 1 Assessment of the Unit Focused Stability Manning System*, final report, U.S. Army Research Institute for the Behavioral and Social Sciences, 2004.

[c] G. D. Zimet, N. W. Dahlem, S. G. Zimet, and G. K. Farley, "The Multidimensional Scale of Perceived Social Support," *Journal of Personality Assessment*, Vol. 52, No. 1, 1988, pp. 30–41.

[d] J. M. Gottman, *What Predicts Divorce? The Relationship Between Marital Processes and Marital Outcomes*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1994; J. M. Gottman, *The Relationship Clinic: A Scientifically Based Marital Therapy*, New York: W. W. Norton & Company, 1999.

Corps Order (MCO) may specify that a workshop must be offered a certain number of times in a year, or state that all Marines in a certain role should participate in a program within a certain time frame. These standards may influence the time course of an evaluation. That is, if a program is required to offer a workshop a certain number of times per year, it may be appropriate to collect data over at least a year.

Analyzing Data

When designing an evaluation, it is useful to think about how the results of both the process and outcome components will be analyzed and interpreted once the data have been collected. There are many options for analyzing and interpreting data, and the best option might depend on the goals of the evaluation or the way that it was designed.

Identifying Metrics or Standards of Performance

One consideration is whether there are specific standards, or *metrics*, that a program is required or aims to achieve. As mentioned before, some programs and services have specific requirements related to their implementation, such as offering a workshop a certain number of times per year. There can also be metrics related to the outcomes of a program or service. For example, the goal might be for no Marines to test positive for alcohol use while on duty. Metrics are useful because they provide a standard against which to compare the performance of a

program. Even if there is not a specific metric established by Marine Corps policy, though, a program or service might set a standard that it would like to achieve (e.g., participation by 70 percent of the eligible population).

Not all of the MOPs or MOEs will have a specific target. It may be that the goal is to reach as many family members as possible, or to increase the capacity for services from year to year, but with no specific target attached to the measure. In this case, an evaluator may choose to simply track the measure over time to determine whether there is a trend or pattern, or examine whether participants experienced an improvement from preparticipation to postparticipation.

Analyzing and Interpreting Data

There are three common ways that data may be analyzed and interpreted as part of an evaluation.

1. **Comparison over time.** When tracking an MOP or MOE over time, there are situations in which the evaluator wants to monitor the variable at set intervals to see whether it generally increases or decreases over time (e.g., does average satisfaction with a program increase over consecutive months?), or to see whether a variable changes from one point to another (e.g., changes in participant attitudes from before taking a course to after taking the course). The choice of time period to monitor MOPs/MOEs is dependent on the particular indicator, but could be weekly, monthly, quarterly, or annually.

   Tracking ongoing trends in MOPs and MOEs is an effective way to monitor how a program or service is functioning over time. Favorable trends—increases in positive indicators (e.g., knowledge), decreases in negative indicators (e.g., problem behaviors), or steady trends at acceptable levels (e.g., consistent high attendance)—suggest that the program or service is serving its intended function with the targeted population. However, it is important to keep in mind that favorable trends do not prove that the program or service caused the trend: It could be that some other factor caused the favorable trend (e.g., people could have improved on their own, even without the program; they might have used another service that helped them with their problem).
2. **Pre/post comparison.** Comparing pre-event data to post-event data is similar to examining performance over time, with two main differences. First, there are only two time points (e.g., before and after using the program or service). Second, when trends are tracked over time, the analysis does not necessarily track the same group of people at each time point. Instead, the analysis often provides a snapshot in time, repeated at multiple time points with whoever is using the program or service at that time. With a pre/post comparison, the data are almost always collected on specific individuals to see whether there were changes within the same person. These changes are then averaged over the entire group of participants. Favorable changes in indicators suggest that the program or service is serving its intended function with the targeted population. However, as with examining trends over time, it is important to keep in mind that favorable changes pre- to post-program do not prove that the program or service caused the change, especially if there is no control or comparison group.
3. **Comparison between groups.** There are times when an evaluator wants to compare MOPs or MOEs between groups of people. This is relevant if the evaluation design includes a control or comparison group. Examining differences between groups on MOPs and MOEs is another way to monitor how the program or service is affecting the

target population. For example, finding that program users are different from nonusers on key indicators (e.g., knowledge) suggests that the program had a positive impact on users. If the evaluation has used a rigorous design, such as an experimental design, it is more likely that these changes are due to participation in the program or service. With quasi-experimental designs, it is important to keep in mind that there may be other explanations for the change (e.g., maybe those with more knowledge of family problems were more likely to seek out training on how to cope with family problems).

Step 5. Using the Results of Program Evaluation and Continuous Quality Improvement

After analyzing the data, the next step is to interpret the results and identify patterns that indicate trends, performance outcomes, and effectiveness. This is the stage at which the data can be compared to a standard or metric, or trends over time can be identified.

Program evaluation can be conducted as a one-time occurrence, but the more common purpose of program evaluation is as part of continuous quality improvement efforts.[27] After the evaluation data have been collected and analyzed, the next step is to determine areas of success, areas with room for growth, and areas of struggle that need to be addressed. With this information, the program or service or branch leadership can identify appropriate actions or changes that will address performance issues in future operation of the program. For example, if the evaluation indicates that the program or service has been implemented as intended and resulted in the intended outcomes, then the plan may be to continue implementation and monitor the process and outcomes. If the results indicate that the program or service is not being implemented as expected or is not reaching the target population, the next steps might be to focus on strengthening areas that were not implemented with fidelity, or improve recruiting or retention strategies to ensure that the program reaches its target population. If the program is being implemented as intended but is not achieving the expected results, it may be an indication that there is a misalignment between program activities and intended outcomes, or a mismatch between the program and its participants (e.g., participants started high on outcome measures, leaving limited room for improvement).

Programs and services that operate across multiple installations may respond to program evaluation results in different ways. Local resources and context likely play a role in determining what actions are feasible for program staff to implement. That said, consulting with program managers across installations might provide some guidance on best practices or improved resource utilization, reducing the burden on program staff for identifying and implementing necessary changes.

As these changes are implemented, ongoing evaluation efforts can help to determine whether the changes are improving the implementation or outcomes associated with the program or service. In this way, program evaluation is part of a larger process of monitoring cur-

[27] D. M. Berwick, "Continuous Improvement as an Ideal in Health Care," *New England Journal of Medicine*, Vol. 320, No. 1, 1989, pp. 53–56; W. E. Deming, *The New Economics for Industry, Government, Education*, Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Educational Services, 1994; S. B. Hunter, P. A. Ebener, M. Chinman, A. J. Ober, and Christina Y. Huang, *Promoting Success: A Getting to Outcomes® Guide to Implementing Continuous Quality Improvement for Community Service Organizations*, Santa Monica, Calif.: RAND Corporation, TL-179-NIDA, 2015.

rent performance, enacting corrective action on or modifications to specific aspects of a program, and assessing subsequent changes in program performance and outcomes.

Summary

This document provided a brief introduction to program evaluation. We described the process of identifying program goals and core components, which can be organized with a logic model. We discussed the different types of program evaluation, including process and outcome evaluations, and reviewed considerations related to designing an evaluation, including data sources, timing of the evaluation, availability of resources, and issues related to the protection of evaluation participants.[1] We described MOPs and MOEs, how they relate to the components of the logic model, and how to define these measures. We briefly described the process of implementing an evaluation and interpreting results and discussed how evaluation results can be used to continue to improve a given program or service.

This document is intended to provide a broad overview of the general principles and processes governing program evaluation. As such, it may serve as a resource to programs and services undergoing a program evaluation and provide context for the activities that may be taking place as part of such an evaluation. For those interested in learning more about program evaluation, Appendix A provides the references for additional evaluation resources. In addition, Appendix B provides a series of worksheets that can be used to plan and implement a program evaluation. Some of these worksheets may be helpful to a program or service when used individually, in such tasks as identifying SMART goals or identifying program core components. They can also be used in combination to guide a program evaluation from beginning to end.

Program evaluation is a tool that can be valuable to programs at many stages. These principles can be helpful for new programs just beginning the process of implementation, and to programs that have been in existence for years. Regardless of a program or service's current performance, an understanding of program strengths, weaknesses, and outcomes can help it to operate at its maximal effectiveness.

---

[1] Federal regulations 45 CFR 46.102 (Protection of Human Subjects 2009) and Secretary of the Navy Instruction 3900.39D define *research* as "a systematic investigation including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge." Human subject means a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information. Program evaluations are considered aspects of research and *must* be vetted through the HRPP *before* beginning. The program, staff, or investigators themselves cannot make study determinations and must use the HRPP Applicability Review Process for all potential research and program evaluations.

APPENDIX A

Additional Resources

Cook, T. D., and D. T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Boston, Mass.: Houghton Mifflin, 1979.

Fitzpatrick, J. L., J. R. Sanders, and B. R. Worthen, *Program Evaluation: Alternative Approaches and Practical Guidelines*, 4th ed., Boston, Mass.: Pearson, 2011.

Holliday, S., S. L. Wrabel, T. Trail, M. Tankard, D. Schultz, L. Werber, and S. O. Meadows, *A Guide to Evaluating Marine and Family Division Programs, Services, and Capabilities: An Introduction to Program Evaluation*, Santa Monica, Calif.: RAND Corporation, TL-264-USMC/1, 2021.

Lawton, B., P. R. Brandon, L. Cicchinelli, and W. Kekahio, *Logic Models: A Tool for Designing and Monitoring Program Evaluations* (REL 2014–007), Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific, 2014. As of November 9, 2017:
https://ies.ed.gov/ncee/edlabs/regions/pacific/pdf/REL_2014007.pdf

Patton, Q., *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2015.

Secretary of the Navy Manual 5210.1, *Records Management Manual*, November 2007 (Rev.), Chapter 1.

Secretary of the Navy Instruction 5211.5E, *Department of the Navy (DON) Privacy Program*, December 28, 2005.

Shadish, W. R., T. D. Cook, and D. T. Campbell, eds., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, 2nd ed., Boston, Mass.: Houghton Mifflin, 2002.

Shadish, W. R., T. D. Cook, and L. C. Leviton, *Foundations of Program Evaluation: Theories of Practice*, Newbury Park, Calif.: Sage Publications, 1991.

Weinick, R. M., E. B. Beckjord, C. M. Farmer, L. T. Martin, E. M. Gillen, J. D. Acosta, M. P. Fisher, J. Garnett, G. C. Gonzalez, T. C. Helmus, L. H. Jaycox, K. A. Reynolds, N. Salcedo, N., and D. M. Scharf, *Programs Addressing Psychological Health and Traumatic Brain Injury Among U.S. Military Service Members and Their Families*, Santa Monica, Calif.: RAND Corporation, TR-950-OSD, 2011. As of October 27, 2016:
https://www.rand.org/pubs/technical_reports/TR950.html

W. K. Kellogg Foundation, *Logic Model Development Guide*, Battle Creek, Mich., 2004b. As of June 20, 2017:
https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide

APPENDIX B

Evaluation Worksheets

In this appendix, we provide a set of worksheets designed to help users develop a program evaluation plan. These correspond to each of the steps outlined in this document.

Step 1. Identifying Program Goals and Core Components with a Logic Model

- *Worksheet 1.1. Identifying SMART Goals*: This worksheet helps programs and services to identify their program goals in a manner consistent with the SMART principles.
- *Worksheet 1.2. Resources/Inputs*: This worksheet assists in the development of a logic model by helping programs and services identify program resources and inputs.
- *Worksheet 1.3. Activities*: This worksheet assists in the development of a logic model by helping programs and services describe their activities, including type, frequency, staff, and intended recipient.
- *Worksheet 1.4. Target Population*: This worksheet helps programs and services develop a detailed description of their target population(s).
- *Worksheet 1.5. Outputs*: This worksheet assists in the development of a logic model by helping programs and services identify the outputs associated with each program activity.
- *Worksheet 1.6. Outcomes*: This worksheet assists in the development of a logic model by helping programs and services identify the outcomes associated with each activity, including the time frame for each outcome.
- *Worksheet 1.7. Impact*: This worksheet assists in the development of a logic model by helping programs and services identify the impacts associated with each outcome at the aggregate level.
- *Worksheet 1.8. Context*: This worksheet helps a program or service identify contextual factors that may affect the way a program is delivered or the effectiveness of the program.
- *Worksheet 1.9. Logic Model*: This worksheet provides a logic model template, which programs and services can use to organize the material developed with Worksheets 1.2–1.8.

Step 2. Planning for Evaluation

- *Worksheet 2.1. Identifying the Focus of Your Process Evaluation*: This worksheet helps programs and services determine what aspects of implementation and usage they want to evaluate with the process component of their evaluation.
- *Worksheet 2.2. Evaluation Design Issues*: This worksheet helps programs and services to review some of the considerations that may shape their evaluation design, such as number of participants and timing of the evaluation.
- *Worksheet 2.3. Designing Your Evaluation*: This worksheet provides programs and services with a template to use in designing process and outcome evaluations, including the time frame, methods and data sources, sample, and needed resources.
- *Worksheet 2.4. Evaluation Gantt Chart*: This worksheet provides the template for a Gantt chart, which programs and services can use to create a time line of evaluation tasks.

Step 3. Operationalizing Evaluation Measures

- *Worksheet 3.1. Operationalizing Measures of Performance*: This worksheet provides a template that programs and services can use to identify MOPs associated with each program input, activity, and output. It also provides a space to identify a definition for each measure, the associated metric or time frame for analysis, and the data source.
- *Worksheet 3.2. Operationalizing Measures of Effectiveness*: This worksheet provides a template that programs and services can use to identify MOEs, the way each MOE will be defined, and any notes for use (e.g., how the data will be analyzed or interpreted).

Step 4. Implementing an Evaluation and Interpreting Results

- *Worksheet 4.1. Review Program Outcomes*: This worksheet helps programs and services to organize and interpret the results of their process and outcome evaluations, including whether the results are favorable and whether the implementation, usage, and effects met expectations for the program.

Step 5. Using the Results for Continuous Quality Improvement

- *Worksheet 5.1. Tracking Areas Identified for Program Improvement*: This worksheet helps programs and services to create an action plan based on the results of their evaluation.

Worksheet 1.1
Identifying SMART Goals

<table>
<tr><td colspan="2">What is the mission or purpose of your program or service?

_____
_____
_____</td></tr>
<tr><td>Specific
• What is expected to change?
• How will this be accomplished?
• Who will experience this change?</td><td></td></tr>
<tr><td>Measurable
• How will you measure progress toward this outcome?
• How will you know whether this goal is being achieved?</td><td></td></tr>
<tr><td>Achievable
• Are these changes feasible for the population?
• Does your program have the resources available to support these changes?</td><td></td></tr>
<tr><td>Relevant
• Are the changes you identified related to the mission or purpose of your program?
• Are they related to the activities your program offers?</td><td></td></tr>
<tr><td>Time-bound
• What is the time frame in which you expect to observe effects?
• Are your program's activities tied to specific milestones or events for Marines or family members?</td><td></td></tr>
<tr><td colspan="2">Based on your responses to the questions above, identify your program goal(s):

1. _____
2. _____
3. _____
4. _____</td></tr>
</table>

Worksheet 1.2
Resources/Inputs

Type	Description
Staff	
Facilities	
Funding	
Other	

Worksheet 1.3
Activities

Type	Frequency/Duration	Staff Needed	Intended Recipient

Worksheet 1.4
Target Population

Description (e.g., age, gender, position, area, risk factor)	Eligibility Criteria	Identification and Recruitment

Worksheet 1.5
Outputs

Activity	Outputs (e.g., amount or volume of activity, quality, satisfaction)

Worksheet 1.6
Outcomes

Activity	Outcome (What is expected to change?)	What target population is expected to change?	In what time frame will the change occur?

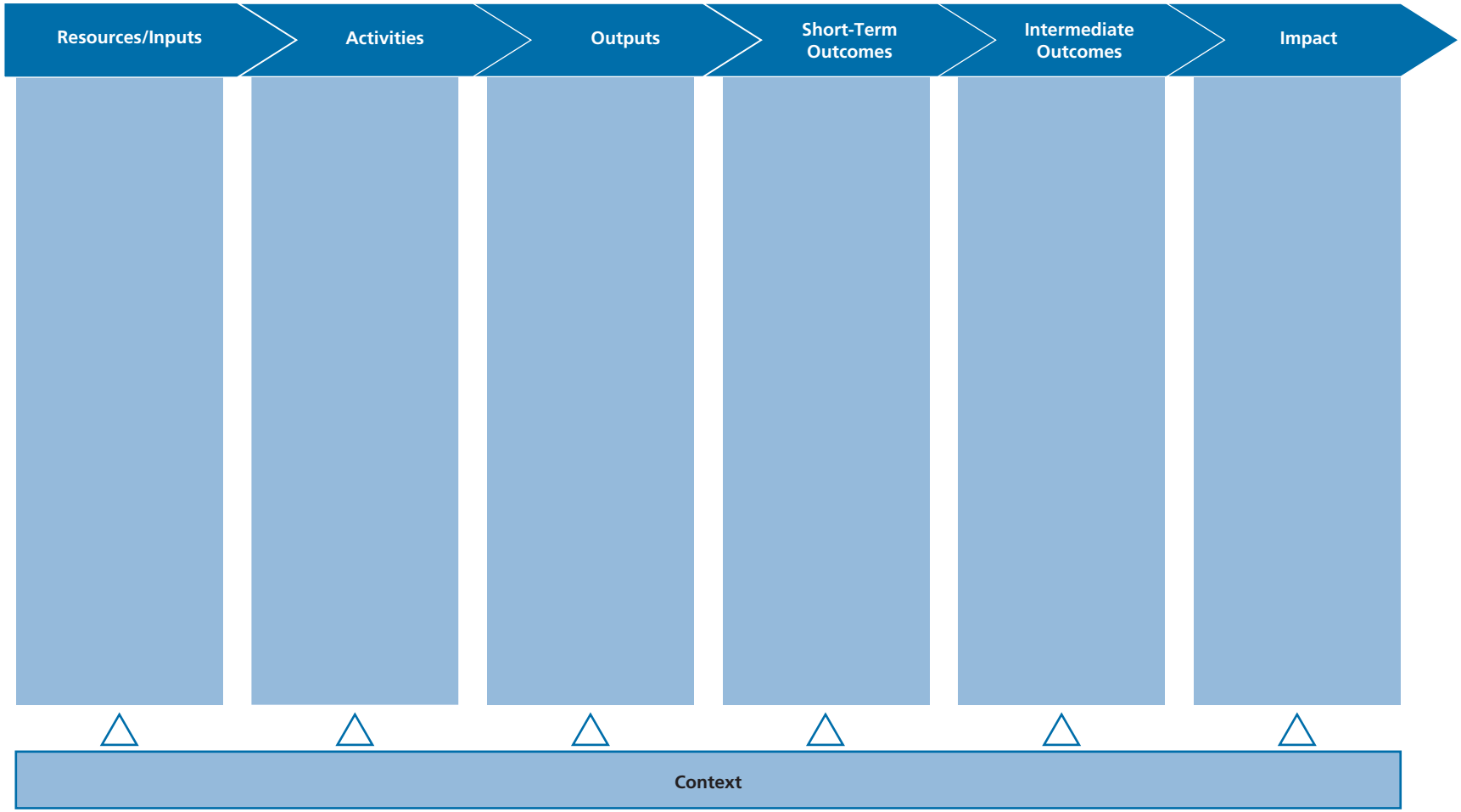
Worksheet 1.7
Impact

Short- or Intermediate-Term Outcome	Expected Impact at the Aggregate Level (e.g., unit, total force, installation)

Worksheet 1.8
Context

Contextual Factor	Area of Impact

Worksheet 1.9
Logic Model



RAND *TL264/1-1.9*

Worksheet 2.1
Identifying the Focus of Your Process Evaluation

☐ Staff qualifications and available resources (resources/ inputs)	Specific aspects (e.g., staff credentials/certifications, availability of resources) • _____ • _____ • _____ • _____
☐ Fidelity of implementation (activities/outputs)	Specific aspects (e.g., adherence to curriculum, appropriate screening for eligibility) • _____ • _____ • _____ • _____
☐ Relevance or usefulness (activities/outputs)	Specific aspects (e.g., activities matched participant needs and interests) • _____ • _____ • _____ • _____
☐ Usage (outputs)	Specific aspects (e.g., attendance, demand met, target population served) • _____ • _____ • _____ • _____
☐ Satisfaction (outputs)	Specific aspects (e.g., among what participants, with which part of the program or service) • _____ • _____ • _____ • _____
☐ Other (outputs)	Specific aspects • _____ • _____ • _____ • _____

Worksheet 2.2
Evaluation Design Issues

Now that you are familiar with some of the issues that may shape your evaluation design, complete the worksheet below. Consider the questions in each section and how they apply to the activity or set of activities that you would like to evaluate for your PSC. Your responses will be useful as you design your evaluation.

Availability of Comparison Group

- Would it be possible to implement a control group that does not participate in this activity?
- If you identified an existing comparison group, what characteristics would you want the group to have (i.e., how should it be similar to the population using your activity)?

Your notes:

Number of Participants

- Will you be using qualitative methods for your process evaluation?
- What is the outcome evaluation design for this activity?
- What degree of change in the outcomes is expected?
- Will you be able to include the population of participants, or will you select a sample (i.e., subset) from this population?
- How many individuals make use of this activity? Over what time frame?
- Do you have access to someone with the statistical skills to help you select the ideal sample size?

Your notes:

Timing of Evaluation

- How long is this activity? If not a set length, how long are participants typically engaged?
- Over what time frame do you expect to observe the short-term effects of this activity?
- Over what time frame do you expect to observe the intermediate effects of this activity?
- On what type of schedule is it possible to collect data? For example, can you collect data (a) before and after a person has participated, (b) only after a person has participated, or (c) on a set schedule (e.g., every 3 months)? Remember to consider the measurement of both process and outcome measures.
- Are there specified reporting standards for this activity (e.g., as outlined in an MCO)?

Your notes:

Evaluation Expertise

- Do PSC staff have evaluation expertise?
- Are there headquarters staff with relevant evaluation expertise?
- Are there other organizations (e.g., academic institutions, community-based organizations) that could assist with an evaluation effort?

Your notes:

Available Evaluation Resources

- What staffing resources does your PSC have available to support the evaluation?
- What financial resources are available?
- What infrastructure is available (e.g., for data collection)?

Worksheet 2.2—Continued

Your notes:

Data Security

- What type of data would you like to collect? Is any of the information particularly sensitive?
- Would you need to collect data over multiple time points, or be able to link responses to a specific person for any other reason?

Your notes:

Evaluating Contracted Services

- What oversight do you have over the contractor, and what activities is it responsible for?
- What data security concerns arise?
- Who will be responsible for evaluation activities, and how will information be communicated between your PSC and the contractor?

Your notes:

Worksheet 2.3
Designing Your Evaluation

Section 1: Time Frame		
Expected start date: _____		
Expected end date: _____		
Section 2: Methods/Data Sources		
Process Evaluation		
Data Source/Method (e.g., attendance records, focus groups, personnel records, satisfaction surveys)	Available now	To be implemented
_____	☐	☐
_____	☐	☐
_____	☐	☐
_____	☐	☐
_____	☐	☐
_____	☐	☐

Data collection schedule:	
Data Source/Method	Timing

Outcome Evaluation		
Data Source/Method (e.g., knowledge tests, surveys of attitudes or skills, behavioral measures)	Available now	To be implemented
_____	☐	☐
_____	☐	☐
_____	☐	☐
_____	☐	☐
_____	☐	☐
_____	☐	☐

Data collection schedule

Data Source/Method	Pre-Participation	During Participation	Post-Participation	Follow-Up (specify)
Example: Knowledge test	*X*		*X*	*X (6 months later)*

Worksheet 2.3—Continued

Section 3: Selecting/Describing Sample		
Process Evaluation		
Data Source/Method	Sample Description	Expected Sample Size
Outcome Evaluation		
Data Source/Method	Sample Description	Expected Sample Size

Section 4: Resources Needed	
A) Staffing Needs	
Evaluation activity (e.g., develop recruitment and retention plan, design data collection forms/instruments, administer surveys)	**Staff member responsible**
_____	_____
_____	_____
_____	_____
_____	_____
B) Infrastructure Needs	
Data collection: _____	
Data storage: _____	
Data analysis: _____	
Reporting results: _____	
C) Additional Costs	
Identify additional costs associated with the evaluation beyond typical PSC implementation (e.g., additional staff, data storage, or management systems).	
_____	
_____	
_____	
_____	

Worksheet 2.4
Evaluation Gantt Chart

	Timeline											
Task												

Worksheet 3.1
Operationalizing Measures of Performance

Input/Activity/ Output	Measure of Performance	Measure Definition	Metric or Time Frame	Data Source

Worksheet 3.2
Operationalizing Measures of Effectiveness

Measure of Effectiveness	Measure Definition	Notes for Use (e.g., comparison group, time frame for data collection/analysis, metrics)

Worksheet 4.1
Review Program Outcomes

Adapted from Acosta et al., 2013, and Hunter et al., 2015

Desired Outcome for MOP/ MOE (Value, Difference, or Change)	Comparison Conducted	Actual Outcome for MOP/MOE		Met Expectations?	Action Needed?
□ MOP □ MOE ***Description:***	□ Continuous comparison over time □ Pre/post comparison □ Comparison between groups □ Comparison with metric ***Description:***		***What is the trend?*** □ Better □ Same □ Worse	***Did this meet your expectations for the program?*** □ Met □ Missed □ Exceeded	□ Yes □ No
□ MOP □ MOE ***Description:***	□ Continuous comparison over time □ Pre/post comparison □ Comparison between groups □ Comparison with metric ***Description:***		***What is the trend?*** □ Better □ Same □ Worse	***Did this meet your expectations for the program?*** □ Met □ Missed □ Exceeded	□ Yes □ No
□ MOP □ MOE ***Description:***	□ Continuous comparison over time □ Pre/post comparison □ Comparison between groups □ Comparison with metric ***Description:***		***What is the trend?*** □ Better □ Same □ Worse	***Did this meet your expectations for the program?*** □ Met □ Missed □ Exceeded	□ Yes □ No
□ MOP □ MOE ***Description:***	□ Continuous comparison over time □ Pre/post comparison □ Comparison between groups □ Comparison with metric ***Description:***		***What is the trend?*** □ Better □ Same □ Worse	***Did this meet your expectations for the program?*** □ Met □ Missed □ Exceeded	□ Yes □ No

Worksheet 5.1
Tracking Areas Identified for Program Improvement

Performance Area	Priority Level	Action(s) to Be Taken	Rationale (if No Action)	Expected Completion	Person or Group Responsible

References

Acosta, J. D., R. N. Ramchand, A. Becker, A. Felton, and A. Kofner, *RAND Suicide Prevention Program Evaluation Toolkit*, Santa Monica, Calif.: RAND Corporation, TL-111-OSD, 2013. As of January 13, 2021: https://www.rand.org/pubs/tools/TL111.html

Berwick, D. M., "Continuous Improvement as an Ideal in Health Care," *New England Journal of Medicine,* Vol. 320, No. 1, 1989, pp. 53–56.

Centers for Disease Control and Prevention, *Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide*, U.S. Department of Health and Human Services, 2011.

Chang, A., and P. Bordia, "A Multidimensional Approach to the Group Cohesion–Group Performance Relationship," *Small Group Research,* Vol. 32, 2001, pp. 379-405.

Chinman, M., P. S. Imm, and A. H. Wandersman, *Getting to Outcomes™ 2004: Promoting Accountability Through Methods and Tools for Planning, Implementation, and Evaluation*, Santa Monica, Calif.: RAND Corporation, TR-101-CDC, 2003. As of January 13, 2021: https://www.rand.org/pubs/technical_reports/TR101.html

Code of Federal Regulations, Title 45, Public Welfare, Department of Health and Human Services, Part 46: Protection of Human Subjects, revised January 15, 2009, effective July 14, 2009.

Daugherty, L., L. Werber, K. Kamarck, L. Harrington, and J. Gazis, *Officer Accession Planning: A Manual for Estimating Air Force Officer Degree Requirements*, Santa Monica, Calif.: RAND Corporation, TL-196-AF, 2016. As of January 13, 2021: https://www.rand.org/pubs/tools/TL196.html

Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury, *Program Evaluation Guide*, Arlington, Va.: Defense Health Agency, 2012.

Department of Defense Instruction 3216.02, *Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research*, November 8, 2011.

Deming, W. E., *The New Economics for Industry, Government, Education*, Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Educational Services, 1994.

Gottman, J. M., *What Predicts Divorce? The Relationship Between Marital Processes and Marital Outcomes*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1994.

———, *The Relationship Clinic: A Scientifically Based Marital Therapy*, New York: W. W. Norton & Company, 1999.

Holliday, S., S. L. Wrabel, T. Trail, M. Tankard, D. Schultz, L. Werber, and S. O. Meadows, *A Guide to Evaluating Marine and Family Division Programs, Services, and Capabilities*, Santa Monica, Calif.: RAND Corporation, TL-264-USMC, 2021. As of January 13, 2021: https://www.rand.org/pubs/tools/TL264.html

Hourcade, J., P. Parette, and H. Anderson, "Accountability in Collaboration: A Framework for Evaluation," *Education and Training in Developmental Disabilities*, Vol. 38, No. 4, 2003, pp. 398–404.

Hox, J. J., and H. R. Boeije, "Data Collection, Primary Versus Secondary," in K. K. Leonard, ed., *Encyclopedia of Social Measurement*, Amsterdam, Netherlands: Elsevier, 2005, pp. 593–599.

Hunter, S. B., P. A. Ebener, M. Chinman, A. J. Ober, and Christina Y. Huang, *Promoting Success: A Getting to Outcomes® Guide to Implementing Continuous Quality Improvement for Community Service Organizations*, Santa Monica, Calif.: RAND Corporation, TL-179-NIDA, 2015. As of January 13, 2021: http://www.rand.org/pubs/tools/TL179.html

Lalayants, M., "Differential Program Evaluation Model in Child Protection," *Child Welfare,* Vol. 91, No. 4, 2012, pp. 9–40.

LeBlanc, P., and C. Lacey, *Peace Works: A Program Evaluation Model*, paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, La., 2002.

Lesesne, C. A., K. M. Lewis, C. Moore, D. Fisher, D. Green, and A. Wandersman, *Promoting Science-Based Approaches to Teen Pregnancy Prevention Using Getting to Outcomes: Draft 2011*, unpublished manual, 2011.

Marine Corps Order 3900.18, *Human Research Protection Program (HRPP)*, January 21, 2011.

Marine Corps Order 5300.18, *Marine Corps Survey Program*, May 21, 2015.

Martin, L. T., C. Farris, D. M. Adamson, and R. M. Weinick, *A Systematic Process to Facilitate Evidence-Informed Decisionmaking Regarding Program Expansion: The RAND Toolkit, Volume 3*, Santa Monica, Calif.: RAND Corporation, RR-487/3-OSD, 2014. As of January 13, 2021: https://www.rand.org/pubs/research_reports/RR487z3.html

McNeil, R. C., "A Program Evaluation Model: Using Bloom's Taxonomy to Identify Outcome Indicators in Outcomes-Based Program Evaluations," *Journal of Adult Education*, Vol. 40, No. 2, 2011, pp. 24–29.

National Center for Education Evaluation and Regional Assistance, *Identifying and Implementing Educational Practices Supported by Rigorous Evaluation: A User Friendly Guide*, Washington, D.C.: Institute of Education Sciences, 2003.

Rossi, P. H., M. W. Lipsey, and H. E. Freeman, *Evaluation: A Systematic Approach*, 7th ed., Thousand Oaks, Calif.: Sage Publications, 2004.

Secretary of Navy Instruction 3900.39D, *Human Research Protection Program*, November 6, 2006.

Smith, M. D., and J. D. Hagman, *Year 1 Assessment of the Unit Focused Stability Manning System*, final report, U.S. Army Research Institute for the Behavioral and Social Sciences, 2004.

Teddlie, C., and F. Yu, "Mixed Methods Sampling: A Typology with Examples," *Journal of Mixed Methods Research*, Vol. 1, No. 1, 2007, pp. 77–100.

Unger, D. G., E. A. Park, P. Antal, P. A. Tressell, K. Rigney, K. DeRasmo, and J. Kassess, "Serving Children with Special Social and Emotional Needs: A Practical Approach to Evaluating Prevention Programs in Schools and Community Settings," *Journal of Educational & Psychological Consultation*, Vol. 11, No. 2, 2000, pp. 273–296.

U.S. Marine Corps, Marine Corps Human Research Protection Program (HRPP) and Institutional Review Board, website, undated. As of January 13, 2021: https://www.tecom.marines.mil/Resources/Marine-Corps-Human-Research-Protection-Program/

Wholey, J. S., H. P. Hatry, and K. E. Newcomer, eds., *Handbook of Practical Program Evaluation* (3rd ed.), San Francisco, Calif.: Jossey-Bass, 2010.

W. K. Kellogg Foundation, *W.K. Kellogg Foundation Evaluation Handbook,* Battle Creek, Mich., 2004.

World Bank, *Doing Impact Evaluation Number 6: Data for Impact Evaluation*, Washington, D.C.: Poverty Reduction and Economic Management, Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation, World Bank, 2009.

World Health Organization, *WHO Evaluation Practice Handbook*, Geneva, Switzerland, 2013.

Zimet, G. D., N. W. Dahlem, S. G. Zimet, and G. K. Farley, "The Multidimensional Scale of Perceived Social Support," *Journal of Personality Assessment*, Vol. 52, No. 1, 1988, pp. 30–41.

The U.S. Marine Corps' Marine and Family Program Division (MF) portfolio includes a variety of programs and services that cut across a number of domains, including behavioral health, personal and professional development, and family readiness. Given increased emphasis on budgetary constraints and efficient spending, assessing the implementation and effectiveness of programs and services has gained significance. Program evaluation is the application of systematic processes for collecting and analyzing information about a program to determine the program's quality of operation and ability to meet intended purposes (i.e., effectiveness).

The goal of this document is to provide an introduction to program evaluation for MF program managers. Although some program managers may have experience with aspects of program evaluation, the process may be less familiar to others. Therefore, this document provides a broad overview of the process of conducting a program evaluation, including identifying program goals and core components, planning for an evaluation, operationalizing program evaluation measures, implementing an evaluation and interpreting results, and how to use the results of an evaluation for continuous quality improvement. For program staff interested in more detail about the process, this review provides a reference list of resources, as well as a series of worksheets that can be used to plan and implement a program evaluation.

A more detailed version of this guide, which includes a set of Excel worksheets that program evaluators may use to guide their own evaluation efforts, can be found in the RAND Marine and Family Division Program, Service, and Capability Evaluation User's Guide (Holliday et al., 2018).

NATIONAL DEFENSE RESEARCH INSTITUTE

www.rand.org

$22.00

ISBN-10 197740770-6
ISBN-13 9781977407702

TL-264/1-USMC