



Assessing Impact to Inform Decisions

A Toolkit on Measures for Policymakers

Scott Savitz, Miriam Matthews, Sarah Weiland



For more information on this publication, visit www.rand.org/t/TL263

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

ISBN: 978-0-8330-9879-5

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2017 RAND Corporation

RAND® is a registered trademark.

Cover image: Tam Kadraj/Fotolia

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND

Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Preface

This report provides a brief overview of how to identify what to measure with respect to programs and policies, how to evaluate measures using specified criteria, and how to identify new measures that can contribute to improved decisionmaking, assessment, planning, and communication. It is a summary of material presented at a series of instructional workshops in 2016 and 2017 to help inform personnel from the Office of the Under Secretary of Defense for Policy about how to better understand measurement issues for policies and programs.

This research was sponsored by the Office of the Under Secretary of Defense for Policy and conducted within the International Security and Defense Policy Center and the Homeland Security and Defense 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 International Security and Defense Policy Center, see www.rand.org/nsrd/ndri/centers/isdp or contact the director (contact information provided on the web page).

Contents

Preface	iii
Figures and Tables	vii
Summary	ix
Acknowledgments	xi
1. Introduction	1
2. Identifying What Should Be Measured Using Logic Models	3
3. Mapping Existing Measures to the Logic Model	7
4. Evaluating Existing Measures	9
5. Identifying and Evaluating New Measures	13
6. Key Challenges in Working with Measures	15
7. Key Points to Remember	17
Appendix	19
Bibliography	21

Figures and Tables

Figures

2.1.	Structure of a Logic Model.....	4
2.2.	A Highly Simplified Example of a Logic Model Applied to a University.....	5
A.1.	Blank Logic Model.....	19

Tables

3.1.	Measures Worksheet.....	7
4.1.	Examples of Criteria and Scales for Evaluating Measures.....	9
4.2.	Measures Worksheet with Criteria Evaluated.....	10
A.1.	Blank Measures Worksheet.....	20

Summary

Well-designed measures are essential for informing plans, decisions, assessments, and communications with key stakeholders. This report provides a very brief overview of how to identify concepts that should be measured, how to evaluate metrics, and how to identify new measures that can contribute to better-informed decisions, plans, and assessments. It is based on the content presented in a series of instructional, interactive workshops that RAND conducted for the Office of the Under Secretary of Defense for Policy in 2016 and 2017. This content was derived from a series of prior RAND reports, as discussed in the body of the document.

Acknowledgments

We would like to thank the many people who contributed to the workshops on which this report is based. Clark Cully and Benj Bain played a central role in sponsoring this project and helping to shape the workshops. Jerry Walsh, Rachel Billingslea, Stephanie Burchard, Michael Donofrio, James Mitre, and Elizabeth Phu all enabled us to engage with their respective offices. Our RAND colleagues who designed materials for the workshops and facilitated them—Kate Anania, Aaron Davenport, Mark Hvizda, Natasha Lander, Andrew Lauand, Lauren Mayer, Henry Willis, Anna Jean Wirth, Emmi Yonekura, and Rebecca Zimmerman—all made valuable contributions. We also greatly appreciate the insights of Seth Jones, Mike McNerney, and Chris Chivvis, who provided oversight for this effort. Our reviewers, Todd Helmus and Chris Paul, provided valuable feedback that helped to improve the document, as did Sarah Meadows.

1. Introduction

Measures provide ways of assessing something. While most people think of measures as quantitative, such as the height of a building, they can also be qualitative, such as whether a process has high, medium, or low efficiency. Measures are essential for informing plans, decisions, assessments, and communications with key stakeholders. In all of these contexts, policymakers or program personnel should be able to characterize what has happened or is happening, as well as how particular choices are likely to influence future events. The measures used may not be explicitly cited as such, and some of them may not be perceived as measures, but they always play a central role in any informed decisionmaking, planning, or assessment.

It is often challenging to select appropriate measures to measure the effectiveness of policies and programs. A clear description of the ultimate goals of those policies or programs is required, as well as a characterization of how the policies or programs will contribute to those goals. Ultimate goals may be somewhat abstract ideas, such as “stability” or “deterrence,” that do not have obvious measures associated with them. Moreover, the effects of a policy or program may be difficult to disentangle from other factors influencing the situation it aims to affect.

The Office of the Under Secretary of Defense for Policy (OSD-Policy), recognizing these challenges, wanted its personnel to be better able to develop, select, evaluate, and use measures. To that end, it asked a RAND team to conduct a series of instructional, interactive workshops on measures for OSD-Policy personnel. Those individuals would then be able to better use measures to provide relevant insights that would inform their own assessments and decisions. Moreover, they would be able to provide more-useful measurement-based information to their top leadership and external stakeholders, fostering better assessments and decisions by others.

This toolkit summarizes the content shared at those workshops, which were held in 2016 and 2017. The intent is to provide policymakers with a very concise overview of key ideas that they can readily use; further information can be found in the references. While many other documents discuss measures, and this material represents a summary of other sources, the fact that so many policymakers have expressed a keen interest in the workshops and in having a ready reference suggests that there is an unmet need for a highly compact source of information on this topic. The workshops were based primarily on an amalgamation of methodologies from a number of prior RAND reports, including the following:¹

¹ Ben Connable, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, Santa Monica, Calif.: RAND Corporation, MG-1086-DOD, 2012; Victoria A. Greenfield, Valerie L. Williams, and Elisa Eiseman, *Using Logic Models for Strategic Planning and Evaluation: Application to the National Center for Injury Prevention and Control*, Santa Monica, Calif.: RAND Corporation, TR-370-NCIPC, 2006; Christopher Paul, Brian J. Gordon, Jennifer D. P. Moroney, Lisa Saum-Manning, Beth Grill, Colin P. Clarke, and Heather Peterson, *A Building Partner Capacity Assessment Framework: Track-*

- *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List, Desk Reference, and Handbook for Practitioners*
- *A Building Partner Capacity Assessment Framework: Tracking Inputs, Outputs, Outcomes, Disrupters, and Workarounds*
- *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*
- *Enhancing U.S. Coast Guard Metrics*
- *Establishing Law and Order After Conflict*
- *Measuring Cooperative Biological Engagement Program (CBEP) Performance: Capacities, Capabilities, and Sustainability Enablers for Biorisk Management and Biosurveillance*
- *Using Logic Models for Strategic Planning and Evaluation: Application to the National Center for Injury Prevention and Control.*

Collectively, these documents discuss four key tasks associated with measures:

- identifying what should be measured, using structures called “logic models”
- mapping existing sets of measures to specific elements of those logic models
- critically evaluating those measures using clearly defined criteria
- identifying new measures to complement or supplant existing measures, and critically evaluating those measures using the same criteria.

The workshops provided an interactive forum for participants to learn how to perform these four tasks. In this report, we briefly describe the same material. The report (like the workshops) is intended to provide a very short overview of the subject, on which many volumes have been written. A few of the many sources of further information—one of which is an annotated reading list—appear in the references section at the end of the document.

The remainder of this report is structured as follows, reflecting the four tasks in the preceding bullet list. Section 2 describes how to develop logic models to identify what should be measured. Since most policies or programs already have some existing measures (even if these are not explicit), section 3 explains how to map these measures to the logic models described in the previous section. Section 4 describes how to evaluate measures using standard criteria, while section 5 describes how to identify and evaluate new measures. In section 6, we present a handful of key challenges that users of measures should recognize, while in section 7, we present some key points to remember. Following this, the appendix provides a brief how-to guide, including blank templates for users to employ.

ing Inputs, Outputs, Outcomes, Disrupters, and Workarounds, Santa Monica, Calif.: RAND Corporation, RR-935-OSD, 2015; Christopher Paul, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*, Santa Monica, Calif.: RAND Corporation, RR-809/3-OSD, 2015a; Christopher Paul, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, Santa Monica, Calif.: RAND Corporation, RR-809/1-OSD, 2015b; Christopher Paul, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence and Persuade: Handbook for Practitioners*, Santa Monica, Calif.: RAND Corporation, RR-809/2-OSD, 2015; Scott Savitz, Henry H. Willis, Aaron Davenport, Martina Melliand, William Sasser, Elizabeth Tencza, and Dulani Woods, *Enhancing U.S. Coast Guard Metrics*, Santa Monica, Calif.: RAND Corporation, RR-1173-USCG, 2015; Seth G. Jones, Jeremy M. Wilson, Andrew Rathmell, and K. Jack Riley, *Establishing Law and Order After Conflict*, Santa Monica, Calif.: RAND Corporation, MG-374-RC, 2005; Stephanie Young, Henry H. Willis, Melinda Moore, and Jeffery Engstrom, *Measuring Cooperative Biological Engagement (CBEP) Performance: Capacities, Capabilities, and Sustainability Enablers for Biorisk Management and Biosurveillance*, Santa Monica, Calif.: RAND Corporation, RR-660-OSD, 2014.

2. Identifying What Should Be Measured Using Logic Models

The first step in evaluating and identifying new measures is determining what they should measure. A common and effective way of evaluating what should be measured is to develop a structure called a logic model, which describes the relationships that link a policy or program's resources with its ultimate goals. In addition to helping to improve the use of measures, a logic model can also aid in policy decisions, by helping to clarify how those decisions can influence different portions of the model. Previous RAND research defines logic models as follows:¹

A logic model typically offers a simplified visual representation of the path of a program's operations, starting with inputs and then progressing to the program's activities, its outputs, its customers, and its intended outcomes. The model may also link the program's operations, what the program actually does either alone or with others to fulfill its mission, to its strategy, which we define as the goals, management objectives, and performance measures.

A sample structure of a logic model is shown in Figure 2.1.

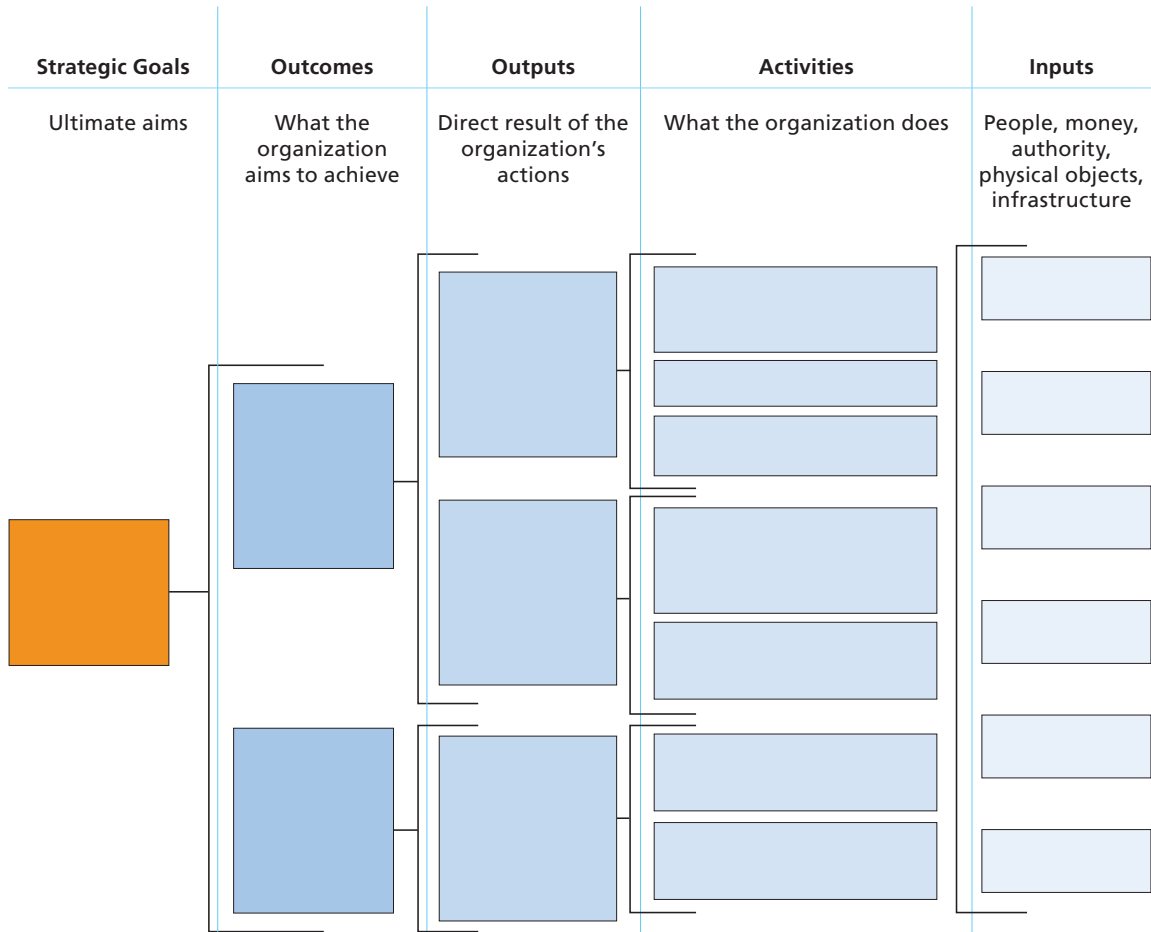
Going from right to left, an organization uses *inputs* (e.g., resources) to perform *activities*. These activities result in *outputs* that achieve *outcomes* (e.g., program achievements, observed changes). These outcomes contribute to *strategic goals* that are being pursued and influenced by a variety of forces, most of them beyond the control of the program or organization. Each block shown within the logic model is termed an *element* of the logic model, while each column (e.g., inputs or activities) is a *category*.

For example, a Department of Defense program may use *inputs* such as funding and personnel-hours to perform *activities* such as training of partner-nation forces. This training is intended to result in more capable partner-nation units (the *output*). It is anticipated that these more capable units will contribute to the *outcome* of more effectively countering an extremist insurgency, which contributes to the ultimate *strategic goal* of fostering stability in a particular nation or region. Obviously, this is a simplified example, but it is meant to illustrate the process.

Alternatively, we could characterize a university using a logic model. A highly simplified example is shown in Figure 2.2. The university leverages diverse *inputs* to conduct *activities* such as teaching, publication, and research. These result in several *outputs*: more capable students, more available information, and new information. These, in turn, contribute to the *outcomes* of dissemination and creation of knowledge. Finally, these collectively contribute to *strategic goals*, including a more successful, informed, and enlightened society. Obviously, the

¹ Greenfield, Williams, and Eiseman, 2006.

Figure 2.1
Structure of a Logic Model



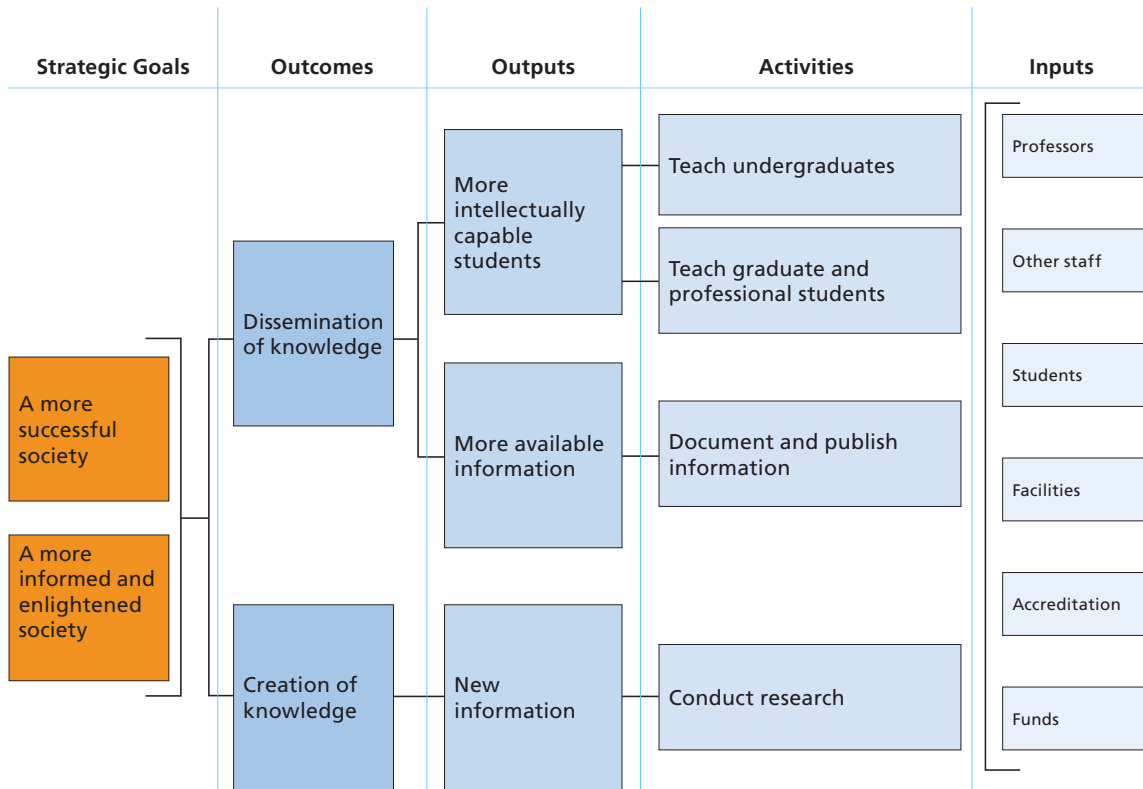
RAND TL263-2.1

university is not wholly responsible for, nor entirely capable of, fulfilling those strategic goals on its own; however, its ultimate aims are to contribute to those strategic goals.

Ideally, what an organization does should ultimately derive from what it is trying to achieve, so development of a logic model for a new policy or program should begin with strategic goals. In practice, even for a new program, other elements of the logic model may already be defined (e.g., funding, authority, activities). When developing a logic model for an existing program or policy, it is important to clarify whether it is intended to be *descriptive*, characterizing the way it is now, or *prescriptive*, the way that policymakers would like it to be.

Logic models come in many varieties and often use slightly different vocabulary to describe categories similar to those listed above. Categories may be added, subtracted, combined, or divided to reflect the desired level of detail. Just as British and American English have different spellings and pronunciations but can be used to communicate information in similar ways, the same is true of different types of logic models. Some examples of logic models can be found in the references mentioned in the previous section of this document. A couple of them, *RAND Program Evaluation Toolkit for Countering Violent Extremism* and *Assessing and Evalu-*

Figure 2.2
A Highly Simplified Example of a Logic Model Applied to a University



RAND TL263-2.2

ating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference, provide lengthy discussion on developing logic models.²

Developing Logic Models

Those developing logic models to evaluate and identify measures can draw on several sources during development, in addition to their own knowledge of the program or policy and its context. One is documentation. Formally disseminated high-level documents, or transcripts of high-level speeches, are particularly important in terms of authoritatively characterizing strategic goals and outcomes. Otherwise, there is a risk of getting into debates about either substance or semantics regarding what the program or organization is actually expected to do. Even key leaders within longstanding organizations sometimes have distinct perspectives on goals (which is one reason why goals sometimes evolve as leadership changes); moreover, goals can evolve over time in response to a changing environment. One of the advantages of developing a logic model is that it can help stakeholders to articulate the relationships that it delineates.

² Paul et al., 2015b, chapter 5; Todd Helmus, Miriam Matthews, Rajeev Ramchand, Sina Beaghley, David Stebins, Amanda Kadlec, Michael A. Brown, Aaron Kofner, and Joie Acosta, *RAND Program Evaluation Toolkit for Countering Violent Extremism*, Santa Monica, Calif.: RAND Corporation, TL-243-DHS, 2017, chapter 2.

An explicit logic model or similar construct helps individuals and small groups reach a shared understanding of goals and how they are to be attained.

Naturally, it is also important to consult with a range of stakeholders and experts. Some of these will be internal to the program, while others will be external individuals such as customers, resource providers, and partners. Depending on the relationships you have with them, and how you anticipate the group dynamics would work, you may want to query a group of individuals together. This promotes discussion and can lead to program insights. It is also very important to get input from people who are actually doing the activities but are not helping directly to develop the strategic plans or the logic models; they will likely have insights regarding how those activities are conducted that no one else will. There may also be data sets, such as spreadsheets characterizing activities or outputs, which can inform the development of logic models.

Developing a logic model is an inherently iterative process that should include review throughout the lifetime of a program. It may involve a series of individuals doing research, sharing what they have with each other, and then having workshop sessions in which they discuss possible items to include in the logic model. Draft logic models can then be reviewed by others, discussed in subsequent workshops, and so on. The process will likely be messy, with disconnects and contradictions among the various stakeholders about what is happening now, what should be happening, and where boundaries of authority and responsibility lie within the organization. This is normal, and part of the reason that iteration is necessary. Elements of the logic model may need to be revised for the sake of clarity or, in some cases, combined or divided.

There are many approaches to developing a logic model, and each individual or group needs to determine the best way forward. Usually, the hardest part is characterizing the items on the far left of Figure 2.1: strategic goals, outcomes, and outputs. Strategic goals are typically identified based on high-level documents, such as the National Strategy of the United States, while outcomes are identified based on organizational mission statements, statements by senior leadership, and insights from subject-matter experts. Outputs may be identified based not only on program documents, but also on interviews with operators who accomplish them (as well as other stakeholders and experts). Working from the left to the right begins with characterizing what the policy or program is intended to accomplish, then linking these goals to activities and inputs (which are usually relatively tangible and intuitive). In some cases, it may be easier to start from the right-hand side, characterizing inputs and activities, so that the team developing the logic model can achieve a quick victory and start to build momentum. There may be cases in which it makes sense to start in the middle and work outward, or to start at both ends and work toward the center. Regardless, development of a logic model requires thoughtful consideration of how to link what is being done (or should be done) with given resources to achieve overall goals. While identifying the elements of the logic model, aim to make them SMART: Specific, Measurable, Achievable, Relevant, and Time-bound.

Logic models typically become more branched as we move to the right. Each outcome is linked to multiple outputs, and each output is linked to multiple activities. However, multiple strategic goals may be collectively supported by a group of outcomes, without clear lines of demarcation indicating that certain outcomes support only one or some of the strategic goals. Also, some inputs (such as money, authority, personnel, or physical assets) may be used for more than one activity.

3. Mapping Existing Measures to the Logic Model

Many programs and policies already have evaluation measures, even if they are not explicitly called measures. Information regarding the characteristics of a program or policy that is shared within an organization, or with external stakeholders, generally contains sets of measures. Compiling a list of such measures is a natural first step in understanding how programs or policies are currently being measured.

The next step is to examine each measure and select the element of the logic model that it measures (this is called *mapping* the measures to the logic model). In some cases, this will be clear; for example, the measure “number of illicit weapons interdicted” can be readily mapped to a specific activity within a logic model, “interdict illicit weapons.” Some elements of the logic model may have multiple measures associated with them. Mapping may also be less obvious than in the above example; such cases typically require discussion and analysis among several stakeholders. There will also be instances in which an existing measure does not measure any element of a logic model. This likely indicates either that the measure is not very useful, or that the logic model should be reviewed to evaluate whether it is missing important elements. Both possibilities should be considered.

It may be useful to develop a measures worksheet like the one shown in Table 3.1, listing the measures in the left-hand column. (Setting up such a table in an electronic spreadsheet format is fast and would enable the results to be easily circulated.) In the second column, enter the element of the logic model being measured, and in the third column, list which category this belongs to (strategic goal, outcome, output, etc.). We will address the remaining four col-

Table 3.1
Measures Worksheet

Measure	Element of the Logic Model It Measures (if any)	Category (input, activity, etc.)	Validity	Reliability	Feasibility	Timeliness
1. Annual number of citations of work by university researchers	Creation of knowledge	Outcome				
2. Annual number of books and journals published by the university press	More available information	Output				
3. Annual number of person-hours spent on research	Teach undergraduates	Activity				

umns in the next section, on evaluating measures. We have inserted some examples of measures into the table, following up on our earlier logic model regarding a university.

In many cases, the relevance of specific measures for decisionmaking will vary depending on who is making what decisions and in what context. For example, there may be measures primarily for use within an office (measuring inputs, activities, and outputs), while some external stakeholders may be more interested in outcome measures. The specifics will vary from situation to situation.

4. Evaluating Existing Measures

Once you have mapped your existing measures to the logic model, the next step is to evaluate the measures using well-defined criteria. As with the choices about how to structure your logic model, there are a number of criteria that you can use.¹ For the sake of accessibility, we will focus on four intuitive criteria:

- **validity**—the extent to which the measure accurately measures an element of the logic model
- **reliability**—how consistently measurements can be made
- **feasibility**—how easily the measurement can be made (the quantity of resources required to make the measurement)
- **timeliness**—how quickly the measurement can be made.

In Table 4.1, we show a simple three-point scale to characterize whether a particular measure rates high, medium, or low using these criteria.

For example, let us assume that the element of a logic model being evaluated is the extent of interethnic violence in a particular area over time. One measure being used to track this might be the number of detected postings on social media that contain ethnically hostile messages. The validity of such a measure would be low: While there might be some correlation between the number of postings and the extent of physical violence, the link is ambiguous.

Table 4.1
Examples of Criteria and Scales for Evaluating Measures

	Validity	Reliability	Feasibility	Timeliness
High	Directly measures the element or a close proxy	Well-defined, objective, and stable	Required data sets are readily available and user-friendly	Shorter than the decisionmaking time line
Medium	Closely related to the element being measured	Some ambiguity, subjectivity, and/or volatility	Required data sets could be collected with limited effort	Similar to the decisionmaking time line
Low	Indirectly related to the element being measured	Considerable ambiguity, subjectivity, and/or volatility	Required data sets would be challenging to collect	Exceeding the decisionmaking time line

¹ For example, the SMART criteria mentioned earlier (specific, measurable, achievable, relevant, and time-bound) can also be applied to measures. The criteria that we described (validity, reliability, feasibility, and timeliness) may be easier for a beginner to use, although VRFT is a harder acronym to pronounce.

Reliability may be medium, since there may be uncertainty regarding the degree to which all postings are being counted, and there may be some ambiguity over whether particular postings contain hostile messages. (It may be difficult to divine whether “I wish all my friends from the other main religious group a festive holiday and encourage them to stay safe” is a positive message or a veiled threat.) If algorithms can search such postings automatically and tally the results, then the measure has high feasibility. If human beings are needed to scroll through each posting and make a determination, the measure has low feasibility. The measure has high timeliness, since postings can be assessed and counted immediately.

To continue with the earlier example of a university, we have expanded on Table 3.2, including evaluation of the measures with respect to the criteria listed earlier, in Table 4.2. The number of citations of work by the university’s researchers is an indicator of the extent to which it is contributing to the creation of knowledge; it is related to what it aims to measure but is not a direct measurement of it. For that reason, we have designated it as having medium validity. This measure has high reliability and feasibility, since an automated search of academic databases should reveal consistent results with limited effort. Its timeliness may be limited by delays in the updating of databases, so the results for a given year may not be known for weeks. If we assume that decisions may need to be made in a comparable time line, that would give it a medium result for timeliness.

The annual number of books and journals published is a good indicator of the extent to which the university is contributing to more available information, giving it high validity. (The logic model does not speak to the quality or insightfulness of that information; a more nuanced real-world example could include a finer characterization, which could also require a more sophisticated measure.) Multiple observers would likely find the same value for this measure, which they could readily derive from university records within a short time frame, giving it high reliability, feasibility, and timeliness.

Finally, the annual number of person-hours spent on research is a good indicator of the extent of the activity of conducting research, with high validity. However, finding out how many hours each person spends would require a survey that would take time and resources to administer, giving it low feasibility. Although a survey would eventually give a precise answer, there would likely be questions about whether the people who did not respond to the survey would have given similar responses to those who did, and whether respondents might have exaggerated the hours they spent (perhaps unintentionally), so observers might judge the value

Table 4.2
Measures Worksheet with Criteria Evaluated

Measure	Element of the Logic Model It Measures (if any)	Category (input, activity, etc.)	Category			
			Validity	Reliability	Feasibility	Timeliness
1. Annual number of citations of work by university researchers	Creation of knowledge	Outcome	M	H	H	M
2. Annual number of books and journals published by the university press	More available information	Output	H	H	H	H
3. Annual number of person-hours spent on research	Conduct research	Activity	H	M	L	M

to be different from the one the survey reported, reducing reliability. Finally, a survey could take weeks to administer and collect data, resulting in a medium value for timeliness.

Considerations in Adjusting Evaluations Using Criteria

The criteria listed in Table 4.1 (validity, reliability, feasibility, timeliness) could also be adjusted based on situational specifics. For example, feasibility may be a function of the extent of resources available: What is highly feasible for a well-endowed program may have low feasibility for one with few resources. When characterizing a slowly evolving phenomenon (such as the effectiveness of a public-health program), a high value for timeliness could correspond to one year, with a medium value corresponding to two to three years, and a low value corresponding to a decade or more. It depends on the time line over which key decisions need to be made with respect to the program.

Naturally, more precise scales could be used, with five or more values (e.g., low, low-medium, medium, medium-high, high). However, such scales may not be more accurate, given that there may be ambiguity and some subjectivity about the extent to which a measure meets one or more of the criteria. Unfortunately, the evaluation of measures may itself involve limited reliability; there may be some disagreements among subject-matter experts regarding where a particular measure rates with respect to one or more of these criteria. If these experts tend to extreme disagreements, with some designating a measure as “high” and others as “low,” that may be indicative of a lack of clarity: The measure, or the element of the logic model to which it maps, may be imprecisely worded. Alternatively, there may be deep underlying differences in various individuals’ understanding of some aspect of the policy or program. Discussing why the parties disagree can lead to clarification, and sometimes to important insights regarding their different perceptions.

An easy way to capture the results of this evaluation is to enter it into the last four columns of Table 3.1 (or an electronic version of it). One of the advantages of doing this in a spreadsheet is that it is easy to sort the measures based on how well they perform on different criteria. For example, it is easy to sort them to see which measures rate highly on all four criteria, and which ones have low ratings for some or most of them.

Naturally, these four criteria are not all equal in terms of importance. A measure that has low validity is unlikely to be a desirable one, regardless of its other merits. The relative importance of the others is somewhat specific to the context of the policy or program. For example, high reliability may be particularly important if a program or policy is highly controversial, but medium reliability may be acceptable in other circumstances. The degree to which feasibility matters may be a function of whether there is the potential to garner more resources, while the importance of timeliness may be a function of whether there are discrete or continuous decision cycles.

5. Identifying and Evaluating New Measures

A review of existing measures, once these have been mapped and evaluated, can reveal two main categories of deficiencies. The first is whether some elements of the logic models are not currently being measured. If these elements of the logic models are relevant for decisionmaking, planning, assessment, or communication, there is a need to identify additional measures that measure these elements of the logic model. The second is that existing measures may have low values for validity, reliability, feasibility, and/or timeliness. While validity is the most important criterion, low values for the others can also contribute to a search for alternative or complementary measures. For example, a measure with low reliability may be so subjective as to be uninformative. One with low feasibility may require more resources for data collection than the program can accommodate, while one with low timeliness may not be available until after critical decisions have been made. A broad rule of thumb is that if a measure has a low value with respect to any of these criteria, it probably should not be used, unless that measure is the least bad way in which to measure the element.

New measures can be variations on existing measures. For example, an existing measure may count the absolute number of something; in many cases, dividing that absolute number by an appropriate denominator may provide a more informative percentage or rate. For example, the number of provincial capitals that have achieved a given level of security may be more informative if that number is divided by the total number of provincial capitals. In some cases, there may be ways of extrapolating from an existing measure to a new measure that is more comprehensive or that describes similar phenomena. For example, a measure of interest may be the number of annual exercises involving multiple services that meet or exceed a given size (measured in terms of the number of personnel involved). Another useful measure in this vein would be the number of bilateral exercises involving specific partner nations that meet or exceed that size.

Alternatively, new measures can be entirely novel. These are often developed by brainstorming, whether individually or in groups. Naturally, there are no good general rules for developing such measures, other than to be creative. These measures can then be evaluated using specific criteria, as described in the previous section, at which point those that have faults can be winnowed out.

6. Key Challenges in Working with Measures

There are a number of challenges involved in working with measures. We summarize a few of these below. (Others are described in many of the references listed at the end of this document.)

- **Analysts and policymakers may be tempted to focus on input and activity measures, without adequately developing measures that measure outputs, outcomes, and strategic goals.** It is usually easy to characterize a policy's or program's effectiveness in terms of the inputs (such as funding or person-hours) that went into it, or the extent of activity undertaken. It is harder to characterize its impact in terms of outputs, outcomes, or strategic goals (such as "strengthening deterrence" or "reducing intercommunal tensions"). Given limited resources with which to develop and implement measures, as well as a desire to simplify communications, it will always be tempting to use and disseminate measures that measure inputs and activities. Awareness of the need to use measures that measure the impact of a policy or program can help to counterbalance this tendency. While it may not be necessary to have explicit measures that measure every individual element of a logic model, measures that relate to the outputs, outcomes, and strategic goals of a program can play a key role in informing decisions about how or whether to continue a program or policy.
- **Analysts and policymakers may be tempted to use single measures to characterize complex phenomena.** Given limited time lines in which to communicate and make decisions, there may be a desire to use a single measure that aggregates a number of others: for example, designating a program as "green," "yellow," or "red." This may not always provide enough detail for policymakers to make informed decisions. If such abridgment is necessary due to limited leadership availability, it should always be based on a traceable analysis of multiple supporting measures. That analysis can be provided to people who immediately support leadership.
- **Complex policies or programs may have multiple, competing outcomes or strategic goals.** In some cases, this may require multiple logic models. In other cases, it may require assessments of trade-offs among conflicting outcomes or strategic goals. For example, a policy or program may be intended both to improve the security environment in another nation, and to enable that nation's own police and military forces to operate independently. Evaluation of the program's or policy's success needs to take into account the relative importance of these two aims, since they may be in conflict with one another.
- **Complex interaction of multiple actors' policies and programs in a dynamic environment.** Policies and programs are almost never operating in isolation and can only sometimes be tested in a rigorous experimental fashion in which one or two changes are

made while all other aspects of the environment are held constant. Natural phenomena ranging from earthquakes to droughts, the actions of adversarial forces, and the policies/programs of other agencies or allied nations can all confound efforts to determine the impact of a policy or program. Reasoned analyses about why the values of measures are changing, or how they are likely to change if policies shift, need to take into account the actions and reactions of other parties. They also need to incorporate uncertainty regarding both the present state and the future, to include both human dynamics and the natural environment.

7. Key Points to Remember

In this brief document, we provide an overview of how to develop logic models, how to map measures to those logic models, how to evaluate measures, and how to identify new measures. Naturally, this is a starting point for developing skills in all of these areas. The items mentioned in the bibliography section that follows can help readers to further improve their skills and insights in these areas. Ultimately, the most important contributor to becoming better in these areas is practice.

A few key points to remember:

- To develop a good set of measures, it is necessary to first develop logic models or similar constructs that explicitly link inputs with ultimate goals.
- Developing logic models requires iterative review of documents, as well as interviews and consultation with subject-matter experts and stakeholders.
- Measures relating to outputs, outcomes, and strategic goals are hard to develop, but they are also critical for assessment and future resource-allocation decisions.
- If an existing measure does not measure an element of the logic model, it may be because the measure is not very useful, or the logic model is missing an element.
- Measures can be qualitatively evaluated using standard criteria, such as validity, reliability, feasibility, and timeliness, among others.
- Complex phenomena typically require sets of measures to characterize them, rather than single measures.

Finally, readers should be cognizant of the fact that logic models and measures are never perfect ways of characterizing the policies or programs that they represent. These tools inform human judgment and should evolve over time, to reflect greater insights, the changing nature of the phenomena that they characterize, and the shifting informational needs of decision-makers. Once logic models and sets of measures have been developed, they can be revisited on a regular basis, perhaps annually. This work can be paused until the next iteration based on decisionmaking needs and cycles.

Bibliography

The following documents represent a small subset of the many sources available regarding logic models and measures, as well as specific instances in which they were developed. They represent a starting point for further reading, rather than a comprehensive guide.

Banks, Robert, "A Resource Guide to Public Diplomacy Evaluation," in *CPD Perspectives on Public Diplomacy*, Los Angeles, Calif.: Figueroa Press, 2011.

Centers for Disease Control and Prevention, Division for Heart Disease and Stroke Prevention, "Evaluation Guide: Developing and Using a Logic Model," Washington, D.C.: U.S. Department of Health and Human Services. As of June 29, 2017:

https://www.cdc.gov/dhdsp/programs/spha/evaluation_guides/logic_model.htm

Connable, Ben, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, Santa Monica, Calif.: RAND Corporation, MG-1086-DOD, 2012. As of June 29, 2017:

<http://www.rand.org/pubs/monographs/MG1086.html>

Donaldson, Stuart, "Theory-Driven Program Evaluation in the New Millennium," in S.I. Donaldson and M. Scriven, eds., *Evaluating Social Programs and Problems: Visions for the New Millennium*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2003, pp. 111–142.

Doran, George T., "There's a S.M.A.R.T. Way to Write Management's Goals and Objectives," *Management Review*, November 1981, pp. 35–36.

Funnell, Sue, and Patricia Rogers, *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*, Thousand Oaks, Calif.: SAGE Publications, 2011.

Greenfield, Victoria A., Valerie L. Williams, and Elisa Eiseman, *Using Logic Models for Strategic Planning and Evaluation: Application to the National Center for Injury Prevention and Control*, Santa Monica, Calif.: RAND Corporation, TR-370-NCIPC, 2006. As of June 29, 2017:

http://www.rand.org/pubs/technical_reports/TR370.html

Hatry, Harry P., *Performance Management: Getting Results*, Washington, D.C.: Urban Institute Press, 2007.

Helmus, Todd, Miriam Matthews, Rajeev Ramchand, Sina Beaghley, David Stebbins, Amanda Kadlec, Michael A. Brown, Aaron Kofner, and Joie Acosta, *RAND Program Evaluation Toolkit for Countering Violent Extremism*, Santa Monica, Calif.: RAND Corporation, TL-243-DHS, 2017. As of June 29, 2017:

<https://www.rand.org/pubs/tools/TL243.html>

Jones, Seth G., Jeremy M. Wilson, Andrew Rathmell, and K. Jack Riley, *Establishing Law and Order After Conflict*, Santa Monica, Calif.: RAND Corporation, MG-374-RC, 2005. As of June 29, 2017:

<http://www.rand.org/pubs/monographs/MG374.html>

Kenrick, Douglas T., Steven L. Neuberg, and Robert B. Cialdini, *Social Psychology: Unraveling the Mystery*, Third Edition, New York: Pearson, 2005.

Knowlton, Lisa Wyatt, and Cynthia C. Phillips, *The Logic Model Guidebook: Better Strategies for Great Results*, 2nd ed., Washington, D.C.: SAGE Publications, 2013.

"Logic Model Workbook," Washington, D.C.: Innovation Network, Inc. As of September 7, 2016:

http://www.innonet.org/client_docs/File/logic_model_workbook.pdf

Paul, Christopher, Brian J. Gordon, Jennifer D. P. Moroney, Lisa Saum-Manning, Beth Grill, Colin P. Clarke, and Heather Peterson, *A Building Partner Capacity Assessment Framework: Tracking Inputs, Outputs, Outcomes, Disruptors, and Workarounds*, Santa Monica, Calif.: RAND Corporation, RR-935-OSD, 2015. As of June 29, 2017:

http://www.rand.org/pubs/research_reports/RR935.html

Paul, Christopher, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*, Santa Monica, Calif.: RAND Corporation, RR-809/3-OSD, 2015a. As of June 29, 2017:

http://www.rand.org/pubs/research_reports/RR809z3.html

———, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, Santa Monica, Calif.: RAND Corporation, RR-809/1-OSD, 2015b. As of June 29, 2017:

http://www.rand.org/pubs/research_reports/RR809z1.html

Paul, Christopher, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence and Persuade: Handbook for Practitioners*, Santa Monica, Calif.: RAND Corporation, RR-809/2-OSD, 2015. As of June 29, 2017:

http://www.rand.org/pubs/research_reports/RR809z2.html

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach*, Thousand Oaks, Calif.: SAGE Publications, 2004.

Savitz, Scott, Henry H. Willis, Aaron Davenport, Martina Melliand, William Sasser, Elizabeth Tencza, and Dulani Woods, *Enhancing U.S. Coast Guard Metrics*, Santa Monica, Calif.: RAND Corporation, RR-1173-USCG, 2015. As of June 29, 2017:

http://www.rand.org/pubs/research_reports/RR1173.html

Shadish, William R., Thomas D. Cook, and Laura C. Leviton, *Foundations of Program Evaluation: Theories of Practice*, Newbury Park, Calif.: SAGE Publications, 1991.

“Using Logic Models to Bring Together Planning, Evaluation, and Action: Logic Model Development Guide,” Battle Creek, Mich.: W. K. Kellogg Foundation, 2004. As of September 7, 2016:

<http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide>

Williams, Valerie L., Elisa Eiseman, Eric Landree, and David M. Adamson, *Demonstrating and Communicating Research Impact: Preparing NIOSH Programs for External Review*, Santa Monica, Calif.: RAND Corporation, MG-809-NIOSH, 2009. As of June 29, 2017:

<http://www.rand.org/pubs/monographs/MG809.html>

Young, Stephanie, Henry H. Willis, Melinda Moore, and Jeffery Engstrom, *Measuring Cooperative Biological Engagement (CBEP) Performance: Capacities, Capabilities, and Sustainability Enablers for Biorisk Management and Biosurveillance*, Santa Monica, Calif.: RAND Corporation, RR-660-OSD, 2014. As of June 29, 2017:

http://www.rand.org/pubs/research_reports/RR660.html

Well-designed measures are essential for informing plans, decisions, assessments, and communications with key stakeholders. This toolkit provides a very brief overview of how to identify concepts that should be measured, how to evaluate measures, and how to identify new measures that can contribute to better-informed decisions, plans, and assessments. It is based on the content presented in a series of instructional, interactive workshops that RAND conducted for the Office of the Under Secretary of Defense for Policy in 2016 and 2017. This content was derived from a series of prior RAND reports, as discussed in the body of the document.



NATIONAL DEFENSE RESEARCH INSTITUTE

www.rand.org

\$14.00

ISBN-10 0-8330-9879-9
ISBN-13 978-0-8330-9879-5

