MEASURING FOR RESULTS

Key Concepts for Understanding the Performance of DHS Programs and Activities

1 October 2014

TASK LEAD
Erin Goodykoontz

TASK TEAM
Jerry Carbonette
Kristen Klein, PhD
Sarah Nashed
Adrian Smith
HOMELAND SECURITY STUDIES AND ANALYSIS INSTITUTE

The Homeland Security Act of 2002 (Section 305 of PL 107-296, as codified in 6 U.S.C. § 185) authorizes the Secretary of the Department of Homeland Security (DHS), acting through the Under Secretary for Science and Technology, to establish one or more federally funded research and development centers (FFRDCs) to provide independent analysis of homeland security issues. Analytic Services Inc. operates the Homeland Security Studies and Analysis Institute (HSSAI) as an FFRDC for DHS under contract HSHQDC-09-D-00003.

HSSAI provides the U.S. Government with the necessary expertise to conduct: crosscutting mission analysis, strategic studies and assessments, development of models that baseline current capabilities, development of simulations and technical evaluations to evaluate mission trade-offs, creation and evolution of high-level operational and system concepts, development of top-level system and operational requirements and performance metrics, operational analysis across the homeland security enterprise, and analytic support for operational testing evaluation in tandem with the government’s acquisition process. HSSAI also works with and supports other Federal, state, local, tribal, public and private sector organizations that make up the homeland security enterprise.

HSSAI’s research is undertaken by mutual consent with the DHS and is organized as a set of discrete tasks. This report presents the results of research and analysis conducted under

Task 13-02.01.02, Performance Management Methodology Development

The purpose of this task is to research and analyze concepts and methods for measuring performance in the DHS context. It is intended to produce a resource to inform the way that DHS program managers and other stakeholders think about challenging topics in performance measurement, particularly the definition of program outcomes linked to strategic goals or program objectives.

The results presented in this report do not necessarily reflect official DHS opinion or policy.
MEASURING FOR RESULTS

Key Concepts for Understanding the Performance of DHS Programs and Activities

1 October 2014

TASK LEAD
Erin Goodykoontz

MANAGER, POLICY AND PROGRAM ANALYSIS DIVISION
Howard A. Smith, PhD

TASK TEAM
Jerry Carbonette
Kristen Klein, PhD
Sarah Nashed
Adrian Smith

DEPUTY DIRECTOR, HSSAI
Robert V. Tuohy

RESEARCH DIRECTOR
M. Jo Velardo, PhD

For more information, contact us at performance@hsi.dhs.gov
# Contents

Acronym List ........................................................................................................... vii
Preface ...................................................................................................................... ix

PART I:  
A Case and a Conceptual Framework for Program Performance Measurement ................................................................. 1

1. Introduction ....................................................................................................... 3
   What Are Measures? ......................................................................................... 4
   What Is Measuring for Results? ...................................................................... 4
   Why Is It Important? ....................................................................................... 5
   What Are the Challenges? ............................................................................. 6
   About This Resource ..................................................................................... 7

2. A Conceptual Framework .............................................................................. 11
   Purpose ............................................................................................................ 13
   Inputs ............................................................................................................. 14
   Processes ...................................................................................................... 14
   Outputs ......................................................................................................... 15
   Outcomes ..................................................................................................... 15
   External Factors ........................................................................................... 16
   Benefits of Using the Framework ................................................................. 17
   What’s Next? ............................................................................................... 18

PART II:  
Measuring for Results ........................................................................................ 21

3. The Purpose-Driven Program: Identifying Desired Outcomes ................................................................. 23
   Where to Find Your Program’s Purpose ....................................................... 25

4. Understanding Outcomes ............................................................................ 31
   Types of Outcomes: Thinking Vertically and Horizontally ....................... 32
   Horizontal Dimension .................................................................................. 33
   Vertical Dimension ....................................................................................... 34
   Measuring across Both Horizons ................................................................. 36
5. Defining Your Program’s Outcomes ...........................................39

Key Terms and Concepts: Constructs and Indicators.........................39
Measurement Challenges......................................................................44
Straightforward Outcomes.................................................................45
Fuzzy Outcomes.................................................................................45
Hidden Outcomes................................................................................46
Measurement Challenges at the Intersection of Fuzzy and Hidden Concepts.................................................................................48
Methods for Measuring Challenging Concepts.....................................49

6. Narrowing the Field: Deciding Which Outcomes and Indicators to Measure..............................................................53

The Return on Investment (ROI) of Measurement..........................53
Constraints on Measurement.................................................................54
Externally Imposed Limitations..........................................................54
Resource Limitations............................................................................55
Measurement Fatigue............................................................................56
A Constraint – and an Opportunity......................................................56
Deciding What to Measure: Scoping Your Indicators.......................57
Discriminating Power: Success or Failure..........................................58
Preordained Indicators.........................................................................59
Who Will Use the Information, and How?..........................................60
Audience Level of Knowledge..............................................................60
Primary and Secondary Data: Some Key Tradeoffs...........................63

7. Special Challenges in Measurement ...........................................67

The Causality Issue...............................................................................68
Unintended Consequences.....................................................................72
Shared Program Ownership..................................................................74

8. Conclusion.......................................................................................79

APPENDIX A: Glossary.........................................................................83

APPENDIX B: Authorities, Guidance, and GAO and DHS Office of the Inspector General (OIG) Commentary on DHS Program Measurement......................................................89

DHS OIG Reports................................................................................90
GAO Reports.........................................................................................91

APPENDIX C: The Program Logic Model ..........................................93

APPENDIX D: Bibliography ...............................................................97
# Acronym List

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBP</td>
<td>Customs and Border Protection</td>
</tr>
<tr>
<td>DHS</td>
<td>Department of Homeland Security</td>
</tr>
<tr>
<td>DHS ETMS</td>
<td>DHS Enterprise Talent Management System</td>
</tr>
<tr>
<td>DHS OIG</td>
<td>DHS Office of the Inspector General</td>
</tr>
<tr>
<td>FAMS</td>
<td>Federal Air Marshal Service</td>
</tr>
<tr>
<td>FBI</td>
<td>Federal Bureau of Investigation</td>
</tr>
<tr>
<td>FEMA</td>
<td>Federal Emergency Management Agency</td>
</tr>
<tr>
<td>FLETC</td>
<td>Federal Law Enforcement Training Center</td>
</tr>
<tr>
<td>FY</td>
<td>Fiscal year</td>
</tr>
<tr>
<td>GAO</td>
<td>Government Accountability Office</td>
</tr>
<tr>
<td>GPRA</td>
<td>Government Performance and Results Act of 1993</td>
</tr>
<tr>
<td>GPRAMA</td>
<td>GPRA Modernization Act of 2010</td>
</tr>
<tr>
<td>HSSAI</td>
<td>Homeland Security Studies and Analysis Institute</td>
</tr>
<tr>
<td>ICE</td>
<td>Immigration and Customs Enforcement</td>
</tr>
<tr>
<td>ISA IPC</td>
<td>Information Sharing and Access Interagency Policy Committee</td>
</tr>
<tr>
<td>KPI</td>
<td>Key performance indicator</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>OMB</td>
<td>Office of Management and Budget</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>QHSR</td>
<td>Quadrennial Homeland Security Review</td>
</tr>
<tr>
<td>ROI</td>
<td>Return on investment</td>
</tr>
<tr>
<td>S&amp;T</td>
<td>Science and Technology Directorate (DHS)</td>
</tr>
<tr>
<td>SIPOC</td>
<td>Suppliers, Inputs, Processes, Outputs, and Customers</td>
</tr>
<tr>
<td>SLTT</td>
<td>State, local, tribal, territorial government</td>
</tr>
<tr>
<td>SPOT</td>
<td>Screening of Passengers by Observation Technique Program</td>
</tr>
<tr>
<td>TSA</td>
<td>Transportation Security Administration</td>
</tr>
<tr>
<td>USCG</td>
<td>U.S. Coast Guard</td>
</tr>
<tr>
<td>USCIS</td>
<td>U.S. Citizenship and Immigration Services</td>
</tr>
</tbody>
</table>
Preface

In a memorandum dated April 22, 2014, Department of Homeland Security (DHS) Secretary Jeh Johnson argued for “better traceability between strategic objectives, budgeting, acquisition decisions, operational planning, and mission execution.” This memorandum also highlighted the need for a process, such as DHS’s Integrated Investment Life Cycle Management, that connects and explains the relationships “between interrelated strategy, capabilities and resources, programming and budgeting, and major acquisition oversight processes.” As a result of this memorandum and other shaping forces, DHS leadership is prioritizing strategic performance measurement to help assess and shape the results of DHS programs. For instance, DHS Science and Technology (S&T) Directorate Undersecretary Reginald Brothers testified a few months after the memorandum’s release about the importance of long-term strategic planning and concrete metrics for success in S&T.

Beyond the Secretary’s memorandum, the Government Performance and Results Act of 1993 (GPRA) and the GPRA Modernization Act of 2010 (GPRAMA) also motivate DHS performance measurement. Notably, GPRA and GPRAMA mandate strategic planning and performance measurement at the department or agency level, rather than the individual program or activity level. However, performance measurement at these lower levels is just as important.

---

2 Ibid.
importance as measurement for DHS as a whole. By measuring individual programs’ results, program managers can contribute sound and meaningful performance data to DHS’s performance measurement and management efforts and also drive improvements and other program management activities for their own programs.

Despite the value of DHS program performance measurement, guidance on the subject is currently sparse. Some instructions exist on specific aspects of performance measurement (most notably, the *DHS Performance Measure Verification and Validation Handbook*) or on performance measurement in specific domains (see, for instance, the *DHS Information Sharing and Safeguarding Strategy*). However, unified guidance for DHS program performance measurement does not appear to be readily available.

Given the importance and current scarcity of DHS-wide guidance on program performance measurement, we developed this resource for use among DHS program managers and other stakeholders interested in performance measurement, such as DHS leadership or those responsible for creating a performance measurement approach for an organization. We hope that this resource will facilitate discussion and help establish a consistent thought framework for performance measurement throughout the department.

This resource is intended to serve as both a guide and a reference to inform users on how to effectively measure performance throughout a program or activity’s life cycle. It does not impose any new requirements on program managers; rather, this resource may be viewed as complementary to what program managers and other stakeholders are already doing in line with GPRA, GPRAMA, and other authorities on performance measurement.

This resource will provide an intuitive framework to help you think through the many elements of your program and ultimately ensure that you’re measuring the results that matter. This resource also offers some useful tips to help you measure some of DHS’s most difficult-to-measure outcomes, such as deterrence, risk, and resilience. Taken together, this guidance can help program managers to establish a results-based measurement system that reflects alignments between program outcomes and strategic objectives,
goals, or missions. Developing measures that reflect these alignments is consistent with the spirit of the Secretary’s “unity of effort” memorandum and helps to ensure that your performance measures are meaningful and traceable.

Performance measurement is a broad topic that is closely related to, although importantly distinct from, similar concepts such as performance management or metrics development (see chapter 1 for more on these concepts). The purview of this resource is the “middle lane” of performance measurement. Although it includes a cursory discussion of some introductory concepts and issues in performance management and metrics, the vast majority of this resource is focused on how to think about performance measurement. That is, how do you identify, define, and scope the program outcomes you want to measure? These conceptual-level decisions can be quite challenging; it’s important to give them equal consideration, rather than short shrift, as compared to performance management decisions and the more tactical process of developing and deploying metrics.

---

Figure 1. A pyramid showing the relationships between performance management, performance measurement, and metrics development and deployment. The purview of this resource is primarily in the performance measurement block.
RESOURCE ROAD MAP

Readers may use this document in whole or in part to inform their performance measurement decisions. Accordingly, we provide a couple of different “views” for readers to understand how the guidebook is organized and quickly navigate to the most relevant passages for their specific needs. The table of contents above provides descriptive chapter titles to give readers a quick-look capture of each chapter’s contents. In addition, the following paragraphs provide a slightly more descriptive, high-level overview of each chapter.

Part I of this resource presents a case and a conceptual framework for results-based program performance measurement. In chapter 1, we discuss key terms such as performance measurement and performance management and discuss the importance of measuring for results. In chapter 2, we present a conceptual framework that can help you think through the many elements of your program and ensure that you are measuring outcomes (among other things).

Part II of this resource takes a deeper dive into concepts and principles that will help you measure the outcomes that matter. In chapter 3, we discuss the importance of measuring outcomes that are aligned with your program’s purpose and refer you to key resources and considerations for identifying and defining your program’s purpose. In chapter 4, we discuss some critical differences between outputs and outcomes, as well as the major types of outcomes you can measure for your program. Chapter 5 presents key concepts and guidance to help you define and measure some of the more difficult-to-measure outcomes you may encounter in the DHS context. In chapter 6, we discuss various measurement costs and constraints and offer some practical considerations to help you make scoping decisions for your measurement system. In chapter 7, we provide an overview and some DHS examples of challenges associated with performance measurement, along with some guidance and additional resources to help you think about and address these challenges. Finally, chapter 8 brings the concepts together by returning to the conceptual framework from Part I and walking through a summary example that integrates numerous concepts found throughout this resource.
A Case and a Conceptual Framework for Program Performance Measurement
In recent years, federal programs have encountered a new paradigm emphasizing results-focused program performance measurement and management. Many forces have contributed to this trend, including a culture of continual improvement, federal budget uncertainty and constraints, increased calls for government accountability and transparency, and society’s demands for government efficiency in providing better services while simultaneously reducing the deficit. In addition to these factors, recent policy guidance mandates a number of activities for federal departments and agencies in measuring and managing their programs’ performance. The most notable of these policies are the Government Performance and Results Act of 1993 (GPRA) and the GPRA Modernization Act of 2010 (GPRAMA), which mandate – among other things – that federal departments and agencies establish clear alignments between their performance goals and strategic priorities.\(^7\) Numerous Government Accountability Office (GAO) reports followed the release of this guidance, documenting challenges and opportunities for performance measurement and management across multiple federal departments and agencies.\(^8\)

In this context, it has become increasingly important for program managers to measure for results, to understand their programs’ goals, resources, and activities and be able to realize their true impact. The purpose of this resource is to help DHS program managers and other stakeholders with the critical and challenging task of developing sound, outcome-focused performance

---


\(^8\) See appendix B for additional discussion of the applicable authorities and guidance regarding performance measurement.
measures for their programs. In Part I, we make a case and present a conceptual framework for results-based program performance measurement.

**What Are Measures?**

When we talk about measures, we are referring to the quantitative and qualitative information collected on key aspects of a program. Measures can focus on things such as quality, cost, timeliness, customer satisfaction, or efficiency. For example, managers can use measures to:

- Report how many things are produced or services are provided
- Communicate whether a program is operating efficiently
- Report whether or not customers are satisfied
- Describe the quality of the products or services provided
- Demonstrate whether the program is achieving the intended outcomes

Measures can be used to inform key program management decisions such as setting program priorities, allocating resources, or identifying program problems and taking corrective action to solve them. Measures enable program managers to make evidence-based decisions that can improve or sustain their program’s results. Performance measurement is therefore an essential element of effective program management.

**What Is Measuring for Results?**

When we say measuring for results, we’re adapting GAO’s usage of the term managing for results to a performance measurement context. Measuring for results is about monitoring the vital indicators of a program’s outcomes in a way that enables decision makers to be aware of and improve their results. We use this term to remind the reader that the primary focus of performance measurement should be assessing outcomes that reflect the goals (intended or desired outcomes) of the program. Measuring for results enables program managers to assess whether the program’s actual results are aligned with its intended results, and make any necessary changes to improve this alignment. Measuring for results is a common-sense approach to performance measurement that links the program’s purpose with its actual results.

---

WHY IS IT IMPORTANT?

In order to know and show a program’s results, program managers first need to measure results. Leadership and other program stakeholders will be looking for “proof” that a program is producing the results it’s intended to produce. Solid data on important program results can provide the evidence stakeholders seek. Moreover, in the current fiscal environment, a program’s continued existence may hinge on the availability of such outcome data. If key program outcomes are aligned with leadership’s strategic priorities, the program may have a better chance of surviving any budget cuts that may occur, enabling the program to achieve its objectives.

Despite the importance of measuring program outcomes, other program elements – particularly outputs – may lend themselves more readily to measurement. Program outcomes are often relatively abstract concepts, such as resilience or deterrence. It can be difficult to identify valid and meaningful measures of these complex, multifaceted concepts. Moreover, even if valid and meaningful measures of outcomes can be found, these measures can often be quite difficult, time consuming, or expensive to implement. By contrast, program outputs – for instance, drug interdiction quantities or the number of grants administered – are often fairly simple to identify and measure. This can sometimes lead program managers to measure the “low-hanging fruit” – outputs – at the expense of measuring true program outcomes.

The problem is that outputs are not always simply connected to the program’s outcomes.\(^\text{10}\) If a program administers more grants, does this result in grantees being more resilient? Do drug interdictions have any effect on deterrence? If program managers only measure outputs, the data will not tell them or their stakeholders what they really need to know: whether the program is accomplishing what it was intended to accomplish. Measuring program results, and measuring them well, is vital to effective performance measurement – and ultimately to effective performance management.

For instance, if program managers have only output data, they may see a negative pattern and propose program overhauls to try to reverse this trend. However, such output-focused changes may amount to “rearranging deck chairs on the Titanic” if the changes have little or no effect on the desired program outcomes. Or – perhaps worse – maybe the outcomes were fine to begin with, and program changes weren’t necessary at all; rather, the output


WHY MEASURE OUTCOMES?

Measuring outcomes is critical for assessing the success or progress of your program toward its goals. Although output measures can tell you what goods or services your program is producing, they can’t tell you whether those goods or services are getting you any closer to achieving your program’s goals. Sound measures of the outcomes that matter can provide you with evidence of your program’s achievements.
data was misleading because the outputs were poorly aligned with key program outcomes. These scenarios are not improbable, and they highlight the importance of measuring for results.

Program managers can be most effective when they have a clear idea of their program’s purpose and measures that tell them whether the program is fulfilling that purpose. Results-based measures provide program managers the tools they need to observe and achieve these results.

WHAT ARE THE CHALLENGES?

Although a clear value proposition and mandate for sound program performance measurement exist, these “why should we” arguments do not answer the “how to” questions that program managers are likely to face when implementing measurement systems for their programs that are linked to strategic goals. Some of these questions include:

- How do I determine what is actually important to measure?
- How do I know if my measures are outputs or outcomes? How are the two different?
- Is it worth measuring inputs or processes?
- What external factors are likely to affect my measurements?
- How can I properly identify the expected outcomes for my program?
- What if I have no control over the intended purpose or strategic direction of my program? Should I still measure my program’s outcomes?
- What can I do to understand how my program fits into larger strategic goals across DHS?
- If I’m measuring a challenging topic like resilience, how can I break the topic down to better understand it? What methods can I use to measure these fuzzy concepts?
- What methods can I use to measure outcomes that are not directly observable, such as undetected border crossings?
- How do I know that the measures I created are actually valid for what I am trying to measure?
- What are some likely limitations that may constrain my measurement efforts? How can I overcome these limitations, or work within them?
- How many measures should I use?
- What data sources can I use to measure performance?
- Once I have data on my outcome variables, how do I determine whether it was my program that caused those effects, or some external factor(s)?
If my program is just one of several programs or activities contributing to the same outcome, how do I know the effect my program had? What if I am not the program’s only owner?

In addition to these general challenges, DHS has some additional challenges related to its unique characteristics. DHS is the newest federal department, established in 2002. It is one of the largest Cabinet agencies, with more than 240,000 employees spread across 22 different headquarters directorates, offices, and operational components. The department is also responsible for a wide array of missions and functions. It has operational responsibilities in the areas of law enforcement, transportation security, border security, immigrations and customs, and special security events, among others. The diversity of DHS components’ missions and cultures, as well as their diverse stakeholder bases, yield an equally diverse set of results for DHS program managers to measure and manage, some of which are influenced by factors outside of their control or are quite difficult to measure.

For instance, some DHS programs result in tangible outputs, such as the number of passengers screened by the Transportation Security Administration (TSA) on domestic flights. However, the outcomes of this set of activities at TSA are much more difficult to measure. For example, how many terrorists are deterred from attacking the air transportation sector because of TSA security measures? How much safer is air travel?

This resource can help program managers to understand and overcome the many challenges they face in developing, implementing, and executing sound program performance measures.

**About This Resource**

Fundamentally, this resource seeks to help DHS program managers use measurement to tell their program’s story in a manner that communicates the program’s intended results, when those results are achieved, and implications for DHS strategic priorities and missions. It is intended to serve as a tool for DHS managers as they design and implement measurement systems for their programs. This resource will describe key principles and widely

---


12 A recent study by HSSAI identified four types of measurement challenges that are prevalent in the homeland security environment (see chapter 5 for more on these challenges). Jennifer Jacobs et. al., *Measuring Security: Approaches for Applying Relevant Metrics to Homeland Security Challenges*, (Falls Church, VA: Homeland Security Studies and Analysis Institute, October 30, 2013).
accepted, repeatable practices for developing outcome-focused performance measures. These principles and practices can be tailored and applied to the wide range of tasks and activities of the homeland security enterprise.

To create this resource, we consulted numerous academic, government, and professional sources, including our own past work on performance measurement, which helped us to understand a variety of measurement methodologies. In addition, we investigated the applicable authorities and guidance on performance measurement for government agencies. For interested readers, these sources are provided at the end of each chapter, in an “Additional Resources” section. We have also included a glossary in appendix A as a quick-look reference for key terms and their definitions. The key terms from the glossary are in bold font where we define or introduce them throughout the document.

Throughout the document, we draw on a single conceptual framework, first described in chapter 2, to help program managers make sense of all the moving parts in their programs – inputs, processes, and outputs – and how these relate to goals or objectives and, most importantly, program outcomes. This framework also calls out the important distinctions between these program elements and helps illustrate why results are so important. Following chapter 2, Part II of this resource (chapters 3-8) presents key principles and practices – as well as some how-to guidance – for measuring program results.
ADDITIONAL RESOURCES

FEDERAL GUIDANCE


GENERAL PERFORMANCE INFORMATION


To measure for results, a program manager must first understand the basic elements of a program. Understanding these elements—and the differences and connections between them—will help program managers to measure the right things and communicate what it is that they are measuring. A common misstep in performance measurement occurs when program managers believe they are measuring results, but they are actually measuring a different element of the program, such as outputs or activities. This can lead to misunderstanding and miscommunication about the program’s progress or success in achieving its purpose. In this resource, we use a basic conceptual framework that can help program managers correctly identify and understand the major differences between program elements, so that they can measure the right things.

This framework is a variation of a program logic model. A logic model illustrates the relationships between a program’s purpose, inputs, processes, and outputs. It also depicts how inputs, processes, and outputs lead to outcomes that reflect the program’s intended purpose. Figure 2 depicts our representation of an adapted program logic model. We are using this framework to illustrate the basic elements of a program, as well as the fundamental relationships between them. Most importantly, note the

---

13 Different variants similar to program logic models also exist in the literature, including Suppliers, Inputs, Process, Outputs, and Customers (SIPOC) methods developed for private industry that analyze a program’s suppliers, inputs, processes, outputs, and customers. SIPOC originated as part of the Total Quality Management programs. In the government context, supplier can be considered roughly equivalent to mission and customer can be thought of as parallel to outcomes. For more on SIPOC, see “SIPOC (Suppliers, Inputs, Process, Outputs, Customers) Diagram,” ASQ Service Quality Division, accessed September 24, 2014, http://asq.org/service/body-of-knowledge/tools-sipoc.

14 See appendix C for more discussion of the program logic model.
2. A Conceptual Framework

alignment between the program’s intended purpose (missions, goals, and objectives) and the measured program results, or outcomes. Establishing this alignment should be the aim of any results-based performance measurement system.

In the following paragraphs, we’ll briefly explore each of the framework elements, as well as their interrelationships and roles in program performance measurement. We’ll then discuss some uses of the framework and its benefits. In later chapters in Part II, we will provide a more detailed look at two framework elements: the program's purpose and outcomes.

Importantly, all of the following framework elements could and often should be measured by program managers and other stakeholders. As a DHS program manager, the specific elements you choose to measure will often depend on the purpose of your measurement. In the following paragraphs, we’ll briefly cover some of the possible use cases for measures of each of these elements.
Purpose

A purpose is a statement about what the program intends to accomplish or its desired outcomes. It may be reflected as a mission, goal, or set of objectives. From a programmatic point of view, the program purpose is the most vital aspect of the framework and should be clearly aligned with overall strategy. The purpose defines the desired outcomes of and fundamental reason for the program and is thus the impetus for the entire programmatic endeavor. Without a clear purpose, it is impossible to define or determine the success of a program. Yet the clarity of purpose can be highly variable across DHS programs and contexts.

In some cases, the purpose of a program is clear, but establishing whether you’ve accomplished the purpose is nebulous. For example, the purpose of the DHS “If You See Something, Say Something™” campaign is “to raise public awareness of indicators of terrorism and terrorism-related crime, and to emphasize the importance of reporting suspicious activity to the proper local law enforcement authorities.” Whether you’ve raised public awareness could be difficult to ascertain; however, with the use of appropriate, well-defined, and well-understood metrics, an acceptable determination could most likely be reached.

Some programs that have a straightforward purpose that is relatively simple to identify and measure. For example, “The Federal Law Enforcement Training Center (FLETC) provides career-long training to law enforcement professionals to help them fulfill their responsibilities safely and proficiently.” A possible metric for this purpose could be the number and type of graduates from FLETC each year. Note that this metric is an output measure, rather than an outcome measure. We will expand on the distinction between outputs and outcomes in greater detail below, as well as in chapter 4.

---

17 An example of such a metric might be the change in suspicious activity reporting rates along with metadata as to whether or not the suspicious activity reporter had been exposed to the “If You See Something, Say Something™” campaign.
2. A Conceptual Framework

**Inputs**

The next element in the framework is inputs. Inputs are the resources that are used to undertake activities, produce outputs, and influence outcomes. Some types of resources include funding, personnel, and capital investments such as information technology.

In the DHS context, for instance, inputs might include the number of Border Patrol agents assigned to the Southwest border. Program managers might want to measure inputs for several reasons. Among other things, input data can inform the planning, budgeting, and resource allocation necessary to carry out the program’s processes and activities, assuming that the program manager knows how many agents are necessary to execute the mission. Keeping accurate and timely measurements of input data can promote sound decisions in managing the program’s processes.

**Processes**

Following inputs in the framework is processes. A process is a program activity, workload, or other effort. These may include initiatives, procedures, workflows, or other programmatic activities.

The organizations within DHS accomplish a number of activities in the pursuit of their missions. Measuring or tracking these processes and their characteristics allows the program manager to optimize performance. For example, one of FEMA’s roles is to process disaster assistance requests. There are several steps in the request process, some completed by the victim and some by FEMA. Each of these steps has associated metrics that are monitored. The goal of measuring this process is to ensure that accurate decisions are made as quickly as possible to allow individuals and communities to recover from disaster. Other examples might include CBP’s border patrolling activities, as well as airport screening activities undertaken by TSA.

---

21 Another more complex input might be the baseline of illicit activity to be stopped by Border Patrol agents at the border.
## Outputs

After processes in the framework come outputs. An **output** is “amount of products or services delivered.”23 These typically include the products and services that result from program activities.24

The following are a few examples of outputs for DHS:

- The number of illegal aliens apprehended by U.S. Immigration and Customs Enforcement (ICE) in the interior of the United States
- The number of prohibited items confiscated by the Transportation Security Administration (TSA) from carry-on luggage at U.S. airports
- The number of citizenship applications processed by U.S. Citizenship and Immigration Services (USCIS)
- The number of distressed boaters and mariners rescued by the U.S. Coast Guard (USCG)

Measuring outputs gives program managers the ability to determine if their program processes are operating correctly and efficiently. These measurements also provide the information necessary to help optimize any underperforming processes, including the ability to identify any gaps or opportunities for increasing efficiency or effectiveness in processes. Output measures can also be useful for motivating and rewarding employees, as outputs are often more immediate, salient, and controllable (at least by any individual employee or team) than outcomes.25

## Outcomes

The final element in the process chain of the framework is outcomes. **Outcomes** are “the desired results of a program.”26 They are also interpreted as the impact of a program.27 Measuring outcomes can occur at different times during or after the life of a program, and baseline measures of outcome variables from *before* a program’s existence can also be useful for comparison purposes.28 Outcomes can also be more or less immediate, and you could measure a variety of outcomes that you would expect to see emerge at different times. Outcomes may be:

---

24 A more in-depth discussion of outputs can be found in Part II.
27 Metrics for Measuring, 18.
28 A more in-depth discussion of outcomes can be found in Part II.
- **Near term**: the outcomes most closely related to the program’s activities and results.  
  - **Medium term**: the outcomes that are based on the application of the near-term outcomes and their cumulative effects.  
  - **Long term**: the high-level outcomes that may not be realized for some length of time, but are the direct and indirect outcomes of the near- and medium-term outcomes.

As an example, let's consider some outcomes of a specific DHS program. Customs and Border Protection (CBP) “is charged with keeping terrorists and their weapons out of the U.S. while facilitating lawful international travel and trade.” These intended purposes, or responsibilities, could also be viewed as desired outcomes. In the near term, this could mean that terrorist attacks are stopped or thwarted at the border and that rates of travel and trade remain consistent from year to year. In the medium term, CBP could aggregate the results of its activities to present the impact of its work on border trends, such as illegal immigration. Finally, for the long term, CBP could relate the outcomes of its activities to broader strategic missions of DHS – preventing terrorism and enhancing security, securing and managing the borders, and enforcing and administering immigration laws. More information on considering time-phased outcomes can be found in chapter 4.

### External Factors

Although not directly in the process chain, external factors influence each element. External factors are factors outside of a program that may influence the program’s inputs, processes, outputs, or outcomes. External factors can range from politics to budgetary constraints to public opinion and can even include other programs that may impact one’s own program or its intended outputs/outcomes. The role of external factors in program performance measurement is discussed in more detail in chapter 7.

One external factor for DHS is the input of Congress; the department reports to 119 congressional committees. This is far more than other departments,
such as the Department of Defense. It requires DHS to spend a lot of time and resources measuring things that are of interest to external stakeholders. These things may or may not be outcomes that are relevant to measuring for results. Another external factor is the fiscal environment, which may affect multiple program elements, particularly inputs. Finally, legal and statutory requirements can also be viewed as an external factor bounding program inputs and activities.

**Benefits of Using the Framework**

The program framework we’ve adopted in this resource can help you, as a program manager, to identify the important program elements you may need or want to measure. By walking through your program’s framework, you have essentially documented the entire landscape of variables or factors that are relevant to your program – its inputs, processes, outputs, and outcomes, and any external factors that affect these elements in some way.

Once you have identified the variables that are relevant to your program, you can find ways to measure the ones that matter most. How many resources do you have? How much activity is happening in your program? How many products have you generated? What is the direction (positive or negative) of public opinion, and how strong is it? Then you can examine the data to determine how your program is actually doing. Are the outcome variables showing the kind of patterns you want – for instance, has crime decreased since your program was launched? If so, can you really attribute this pattern to your program’s activities and outputs, or might some other external factor be responsible? Conversely, if crime has increased since your program was launched, is there reason to believe that it would have increased even more without your program?

The process of developing this sort of framework for your program can also help you to establish a “line of sight” that shows the alignment of all the program elements, including outcomes, to strategy at all levels. Your program’s resources feed into its activities, which in turn should produce outputs that contribute to key outcomes: the objectives, goals, and missions

---

A VARIABLE is any factor that can take on more than one value. Resource levels, activity levels, products generated, and public opinion are all variables, because they can have more than one value (e.g., positive, negative, or neutral).

---


MEASURING OTHER PROGRAM ELEMENTS

Although this resource focuses on measuring outcomes, all of the other program elements can and often should be measured, as well. Measures of program inputs, activities, and outputs provide valuable information that program managers can use in refining and improving the program’s operations. These refinements can ultimately lead to better results.

Program managers should seek to understand and (as appropriate) measure how their programs’ inputs, activities, and outputs contribute to results.

The ideas behind this “line of sight” concept relate loosely to the ideas of value stream mapping, another management methodology.

If you’ve established an alignment between these program elements, you’re better equipped to argue that your program is achieving its intended purpose. This is what measuring for results is all about: showing that your program is contributing to the outcomes that matter. A lack of alignment in either the desired or actual program elements can reveal problems with program design, planning, or implementation that may impact the program’s overall success in achieving its intended purpose. Understanding your program in the context of the framework can help uncover these issues and highlight potential intervention points where misalignments exist.

Ultimately, you can use the framework to tell the story of your program – its resources, activities, outputs, and impact on relevant strategic objectives, goals, and missions. The framework can help you articulate and communicate your program’s logic and important results with peers and decision makers. You can also use data on the program outputs and outcomes to improve and inform your program inputs and activities, as well as the broader objectives, goals, and missions your program is intended to support. This “improve and inform” use case for the framework deals more with program evaluation and management than performance measurement. However, it is worth noting here as a key benefit of performance measurement with the framework we’ve adopted. Walking through this framework encourages systematic consideration of the critical whats, hows, if-thens, and whys of your program. It also allows you to document the answers to these questions in a clear and comprehensive diagram that can be easily communicated to other stakeholders.

What’s Next?

The first part of this resource provided an introduction to performance measurement, as well as an overview of a conceptual framework that can guide your efforts to measure your program’s performance. Although all the elements of this framework can be useful, two elements are particularly relevant to measuring for results: the program’s purpose and outcomes.

In Part II of this resource, we explore both of these topics in greater detail. The bulk of Part II focuses on principles, approaches, and considerations for developing and implementing results-based performance measures.

---

The ideas behind this “line of sight” concept relate loosely to the ideas of value stream mapping, another management methodology.
ADDITIONAL RESOURCES


Measuring for Results
Not all governmental programs lend themselves easily to measurable goals. For some it will be very difficult, and for a few, perhaps impractical altogether. Nonetheless, managers should resist the temptation to decide too quickly that a particular program is unsuitable for measurable goals. The fundamental question is, what is the difference between a successful program and a failure? Between a well-run operation and one that is mismanaged? How can we tell the difference, and how should that be defined? If the difference cannot be defined, then is that not just an invitation to waste, inefficiency, and ineffectiveness?


In Part I of this resource, we introduced the concept of measuring for results – or results-based performance measurement – and presented a conceptual framework that shows the many kinds of program elements and results that you could feasibly measure: inputs, activities, outputs, and outcomes. In Part II, we turn to the “how to” of results-based performance measurement, outlining some concepts, principles, and practices that you can use to measure your program’s results.

In chapter 3, we provide some guidance for identifying or defining your program’s purpose – the results your program should have, and thus the outcomes you should seek to identify in your performance measures. In chapter 4, we describe some of the important differences between outputs and outcomes, and discuss why measuring outcomes is so important – and so challenging. Chapter 5 presents some foundational concepts to help you define and understand the outcomes you want to measure, and discusses some common measurement challenges related to these issues. In chapter 6, we discuss some considerations that can help you narrow the range of outcomes to measure and thus reduce your measurement costs. And in chapter 7, we provide an overview of several advanced measurement topics that often arise in the DHS context.

Part II provides a useful set of principles and practices for designing and implementing sound, results-based performance measures. These principles and practices are drawn from a vast knowledge base, built through decades of measurement theory, research, and practice. Given the enormous amount of information on measuring for results, this resource focuses on the most useful takeaways for DHS program managers. As in Part I, a list of additional resources is provided at the end of each chapter for those who seek a “deeper dive” into a specific topic.
3. The Purpose-Driven Program: Identifying Desired Outcomes

In order to measure for results, the first question you should ask is: What are the results I’m looking for? That is, what outcomes should this program have? We refer to the desired or intended outcomes of a program as the program’s purpose. As defined in chapter 2, a program’s purpose is a statement about what the program intends to accomplish or its desired outcomes. Although a program’s purpose is an expression of intended or desired outcomes, it is generally derived from strategic guidance. A program’s purpose may be defined in terms of objectives, goals, missions, or other formal or informal statements of intended/desired outcomes for the program.

At this point, it may be useful to briefly return to the definitions of purpose and outcomes. These concepts are closely related but importantly distinct. A program’s purpose refers to the intended or desired outcomes of the program. A program’s outcomes, by contrast, are the actual results or outcomes of the program. In measuring for results, the goal is to assess whether the actual outcomes of your program indicate progress toward or achievement of the intended or desired outcomes of your program (its purpose).

In this chapter, we focus on key concepts and considerations for identifying or defining your program’s purpose – the intended or desired outcomes of your program. Accordingly, we use outcome terminology throughout this chapter. This terminology reflects the importance of conceptual alignment between your program’s intended/desired outcomes (purpose) and its actual outcomes. By identifying your program’s purpose correctly at the outset, you will be better positioned to measure its actual progress toward these goals.

For instance, you might be responsible for a program that is intended to secure .gov networks. This is the program’s purpose: its desired or intended outcome. To determine whether the program is actually achieving or progressing toward this outcome, you could measure a variety of indicators that assess the security of .gov networks. For instance, the rate of unsuccessful network intrusion/breach attempts that occur during FY15 could be used as one measure of .gov network security. (For more on outcomes and indicators, see chapters 4 and 5.) In this example, you can see that a program’s outcome measures should be defined in terms of its purpose; that is, purpose and outcomes should be conceptually aligned. Whether the program is actually achieving its purpose is something you can assess once you’ve collected data on those purpose-driven outcome measures.

The breadth and range of intended/desired program outcomes you could measure span multiple time horizons, strategic levels, and organizational layers (see chapter 4). To identify these outcomes, you can look at any number of program or organizational documents, talk to a multitude of different program stakeholders from employees to customers to leadership, or engage in hours of armchair speculation and brainstorming about the kinds of things the program should impact. In some cases, Congress, DHS leadership, or other stakeholders will already have identified some outcomes that are of interest to them, and thus are the “outcomes that matter.” With many sources of potential program purpose statements, it is possible to become a bit overwhelmed as you try to identify the kinds of results you should measure.

Moreover, due to resource constraints, you will often be unable to measure all the desired/intended outcomes that matter. The higher-level, longer-term outcomes in particular may be beyond your budget to measure – although some immediate, lower-level outcomes can also be rather complex and costly to measure (more on this in chapter 6). Regardless of these challenges, even if you can’t afford to measure all the desired outcomes of your program, it’s still important to have a good understanding of the full range of these intended outcomes.

For one, understanding the full range of your program’s intended outcomes can help you ensure that you’re not leaving anything important out when you decide which of those outcomes to measure. By identifying all the possibilities for your program’s results, you have a more comprehensive picture of the options available to you for measurement and can make better informed decisions when weighing among these alternatives.
Understanding all the possible outcomes is also important for establishing a clear line of sight from the lowest-level, most immediate outcomes of your program all the way up to the most distant, strategic and long-term outcomes. This line of sight essentially captures the alignment of your program’s outputs and outcomes (and even its inputs and activities) with the broader and longer-term goals, missions, or other purposes of DHS and the homeland security enterprise. Establishing this line of sight is important for many reasons.

For one, those who monitor the budgets and set strategic priorities for the organization will often think in terms of the higher-level strategic or mission outcomes of a program – or suite of programs – rather than lower-level, near-term results. If you can establish a line of sight that shows the alignment of your program’s lowest-level, nearest-term outcomes with the high-level, long-term outcomes that leadership and other decision makers often give the most attention, you may be better positioned to make a case for your program’s value.

A line of sight from the low-level, near-term outcomes of your program to longer-term, high-level results can also help you and your program’s stakeholders understand how the program fits into the bigger picture. It can be all too easy for employees, stakeholders, and managers to get caught up in the day-to-day tactical decisions, activities, and outputs that keep the program churning, and lose sight of the reason the program exists. By establishing the alignments between these day-to-day activities, outputs, and outcomes at the lowest and highest levels, you can remind yourself and your program stakeholders of why their efforts matter and the difference they are making.

Where to Find Your Program’s Purpose

As a group, program managers may find themselves in quite different situations as they seek to identify their program’s purpose and, relatedly, its intended outcomes. In some cases, the desired or intended outcomes of the program will already be defined — for example, in program planning documents. In other cases, the desired program outcomes will not yet be identified or defined anywhere, and program managers will have to determine what the program should impact. In most cases, however, program managers will encounter a mix of these situations, with some of their programs’ intended outcomes predefined and others undetermined.
No matter the situation, you can turn to a number of resources to help you identify the intended or desired outcomes of your programs. A featured call-out box provides an overview of some resources that you may find especially helpful.
RESOURCES FOR IDENTIFYING YOUR PROGRAM’S PURPOSE

**Program planning documents:** Any resources that describe the intended purpose and outcomes of the program can be considered planning documents — for instance documents that were written during the development of the program. These are more likely to provide purposes that are very closely linked with your program than the higher-level sources listed below.

**Congressional budget justifications:** Annually, DHS submits a budget justification to the Office of Management and Budget (OMB), which frequently describes the intent behind various programs, both within the components and at DHS headquarters. In many cases, program managers may be asked to provide input for budget justifications.

**Organizational strategic documents:**
- *Quadrennial Homeland Security Review (QHSR):* The QHSR outlines the vision, mission areas, goals, and objectives for homeland security. The first QHSR was published in 2010.
- *DHS Bottom-up Review (2010):* The Bottom-up Review provides an overview of the alignment between DHS activities and organization and the 2010 QHSR mission priorities and goals. Identifies areas where greater alignment is needed.
- *DHS Annual Performance Report:* This report identifies department-level performance measures, targets, and results for DHS missions, goals, and objectives. Describes the performance measurement and management process in DHS.
- *DHS Component, Directorate, Office, or Division Strategic Plans:* These documents can provide missions, goals, objectives, and (sometimes) performance measures and targets for specific organizations within DHS.

**External authorities and guidance:** Programs are influenced by many external policies, pieces of legislation, mandates, orders, or other guidance and authorities that establish desired outcomes, expected activities, or other goals and purposes for your program specifically, or for programs that address the issues your program is intended to address.

**Stakeholders:** Program stakeholders, including employees, customers, funders, and leadership, may have unique perspectives or institutional knowledge of the program’s intended purpose and outcomes that aren’t formally documented anywhere. Focus groups, interviews, and informal discussions with these stakeholders can provide valuable insights.

**Leadership and dynamic strategic priorities:** Organizational leadership may have valuable, early knowledge of new or shifting strategic priorities that are relevant to your program. Keep an ear to the ground for changes to your program or broader organization’s objectives, goals, and missions.

**Past program performance data:** What kinds of outcomes have been measured for this program in the past? Are these outcomes still relevant?

**GAO, IG, and similar reports:** Have reports identified any issues or problems with your program’s outcomes and impacts? What did these reports recommend your program could do better?
DEFINING YOUR PROGRAM’S OUTCOMES

In some cases, you will have more latitude to define the desired outcomes of your program. A variety of resources provide guidelines for defining useful and meaningful goals and objectives. For example, the SMART criteria are commonly used:

- **Specific:** Goals should be targeted to specific domain(s)
- **Measurable:** Quantifiers or indicators can be identified for progress toward the goal.
- **Assignable:** Responsible parties are clearly identified.
- **Realistic:** Feasibility constraints are taken into consideration in setting goals.
- **Time-related:** There are milestones and deadlines for goal progress/achievement.


If you can’t find your program’s intended outcomes in these resources – for instance, if your program is brand new – you may have more latitude to define your intended program results from scratch, perhaps in conjunction with leadership and other key stakeholders. In this case, it may be helpful to think through the following questions:

- What is the background or history behind the creation of your program?
- What problem(s) or issue(s) was your program created to address?
- Has leadership discussed or considered any milestones – formal or informal – for your program?
- Who are your program’s stakeholders (customers, funders, etc.), and what results do they expect your program to deliver?
- Where does your program fit within the missions, goals, and objectives of your division, office, component/directorate, and DHS more broadly?

Note that many of these questions can be answered – at least in part – by turning to the same resources listed above. In addition to these resources, creative thinking and discussions with different program stakeholders can help you identify the desired, intended, or anticipated outcomes of a new program.
3. The Purpose-Driven Program: Identifying Desired Outcomes

ADDITIONAL RESOURCES

GENERAL GUIDANCE


DHS-SPECIFIC GUIDANCE


4. Understanding Outcomes

When you want to measure your program’s results, there are generally two types of results you can measure: outputs and outcomes. Although we provided a high-level overview and definitions of these terms in Part I, this chapter will focus much more heavily on the important distinctions between outputs and outcomes. One common misstep in measuring for results occurs when outputs are measured and mistaken for outcomes. This chapter will help to clarify the differences between outputs and outcomes, and hopefully help you to guard against this common mistake.

Before tackling these topics, however, we should briefly review the two key terms that will be used throughout this chapter: outputs and outcomes. Outputs are the “amount of products or services delivered.”39 For instance, passengers screened and grants administered are both examples of program outputs. Outcomes are “the desired results of a program.”40 For a program designed to reduce risk, risk could be measured as an outcome. For a program designed to enhance community resilience, community resilience could be measured as an outcome. In the DHS context, risk and community resilience both contribute to national security, so national security could be a higher-level outcome of both of these programs.

As these definitions and examples suggest, there are some important differences between outputs and outcomes that have implications for performance measurement. For one, outcomes are often relatively abstract concepts that can be difficult to measure, whereas outputs tend to be more concrete and thus easier to measure. As noted in chapter 1, this

40 Ibid.
may sometimes lead to measuring only the “low-hanging fruit” — outputs — and shying away from the challenges of measuring outcomes. This is unfortunate, because relative to outputs, outcomes also tend to be more closely connected to higher-level program or organizational goals. This can render outcome data far more meaningful than output data for key program stakeholders and leadership. Outcome data also provides a window into whether the program is achieving its intended purpose; output data only tells you what goods or services your program is producing. These goods and services may or may not be contributing to the broader program goals. Measuring program outcomes is thus challenging but essential. It is possible to overcome these challenges and accomplish the difficult but necessary task of measuring for results.

Of course, there’s also a time and place to measure program outputs. Some of your program’s goals may be pre-established at the output level rather than the outcome level — for instance, a cargo screening program may have an efficiency goal to screen a certain amount of cargo in a set period of time. Measuring program outputs can also help you trace backward through your program logic (think about the conceptual framework here) to identify “breakdown” points that are resulting in less-than-stellar outcome data. Perhaps the quantity or quality of your program’s products and services are not up to standard, and this is harming the more distant outcomes of your program. You will often need to measure both outputs and outcomes that capture key program results.

Despite the differences between outputs and outcomes, many of the same basic principles and practices apply when developing good output and outcome measures. Given this commonality, Part II presents general measurement principles and practices that could apply to both outputs and outcomes. However, Part II focuses primarily on outcomes due to their importance in measuring for results.

Types of Outcomes: Thinking Vertically and Horizontally

Outcomes represent the impact of your program on the objectives, goals, and missions of the program and the broader organization in which it is embedded. The results of your program could appear tomorrow, or they could take years to materialize in any meaningful, observable form. Similarly, your program may affect the lower-level goals of your office or component,

---

but it could also impact the higher-level goals of DHS and the broader security enterprise of the nation, or even the global community. When determining what outcomes your program may affect, it’s important to think vertically – across multiple layers of objectives, goals, and missions held by various organizational levels within the broader homeland security enterprise – and horizontally – across a range of time horizons.

To help illustrate these concepts, let’s take a closer look at the outcomes piece of our conceptual framework (see figure 3).

**HORIZONTAL DIMENSION**

The horizontal dimension of measurement is the time horizon when you would expect your program’s outcomes to materialize. Your program may have an immediate impact on some outcomes, whereas for others it may take weeks, months, or years for the effects to become evident. DHS demonstrates this focus on the time horizon in its discussion of the five DHS mission areas. The *DHS Strategic Plan Fiscal Years 2012–2016* presents performance measures for each mission area and targets for future-year performance (see figure 4).

<table>
<thead>
<tr>
<th>Highlight Mission 4 Performance Measures</th>
<th>Planned Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of intelligence reports rated “satisfactory” or higher in customer feedback that enable customers to manage risks to cyberspace</td>
<td>FY 2012 FY 2013 FY 2016</td>
</tr>
<tr>
<td>80% 80% 80%</td>
<td></td>
</tr>
<tr>
<td>Percent of external traffic monitored for cyber intrusions at civilian Federal Executive Branch agencies</td>
<td>55% 70% 95%</td>
</tr>
<tr>
<td>Financial crimes loss prevented by the Secret Service Electronic Crimes Task Forces</td>
<td>$279M $300M $320M</td>
</tr>
<tr>
<td>Percent of unique vulnerabilities detected during cyber incidents where mitigation strategies were provided by DHS</td>
<td>95% 100% 100%</td>
</tr>
<tr>
<td>Average amount of time required for initial response to a request for assistance from public and private sector partners to prevent or respond to major cyber incidents (in minutes)</td>
<td>&lt; 90 &lt; 60 &lt; 15</td>
</tr>
</tbody>
</table>

Figure 4. QHSR Mission 4 performance measures and FY targets. Reproduced from *Department of Homeland Security Strategic Plan Fiscal Years 2012 – 2016*, February 2012, 14.

This example illustrates a common approach to addressing the horizontal dimension of performance measures, by measuring the same outputs and outcomes at multiple points in time. DHS sets specific future-year targets for activities associated with each of the QHSR mission areas. This approach can be very useful, for instance to evaluate change or improvement on each output or outcome over time. Note, however, this list focuses on operational
activities or outputs that can be quantified with little effort, rather than outcomes that may be more difficult but highly valuable to measure.

In this multiyear target-setting approach, it is also important to consider when you might expect to see the first or strongest impacts of your program on each output or outcome of interest. In the example above, for instance, some results may not be fully realized until FY2016, whereas others you might expect to see immediately. This can be reflected by setting different performance targets based on your theory of program change or impact for each output/outcome. Note also that this range of years for target setting is fairly limited in terms of what constitutes a “long-term” outcome. Indeed, although 2016 may seem long term relative to 2012, long-term outcomes could also be identified 20 or 30 years out.

You can work with stakeholders to identify any outcomes that might not be immediately affected by the program, and consider how long a delay you might expect before results start to appear for each of these outcomes. This sort of analysis can help you determine how long you should expect to wait before you see any potential impacts of your program on these outcomes.

**VERTICAL DIMENSION**

The *vertical dimension of measurement* represents the levels at which you are defining your outcomes. These levels could be defined in terms of the many layers of the organization: your division, office, and component, all the way up to the departmental, national, or even global level. For instance, the U.S. Border Patrol has an overall mission to detect and prevent the illegal entry of aliens into the United States.42 This mission ties to several of the missions of its parent organization, U.S. Customs and Border Protection (CBP). In turn, the missions of CBP directly relate to the larger DHS mission. Figure 5 illustrates these relationships.

In this example, note how the missions of DHS cascade down to a specific component, CBP, and one of its subordinate organizations, the Border Patrol. Taking a different perspective, each activity area of the Border Patrol can be directly linked upward to a mission in CBP. Each CBP mission, in turn, is tied to a broader DHS mission. This chart provides an example of how it is possible to trace the purpose and desired outcome through the levels and layers of an organization.

---

Figure 5. These are notional linkages as determined by HSSAI for explanatory purposes. The mission descriptions are actual language used by DHS, CBP, and Border Patrol.43

Alternatively, you could also define the vertical levels of your program’s outcomes according to whatever hierarchy your organization or program uses to lay out its strategic plan or purpose. For instance, some organizations might identify a broad mission that includes several subordinate goals, each of which involves meeting several lower-level objectives. The DHS Strategic Plan offers an example of a broad goal and its subordinate objectives and associated activities (see figure 6).

Goal: Strengthen and unify DHS international engagement

Improve the Department’s international interactions via enhanced coordination and rigorous review of the DHS international footprint.

Objective: Coordinate international engagement

Develop and execute mechanisms for all components to implement the DHS International Strategy and Regional Engagement Plans and put procedures and policies in place for the coordination of international travel, training, agreements, and other engagement.

Establish mechanisms for regular reporting and information sharing on the DHS international footprint, international arrangements and agreements, international training, technical assistance, and travel.

Objective: Review the DHS international footprint

Conduct a comprehensive review of all permanent and long-term temporary positions abroad.

Establish and conduct subsequent biannual reviews.

Figure 6. An example of a DHS strategic goal and its associated objectives and activities. Taken from the *Department of Homeland Security Strategic Plan Fiscal Years 2012–2016*, February 2012, 28.

The goal, “Strengthen and unify DHS international engagement,” has several associated objectives (two are shown here). The objectives are the necessary elements that must be accomplished to achieve the goal. Each objective then includes more specific actions that are intended to help achieve the objectives. Looking to strategic and organizational documents for your division, office, component, and DHS or the federal government more broadly is an excellent way to determine what outcomes you can and should measure along these vertical dimensions.

**MEASURING ACROSS BOTH HORIZONS**

For both the vertical and horizontal dimensions, it’s important to note that you can – and often should – measure outcomes across multiple different levels. You could have several near-, mid-, and long-term outcomes, some of which are more operational-level objectives or goals, and others strategic-level mission areas. And that’s just for one layer of your organization – you could measure each of these outcomes as defined by your division, office,
and component, each of which may have somewhat different objectives or goals. It is important to note that combined measurement of agency progress toward a higher level goal often rests with its higher-level leadership, rather than individual program managers.

Resource constraints may limit the number and type of outcomes you can feasibly measure, of course. However, it’s often useful to measure a variety of outcomes that span multiple levels – horizontal and vertical – to ensure you are capturing the outcomes that represent the most important and meaningful intended outcomes of your program. This sort of multidimensional, long-term planning can also help with long-term strategic planning, particularly for gaining awareness of outcomes on the horizontal dimension. At the very least, it’s a good idea to sit down with program stakeholders and work to identify the outcomes at all levels that your program should impact – regardless of whether you can measure those outcomes. This will give you and your stakeholders a better understanding of your program’s purpose, and also guide your attempts to identify, scope, and measure the outcomes that matter.

---

44 It is important to note that combined measurement of agency progress toward a higher level goal often rests with its higher-level leadership, rather than individual program managers.
ADDITIONAL RESOURCES


5. Defining Your Program’s Outcomes

Once you’ve identified your program’s purpose and outcomes, the next step is to ensure that you understand those concepts well enough that you can measure them. In the DHS context, understanding and defining the outcomes you want to measure can be simple or complex. It depends on what kinds of outcomes you’re trying to measure.

In this chapter, we’ll present some foundational key terms and concepts that can help you to define and understand the outcomes you need to measure. We’ll also provide an overview of some common challenges to understanding your outcomes, as well as some methods for tackling these challenges.

Key Terms and Concepts: Constructs and Indicators

As we discussed in chapter 4, the outcomes of DHS programs can span multiple time horizons, strategic levels, and organizational layers. As you go farther out and up on these horizontal and vertical dimensions, the outcomes often become more abstract, intangible, and multifaceted. For example, “secure and manage our borders” – one of the 2014 QHSR missions – suggests a much more abstract and intangible outcome than “detaining and removing aliens seeking illegal entry,” one of the Immigration and Customs Enforcement (ICE) Strategic Plan’s objectives for FY2010-2014. Secure and well-managed borders is also a more multi-faceted outcome that includes much

more than just detained and removed illegal aliens. It also includes intercepted drugs, weapons, and other contraband; effective deterrent strategies; and a number of other lower-level outcomes, each of which could be measured to tap into the broader outcome of secure and well-managed borders.

The above examples help illustrate some key terms and concepts for defining and understanding the outcomes of your program: constructs and indicators.

A **construct** is essentially a concept, typically one that is relatively abstract, intangible, and often multi-faceted. Secure and well-managed borders is a good example of an outcome construct. Resilience, risk, and deterrence are also prototypical examples of constructs relevant to the DHS enterprise. They’re all relatively abstract concepts that have multiple elements or aspects and are not readily abstract concepts that have multiple elements or aspects and are not readily measurable or observable.

To measure these kinds of program outcomes, such as risk reduction or increased resilience, you would first need to define and understand what risk and resilience are. Clearly defining your outcome constructs is very important for measurement, so that everyone is on the same page when you go to interpret and report your measures. The abstract, multifaceted nature of constructs can lead to confusion and misunderstanding if you are not careful to define your terms. How do you know when something is “resilient” or “risky”? What exactly does “reduced risk” or “increased resilience” look like? If you aren’t clear about your definitions of these constructs, different stakeholders may interpret your measures in very different ways. You could also accidentally omit some important aspects of the constructs you’re seeking to measure.

Consider, for example, a manager who is running a risk reduction program and decides to measure risk by counting the number of network intrusion attempts on the organization, a measure that does not provide a full picture of risk. What about the number of undetected intrusion attempts or, thinking about another facet of risk, the organization’s profile of vulnerabilities? These are also likely components of the organization’s risk, as are the impacts that a successful intrusion would have on the organization’s key resources and activities.

This example illustrates the importance of fully defining the constructs that represent your program’s key outcomes. If you have a comprehensive understanding of your outcome constructs – that is, how they are defined and what their important facets are – this makes it much easier to develop constructs and indicators.

---

Defining Your Program’s Outcomes

meaningful measures of the constructs. Continuing with the risk example, risk is a relatively intangible and abstract concept. Before stakeholders can point to something and agree that it’s high – or low-risk, they would first need to come to some consensus on what risk “looks like.” How should risk be defined, and ultimately, measured? Fortunately, a standard DHS definition of risk already exists and is fairly commonplace among both internal and external DHS stakeholders, and includes threat, vulnerability, and consequences as subcomponents. (This will not always be the case; see the discussion in the margin on disagreement over construct definitions for more on this.)

Although it’s nice to have a little more granularity on what risk is, these three facets are still fairly high-level, abstract, and intangible concepts. Indeed, threat, vulnerability, and consequence could also be viewed as constructs, all nested within the broader construct of risk. This nesting of multiple layers of constructs is fairly common, particularly when the major construct you need to measure is very abstract and high level. In this case, you may need to delve a few layers deeper into the definition of the construct to identify more concrete descriptors, which together can be used represent the “essence” of the construct. That is, if you looked at all those concrete descriptors as a set, you could say that combination taken together is what risk looks like. These concrete descriptors of a construct are commonly known as indicators.

Indicators – sometimes called observables, operational variables, or operationalizations – are the observable variables used to represent, or provide evidence of, an underlying construct. Because constructs are multifaceted, multiple indicators can often be used to represent a single construct. For instance, continuing with the risk example, you might select human casualties and financial losses as indicators of the consequences facet of risk, because human casualties and financial losses are both major consequences that can result from an incident. In some cases, you may need to delve a few more layers into the nested set of constructs to identify observable indicators. In the case of threat, for example, it might be helpful

Figure 7. Three lower-level constructs nested within the broader construct of risk.

DISAGREEMENT OVER CONSTRUCT DEFINITIONS

Because constructs are abstract, intangible, and multifaceted, stakeholders often disagree about how constructs should be defined. Different stakeholders could reasonably define the same construct in many different ways – and they could all be “right.” Resilience and terrorism are two notorious examples of disagreement and controversy over construct definitions. However, a lack of consensus about definitions doesn’t have to prevent you from measuring constructs.

Rather, do your homework on the construct and its different definitions, engage stakeholders in discussions (as appropriate), and finally, decide how you are going to define the construct for the purposes of measuring your program’s results. The key is to be transparent about your chosen definition and cognizant of any limitations or controversy that may result. If you understand your construct definition, and how and why you chose it, you’ll be well equipped to measure and defend it with key stakeholders.

---


to further define the facets of the threat construct as threat actor intent and capabilities, before identifying threat actor targeting patterns and capability estimates as indicators of those two facets.

Figure 8 shows the risk construct and its three facets (threat, vulnerability, and consequence), as well as the lower-level constructs that comprise some of these facets. The bottom layer of the figure also shows one possible set of indicators that you might use to measure risk. The arrow on the right shows the gradual distinction between the higher-level constructs and the lower-level indicators that represent these constructs. At the indicator level, you are finally getting down to more concrete, tangible, and measurable indicators that together represent the higher-level constructs. Therefore, indicators will typically be the aspects that you can measure to represent the important outcomes of your program.
CONSTRUCT VALIDITY

Much like a construct, construct validity has several important facets that are commonly discussed in the literature – convergent and discriminant validity, predictive and concurrent validity, and content and face validity. These last two aspects of construct validity are especially relevant in the DHS context.

Content validity is the extent to which you’ve identified and measured all the major facets of a construct, such as threat, vulnerability, and consequence for the risk construct. It’s important to note that you don’t necessarily need to measure all the facets or elements of a construct in order to establish content validity. Rather, you should determine which facets of the construct are most important and relevant for assessing your program’s outputs and outcomes, and measure those.

Face validity is the extent to which a measure seems “on its face” to measure the intended construct; for instance, do human casualty estimates seem to be a good indicator of consequences? Face validity can be very important for gaining stakeholder buy-in on performance measures. If the conceptual link between your indicators and constructs is not obvious, stakeholders may doubt that you’re measuring what you claim you’re measuring – even if evidence suggests you are. For example, if you use personality quizzes to estimate DHS employees’ future job performance, these measures may lack face validity – even if personality quizzes have been shown to reliably predict job performance.50

As figure 8 demonstrates, you can use the process of defining your constructs to establish a clear line of sight from your lowest-level measures or indicators all the way up to the highest-level constructs they purportedly represent. This helps you to ensure that your highest-level construct is accurately and adequately represented by the lower-level constructs and indicators that you have identified. By establishing this line of sight, you can quickly identify any indicators or lower-level constructs that seem out of scope, or seem to tap into something different than the higher-level construct you’re trying to measure. It will also be easier to notice any gaps where you’re missing key indicators or lower-level constructs that represent important aspects of the broader construct you’re trying to measure. Engaging with stakeholders and subject matter experts during this process can be very helpful for identifying these outliers and gaps, as well as promoting buy-in to your measures. For example, in figure 8, you might decide to modify the set of indicators to include psychological impacts as one of the consequence indicators, if it’s relevant for your purposes.

Ensuring that your indicators – or measures – accurately and comprehensively “tap into” the construct you’re trying to measure is an important step for any sound measurement system. This step is commonly known as establishing construct validity.50 Construct validity essentially asks the question: Are

---

your performance measures actually measuring the concepts or variables you wanted them to measure? By establishing a clear line of sight between your indicators and the broader constructs that represent your program’s important outcomes, you are accumulating some evidence of construct validity. When you’re working with high-level, abstract concepts, this step is critical for the defensibility of your performance measurement system. Numerous sources tackle this complex concept; the information in the margin provides additional relevant concepts, and readers may refer to the additional resources at the end of this chapter for more detail.

**Measurement Challenges**

These foundational concepts – constructs and indicators – are a useful framework to help you understand the kinds of outcomes you’ll often need to measure in the DHS context. This framework can also help you understand some common measurement challenges that often arise in the homeland security enterprise.

In a recent HSSAI report, Jacobs and colleagues identified four categories of measurement situations in the DHS context.\(^{51}\) Figure 9 depicts these four categories. As you can see, one of these categories is a “straightforward” measurement situation; this category includes easy-to-measure outcomes. The remaining three categories of situations present different types of measurement challenges for DHS. These challenges arise primarily for two reasons, which are shown in the margins of the figure: (a) relevant information is not available to measure the concepts of interest, and (b) a definitive solution does not exist – that is, the concept you’re trying to measure is “fuzzy,” or ill-defined. In this section, we’ll describe both of these issues and the types of measurement challenges they pose. We’ll also provide an overview of some useful strategies and methods for tackling these challenges.

---

**Figure 9. Four categories of measurement situations that arise in the DHS context. Adapted from Jacobs et al. (2013), Measuring Security, 3.**

<table>
<thead>
<tr>
<th>Relevant information is not available</th>
<th>Things that are hidden</th>
<th>Things that are fuzzy and hidden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant information is available</td>
<td>Things that are straightforward</td>
<td>Things that are fuzzy</td>
</tr>
<tr>
<td>Definitive solution exists</td>
<td>No definitive solution exists</td>
<td></td>
</tr>
</tbody>
</table>

STRAIGHTFORWARD OUTCOMES

In some cases, the outcomes you need to measure will be relatively straightforward and composed of low-level constructs that aren’t very far removed from their measurable indicators.52 In the DHS context, it is difficult to find such a straightforward outcome; many will be at least somewhat fuzzy or hidden concepts. However, one example of a relatively straightforward outcome might be DHS stakeholders’ use of information products to inform risk management decision making. This could be measured by assessing the percentage of relevant stakeholders who self-report use of these products to inform their risk management decisions. This indicator is just one small step removed from the construct itself.

FUZZY OUTCOMES

In other cases, the outcome constructs will be much higher level, the sort of abstract, multifaceted, intangible concepts provided as examples earlier: risk, resilience, border security, and so forth. In these cases, the constructs are much farther removed from their indicators, and you may need to define multiple layers of lower-level constructs before you can finally identify some measurable indicators. These sorts of outcomes have been defined by others as fuzzy concepts: concepts that lack a clear, unambiguous definition. Fuzzy outcomes represent a common measurement challenge in the DHS context, for all the reasons described above – they’re abstract, intangible, multifaceted, and often characterized by stakeholder disagreement or variability in construct definitions.53

Several methods are available that can help you to define, understand, and ultimately measure your program’s fuzzy outcomes. We’ve already covered the first method: clearly defining and identifying all the important aspects of your outcomes, along with good indicators of those constructs. By establishing a clear line of sight from your lowest-level indicators and constructs all the way up to the outcome you’re seeking to measure, you’ve completed an essential, preliminary step in measuring fuzzy outcomes. You’ve defined the things you need to measure – and with fuzzy concepts, as you’ve seen, that is no small task. We’ll review additional methods for

measuring fuzzy outcomes – along with other types of outcomes – in a chart at the end of this section.

**HIDDEN OUTCOMES**

Once you’ve identified some useful indicators of the outcome constructs you need to measure, you can start thinking about how to go about measuring those things. In some cases, identifying data on those indicators or measures will be relatively straightforward. Say, for instance, you’re interested in measuring the high-level outcome of **DHS employee morale**, and you’ve identified *employee satisfaction with the current DHS enterprise talent management system (ETMS)* as one important indicator of the **employee morale** construct. To measure this indicator, you would simply need to conduct a survey (or perhaps some interviews or focus groups) among DHS employees that asks about their satisfaction with the current DHS ETMS. In these sorts of cases, the information or data you need to effectively measure your indicators is obtainable – it may take a little effort to get it, but it’s available.

In other cases, however, the information you need to measure your indicators is simply unavailable. Perhaps the indicator requires measuring the probability of some future event; many future probabilities are inherently unknowable, and an estimate would have to suffice. Or perhaps the indicator calls for knowledge of events that are very difficult to detect, such as the number of active terrorist plots in the United States. These sorts of indicators have been termed **hidden concepts**: concepts that cannot be directly measured, even at the indicator level.54

Hidden concepts may manifest themselves in the DHS context in several ways. In some cases, for instance, the accomplishment of a DHS program’s purpose is almost impossible to determine until an adverse event has occurred. An example is the Federal Emergency Management Agency’s (FEMA) National Preparedness Program. The National Preparedness Goal is a “secure and resilient nation with the capabilities required across the whole community to prevent, protect against, mitigate, respond to, and recover from the threats and hazards that pose the greatest risk.”55 Verifying preparedness for a disaster is typically accomplished through training and exercises. For instance, performance during an exercise may be used as a proxy measure to estimate how well relevant entities will perform during a

---


disaster. However, the only sure way to determine preparedness is to assess the performance of emergency managers and responders after an event has occurred. Even then, measuring performance across a sample of multiple events, rather than a single event, is advisable.

Another common example of a hidden concept in the DHS context is interdiction rates. DHS program performance measures must often assess the success of interdiction efforts – for terrorist attacks, illegal border crossings, illegal transport of illicit substances, or other events that may pose a threat to homeland security. It may be tempting to measure interdiction outcomes using a raw count of “success stories” as a performance measure. For instance, you could measure the quantity of illegal drugs seized at border crossings, as this number is relatively easy to glean from field data, reports, and so forth. However, this performance measure would only capture the numerator in what should really be an interdiction rate, or fraction. The raw number of success stories tells you how many events were thwarted, but not how many might have been thwarted. Even if your program detected and intercepted five terrorist plots this year, how many terrorist plots did your program fail to detect that still pose an active threat to national security? Without knowing the latter number – the “failed to detect” events – you can’t really say how successful your program was at stopping terrorist plots. The missing denominator in the interdiction rate is thus the total number of events (terrorist plots, illegal border crossing attempts, etc.) that needed to be thwarted. The interdiction rate would tell a program manager: of all the events that might have been thwarted, how many were actually thwarted.

However, the total number of events (terrorist plots, etc.) that needed to be thwarted is often unknown – otherwise you would have at least attempted to thwart them. The unknowable nature of the denominator in many interdiction rates is what makes them hidden concepts. As you might expect, there are a variety of methods that you can use to estimate the denominator. Indeed, most of the methods for measuring hidden concepts involve estimates or approximations of some kind. If you can’t directly measure a hidden concept, you can indirectly measure it – perhaps by gathering multiple estimates from subject matter experts, or by measuring the next best thing (i.e., a proxy).\textsuperscript{56} We’ll provide a more comprehensive review of some of these methods at the end of this section.

MEASURING DETERRENCE: A FUZZY AND HIDDEN CONCEPT

Jacobs and colleagues discuss an example of how one might go about measuring the deterrent value of a DHS program. Let’s say that a program costs $1 million per year, and its purpose is to prevent the kinds of terrorist attacks that could inflict $10 million worth of damage. If this program is the only program designed to prevent these kinds of attacks, one would have to “believe” that it is actually preventing one or more attacks every 10 years to be worth the investment.

If, however, another program is also intended to prevent these kinds of terrorist attacks, then each of the two programs is only responsible for 50 percent of the work of preventing these attacks. This means that each program only “gets credit” for preventing $5 million worth of damage. Therefore, to justify the costs of the program, you would have to believe that this program is actually preventing one such attack every five years. This example illustrates a combination of the “What Do You Have to Believe?” and break-even analysis methods described in the table at the end of this chapter.


MEASUREMENT CHALLENGES AT THE INTERSECTION OF FUZZY AND HIDDEN CONCEPTS

In some cases, your outcomes will be both fuzzy and hidden. In these cases, you’ll need to work to define and understand all the important and relevant aspects of an abstract, intangible, and probably multifaceted construct – and then, even once you’ve identified some useful indicators of the construct, you’ll still have to find a way to estimate or indirectly measure these indicators. Despite these challenges, it is possible and often immensely valuable to measure these kinds of outcomes. One common example of a fuzzy and hidden concept in the DHS context that clearly illustrates this value is deterrence.

Much like risk or resilience, deterrence is a relatively abstract – or fuzzy – concept that could be understood to have many facets or aspects. Some might view deterrence as having both a cognitive and a behavioral element, for instance, each with many lower-level constructs under those broader umbrellas. Many indicators of deterrence also lack readily available data and thus have to be measured or estimated indirectly. However, deterrence is clearly an extremely valuable outcome to measure in the DHS context, as it is the one of the major purposes of many DHS programs and initiatives.

All of these measurement challenges may seem rather daunting. However, there are many well-established methods for overcoming them. The following chart (adapted from Jacobs et al., 2013) provides an overview of some of these methods and notes the different categories of measurement situations where each may be useful. For additional resources on these measurement methods, readers may refer to the sources provided in endnotes for each method.

57 For more on measurement challenges and strategies for deterrence and prevention programs, see Performance Measurement Challenges and Strategies (Washington: White House, June 18, 2003), 10-11.
### Methods for Measuring Challenging Concepts

#### Hidden Measuring Methods

**Sampling**
- A sample of the visible subset of a hidden set can be taken and the results can then be generalized for the whole population.¹
- CBP currently uses this approach in numerous ways, e.g., using sampling to help estimate the total population of illegal immigrants attempting to cross the border.²

**Inference from Secondary Effects**
- A method that measures secondary effects from the visible data set and then uses those measurements to make inferences about the hidden data set.³
- CBP currently uses this approach by counting footprints to make inferences about illegal immigrants crossing the border.

**Proxy Measurement**
- A method where something very similar to what one wants to measure is used when what one wants to measure cannot be measured.⁴
- CBP currently uses this approach with its automated targeting system, part of the cargo entry system.

**Large Data Set Patterns**
- A method where analysts look for patterns or correlations in large data sets.⁵
- DHS has numerous large data sets it can use to implement this approach.

#### Straightforward Measuring Methods

**Administrative Data from Program or Agency Records and Registrations**
- Administrative data, e.g., from the Immigration and Naturalization Service, can be used to measure behaviors that are not visible, such as illegal border crossings.⁶
- DHS collects vast amounts of data and maintains numerous databases, e.g., Immigration-related databases, FEMA’s Flood Insurance Rate Map databases.

**Statistical Institutions**
- Many institutions, both national and international, collect and disseminate official statistics.⁷
- DHS owns some statistical databases and can utilize databases owned by international organizations.

**Special Technical Equipment**
- Special equipment is sometimes needed to collect data for scientific measurement.⁸
- ICE, CBP and other agencies within DHS use fingerprint-matching technology to analyze immigrants’ fingerprints; DNDO uses radiation detection equipment to collect test data; Coast Guard uses numerous types of scientific equipment, including Doppler current profiler to measure ocean velocities.

**Customer Surveys**
- A set of fixed questions asked to a representative sample of a particular population, at a particular time.⁹
- DHS could use surveys to collect information from the U.S. population.

**Trained Observer Ratings**
- Trained observers rate their perceived observations using rating scales.¹⁰
- DHS could use these ratings for many uses, e.g., to ascertain the conditions of numerous points along the border.

**Self-Assessments**
- A method similar to customer surveys and trained observer rating, with the primary difference being that respondents rate themselves.¹¹
- DHS could use self-assessments in the immigration arena, where current efforts already include asking migrant workers who tried to cross illegally about their efforts.

**Role-Playing**
- Trained participants act as both service customers and testers and are rated based on previously defined categories.¹²
- DHS currently uses a form of role-playing, red teaming, in numerous ways, e.g., using scenarios to test cybersecurity defenses. DHS’s use of role-playing could be expanded for use in other mission areas.

#### Fuzzy and Hidden Measuring Methods

**Natural Experiments**
- A method where an event that occurs in the real world (policies, natural disaster, etc.) creates an “experiment” where analysts can evaluate the before and after impacts of whatever change was caused by the event.¹³
- DHS could conduct these types of experiments by evaluating the before and after impacts of a change (such as a natural disaster or a new policy).

**Evaluation of Large Data Sets**
- A method that is useful if the target to be measured is defined.¹⁴
- This approach may be useful to DHS as long as the decision maker knows what is needed.

**What Do You Have to Believe?**
- A method used frequently by business analysts that allows managers to gauge the necessary precursor conditions to realize success in achieving a specific goal or outcome.¹⁵
- Analysts iteratively ask questions about the requirements necessary to reach an objective.
- DHS could use this approach in both qualitative and quantitative ways.

**Break-Even Analysis**
- A method where the cost of the investment is compared with the cost of what the investment is designed to accomplish.¹⁶
- DHS can use this approach to measure things like its deterrence efforts.

#### Fuzzy Measuring Methods

**Surveys and Self-Assessments**
- A set of fixed questions asked to a representative sample of a particular population, at a particular time.¹⁷
- These approaches may have limited application to fuzzy problems because their methodology is by nature subjective; yet if carefully designed, these methods could be used.

**Contingent Valuation Method**
- This method determines a market value for a public good that is not easily definable in market form by using surveys.¹⁸
- These approaches may have limited application to fuzzy problems because their methodology is by nature subjective; yet if carefully designed, these methods could be used.

**Delphi Technique**
- A method to systematically elicit expert opinion.¹⁹
- In theory, this approach could be used by DHS. However, since the scientific community is critical of the technique, it must be used carefully.

**Analytic Hierarchy Process**
- A method for group decision making where participants contributions are arranged into a hierarchy.²⁰
- Although this approach has some potential uses for DHS, it is likely to face methodological challenges in its use.
Table Notes

4 For more on proxy measurement, see Joshua D. Clinton, “Proxy Variable,” Sage Encyclopedia of Social Science Research Methods, 2004.
5 For more on large data set patterns, see Big Data and Privacy: A Technological Perspective, (Washington: Executive Office of the President, President’s Council of Advisors on Science and Technology, May 2014).
7 See footnote 37 above.
12 Hatry, Performance Measurement, 111; Van Dooren, et. al., Performance Management, 63-64.
13 For more on surveys, see Robert M. Groves et al., Survey Methodology, 2nd ed. (Hoboken, NJ: John Wiley & Sons, Inc., 2009).
14 Hatry, Performance Measurement, 100; Van Dooren, et. al., Performance Management, 64.
17 For more on surveys and self-assessments, see footnotes 45 and 47.
CONSTRUCTS, INDICATORS, AND MEASUREMENT VALIDITY


MEASUREMENT METHODS


You have now gotten through some tough preliminary steps of program performance measurement: identifying and defining your program’s purpose and desired outcomes – that is, the things you want to measure. These steps are often very conceptual and strategic in nature. Identifying the program outcomes that matter, and the constructs and key indicators you would need to measure to assess your progress toward or achievement of those outcomes, is critical to measuring performance.

The next step in the process of performance measurement turns from these conceptual questions toward more pragmatic considerations. Specifically, you must carefully examine your resource constraints and decide how many – and which – outcomes and indicators to measure. In most cases, it will be impractical for you to measure all the indicators of all the desired outcomes of your program. This is because – simply stated – measurement is costly. Therefore, a critical decision point is determining which outcomes and indicators will provide the biggest “bang for your buck.”

**The Return on Investment (ROI) of Measurement**

At its heart, performance measurement is all about reducing uncertainty.\(^{58}\) Without sound evidence of some kind (quantitative or qualitative), it may be challenging to know whether – or how much – programs are contributing to the desired outcomes. A major goal of any measurement system is to reduce

---

uncertainty – that is, to give program managers and other stakeholders greater knowledge of how the program is doing on the outcomes that matter.

Rarely if ever will any measurement system provide 100 percent certainty about a program’s performance, but sound measurement can reduce uncertainty. Often, measurement will reduce uncertainty enough to permit verifiable assertions about how the program is doing, as well as informed decisions about whether and how to improve its performance. When scoping the range of outcomes and indicators to measure, you should weigh the value of reduced uncertainty on those outcomes and indicators against the costs of measuring them.59

This all sounds good in theory, but how exactly can you go about this kind of cost-benefit analysis? What sorts of factors can you consider, and what can you do to increase the value you’re getting out of your performance measurement system? In this chapter, we’ll discuss some common constraints on measurement that you may encounter in the DHS context. We’ll also parse out some of the major costs and benefits of measuring different numbers and types of outcomes and indicators, and provide guidance and considerations for scoping down your set of measures.

**Constraints on Measurement**

As noted above, it will often be impractical for you to measure all the indicators of all the desired outcomes of your programs. Besides the costs of measurement, which we’ll explore in greater detail below, as a DHS program manager, you may encounter some additional constraints on the range of measures you can feasibly implement.

**EXTERNALLY IMPOSED LIMITATIONS**

In some cases, limitations may be externally imposed on the number of performance measures that you can report. Leadership and other stakeholders may only be willing or able to review a relatively small number of performance measures. This target number of performance measures – or even the measures themselves – may be set in formal or informal guidance. As a program manager, you may also know from previous experience the tolerance threshold of leadership or stakeholders for performance data. Of course, you can always collect more data than you report – but this may not

---

be advisable in all cases, particularly if nobody is going to use the data to inform any decisions about the program.60

Another externally imposed limitation on your performance measures is time. Time is always a factor in collecting and analyzing performance measurement data. In some cases, time to collect and analyze your measurement data may be limited by budget cycles or other events. In other cases, it may take significant time to measure the outcomes of a program over the long-term. Program managers should be sure to budget and plan for the time required to measure for results.

Yet another externally imposed limitation is the need to rely on stakeholders to measure something. For instance, TSA outcome measures may require airport or airline associations to conduct a survey of airports concerning security measures. If you are relying on other stakeholders for performance data, it is advisable to anticipate and plan for the challenges you may encounter in collecting or interpreting this data.

One additional limitation you may encounter is pressure to make use of legacy measures, that is, inherited measures that already exist for some specific purpose. Although these legacy measures may be well suited to their original purpose, they may be less well suited to measuring new or different program outcomes. Despite their limitations, legacy measures can be useful in some cases; for instance, you might expand on or slightly modify the existing set of measures.

RESOURCE LIMITATIONS

Resources may also constrain the number or type of measures that you can afford to use. For example, some measures may require analysts to generate complex estimates of terrorists’ capabilities or financial losses from an incident. These estimates may be too costly for some to generate on their own, and relevant data from other sources may not be readily available. In general, you can expect to spend on labor and other “hard” resources to plan, implement, analyze, and report on your performance measurement system. In particular, collecting the data can be a very resource-intensive endeavor, depending on how you go about it. Conducting one-on-one interviews or designing and launching a survey may cost much more than leveraging ongoing field data collection efforts or existing data sets from external


HOW MUCH TO SPEND ON MEASUREMENT?

As a program manager, you may have some control over how much of your programmatic budget to spend on measurement. Some guidance is available that may help you make this decision.

For instance, Hubbard (2010) recommends trying to estimate the value of “perfect information” on the outcome under consideration. Whatever this value is, Hubbard recommends spending 2 to 10 percent of it to reduce your uncertainty on this outcome – that is, to measure the outcome well. This method is based on the premise that at some point, taking additional measures provides diminishing returns. You should seek to informally understand the costs and benefits when changing your measurement system. Hubbard provides more information on how to conceptualize the value of perfect information.

\[
\text{ROI of measurement} = \frac{\text{Value of reduced uncertainty}}{\text{Costs of measurement}}
\]

As another means of budgeting for measurement, Jacobs and colleagues (2013) suggested allocating some portion of the overall management costs of your program to performance measurement.

sources. These kinds of budgetary costs and considerations are discussed in greater detail in the section below. The takeaway, for now, is that resource availability will often pose external constraints on the number and type of measures you take.

MEASUREMENT FATIGUE

Even when there are no external constraints on the scope of your performance measures, it’s generally a good idea to limit them to a manageable number. Particularly when stakeholders are involved in generating or providing performance data, an excessive number of measures can lead to measurement fatigue. Measurement fatigue is exactly what it sounds like – when you attempt to measure too many things, the people providing the data become tired, bored, and/or exasperated. This can greatly reduce the quality and quantity of available data and effectively cancel out any benefits of using so many measures in the first place. Pilot testing and stakeholder engagement early in the process can help you arrive at a reasonable number of measures to avoid measurement fatigue. The balanced scorecard model discussed later in this chapter provides a way to focus attention on the most important information by limiting indicators. This can help avoid distraction related to fluctuations in measures that are interesting, but not important.

A CONSTRAINT – AND AN OPPORTUNITY

Another important constraint on your measurement system may be easily overlooked, but is actually quite useful for working around some of the other constraints you may face. Namely, you may not be able or willing to scope down the number of desired program outcomes to measure – but you generally do have some flexibility at the indicator level to save on costs.

At the outcome level, if you’ve already established the importance of the outcomes you want to measure, you probably shouldn’t try to cut costs by cutting those outcomes from your measurement system altogether. Otherwise you’ll be losing valuable performance information that’s of interest to program stakeholders and could be used to make important decisions about the program’s future.

---


At the indicator level, however, remember that there are often many different indicators you could use to measure the same outcome. Some indicators may be cheaper to assess than others. Moreover, once you’ve chosen the indicators you want to measure, there are innumerable ways you could measure them. Some of those measurement techniques and approaches are cheaper than others. By being flexible and creative at the indicator level, you can measure more of the outcomes that matter to your program and its stakeholders.

Of course, when selecting and measuring your indicators, the goal is always to ensure that you are actually measuring the outcomes you’re trying to measure. This brings us back to the concept of construct validity, which we discussed in chapter 5. If you get too creative with your indicators to try to cut costs, the indicators may start to drift away from the outcome you were originally trying to assess. Whenever you make a change to your indicators and the way you’re planning to measure them, always go back to that key question: *Am I still measuring the outcome I’m trying to measure?*

The remainder of this chapter focuses on exactly these kinds of considerations. When you’re deciding which indicators to measure, and how to measure them, how can you preserve construct validity while saving on measurement costs? In the next section, we’ll discuss some considerations and questions you can ask yourself to decide *what* indicators to measure. In the following section, we’ll explore some options for *how* to measure those indicators, each with its own unique costs and benefits. Taken together, this chapter should equip you to make better informed decisions about which indicators to measure and how to measure them.

### Deciding What to Measure: Scoping Your Indicators

Recall the discussion above on the ROI of performance measurement. We’ve now covered some of the major costs and constraints of performance measurement. The “benefits” side of the ROI equation is the value of reduced uncertainty on the outcomes and indicators you measure. Of course, reduced uncertainty on some outcomes and indicators will be more valuable than others. For instance, the outcomes and indicators that are most directly relevant to your program’s purpose, or perhaps those that will resonate most strongly with leadership and other program stakeholders.

**CONSTRUCT VALIDITY**

is the extent to which your measures are actually assessing the construct you want to measure.
In this section, we’ll overview some of the factors you may want to consider as you decide which indicators will get you the biggest return on investment. Again, we’re primarily focusing on indicators rather than outcomes here, because the indicators are where you’ll typically have more flexibility to pick and choose, and thus save or incur measurement costs. Of course, some of these criteria may also be useful for down-selecting the outcomes you want to measure, if you’re facing serious resource constraints that won’t allow you to measure all the results that matter for your program.

DISCRIMINATING POWER: SUCCESS OR FAILURE

In deciding which indicators to include in your final set of performance measures, it may be especially helpful to consider which ones will maximally distinguish between success and failure on the outcome you’re trying to measure. As the quote at the beginning of Part II asks, “What is the difference between a successful program and a failure? Between a well-run operation and one that is mismanaged? How can we tell the difference, and how should that be defined?”

To illustrate, let’s continue with the risk example from chapter 5. Recall that the program manager decided to use only two indicators to measure the consequences outcome within the broader risk construct. These two indicators of consequences were human casualty estimates and estimated financial losses (as shown in figure 11). Of course, there are many other potential indicators of consequences that a program manager could use here – for example, the estimated psychological effects of an incident on the community. However, in this example, the program manager down-selected from the full set of possible indicators to choose the two that seemed most important and relevant to the program. Perhaps the program is primarily intended to reduce human casualties and financial losses. Or perhaps these two indicators signify the success or failure of a risk reduction program much more powerfully than the other possible indicators of consequence.

In deciding which indicators to measure, you can ask yourself: What indicators will provide the strongest distinction between success and failure on my outcome of interest? The more powerful your indicator data is at discriminating between success and failure, the more weight those indicators may carry in your measurement system. This is particularly true if leadership

---

IS ONE THE MAGIC NUMBER?

When scoping down the number of outcomes and indicators to measure, it may be tempting to rely on a single indicator to measure a program’s overall success. It’s generally cheaper and simpler to measure one thing vs. many. There’s also some aesthetic appeal in making programmatic decisions based on a single number; it seems to simplify the problem space for decision making.

However, there are some drawbacks with using just one indicator to measure your program’s performance:

- You may not be measuring all the outcomes that matter if you have just one indicator.
- Single indicators are often aggregates or averages of many different factors. This obscures the strengths (highs) and weaknesses (lows) of the program.
- Too much pressure to perform on a single indicator can lead to unintended consequences, such as denominator management.

To avoid these issues, consider using a carefully-selected set, or dashboard, of indicators. Just like a dashboard on a car or airplane, multiple gauges or indicators of key outcomes can help you pin down what’s going wrong — and what’s going right — in your program.


---

6. Narrowing the Field: Deciding Which Outcomes and Indicators to Measure

and other program stakeholders similarly view those indicators as powerful discriminators.

**PREORDAINED INDICATORS**

In some cases, you may find that the outcomes and indicators you need to measure have been preordained for you. Program or organizational leadership, congressional or executive mandates, and other authorities may issue requirements that narrowly define specific constructs and indicators for you to measure and report. In these cases, the obvious benefit of measuring these outcomes and indicators is that you’re fulfilling the requirement and thus keeping your bosses, auditors, or other authorities informed.

Hopefully, however, the outcomes and indicators you’re required to measure are a value-add for other reasons. The best-case scenario occurs when these mandated variables already “fit” into the set of constructs and indicators you’ve identified as meaningful and intended outcomes for your program. This essentially means that the authorities are requiring the right measures, asking you to assess the outcomes that truly matter, using the indicators that best represent those outcomes.

Of course, this may not always be the case. Mandated measures may become outdated as programs evolve, and those creating the mandates may not always have a complete picture of the program’s purpose or intended outcomes. In these cases, it’s useful to understand these limitations so that you can more appropriately interpret and report performance data on the mandated measures. In some instances, you may even be able to use your knowledge of these limitations to argue for a change in the reporting requirements. In any event, understanding how mandated outcomes and

---

64 Legacy measures, described above as a type of externally imposed limitation, are one example – though not the only example – of preordained indicators.

65 It is also possible that mandated measures may be an indication of the program’s intended purpose or outcomes. If these don’t seem to align, there might be a powerful argument to change the measures.

---

Adapted from [http://balancedscorecard.org/Resources/About-the-Balanced-Scorecard](http://balancedscorecard.org/Resources/About-the-Balanced-Scorecard)

---

**THE BALANCED SCORECARD**

Originating in the business community, one common method for identifying the most important things to measure in an organization or program is the balanced scorecard, developed by Robert Kaplan and David Norton. The balanced scorecard approach advocates that you consider and measure key performance indicators (KPIs) from four different perspectives or domains, as shown in the diagram above.

As you can see, some of the perspectives (e.g., financial/stewardship and internal business process) in the balanced scorecard approach may be more conducive to measuring inputs, processes, and outputs rather than outcomes. However, all four perspectives could implicate some key outcomes and are worth considering as part of a broader program performance measurement system.
indicators fit (or don’t fit) into the scope of your key program outcomes and indicators is a useful step toward effective measuring and reporting.

WHO WILL USE THE INFORMATION, AND HOW?

Another way of deciding which outcomes and indicators to measure is to consider who will use the performance data on those outcomes and indicators, and how. This kind of assessment often requires close consultation with program stakeholders, particularly those who have the power to effect meaningful change to or within the program.

Will leadership look to certain outcomes and indicators to make budgetary decisions about the program? Will program employees be persuaded by certain indicators or outcomes to change the way they do business and improve the program from the ground up? Will program customers see the data as evidence that someone is actually listening to their feedback? Will data on certain indicators or outcomes help persuade other program owners that the successes or challenges you’ve been claiming are real?

Working through these kinds of questions can provide you with a multi-stakeholder perspective on the value of measuring particular outcomes and indicators. At the end of the day, of course, you may have to make some tough decisions about the stakeholders whose questions most urgently need answers. Once you’ve made those decisions, though, you can scope your measurement system accordingly.

THE COST OF ESTIMATES

Estimates can be very useful indicator measures in the DHS context. However, their costs are highly variable.

Because estimates often rest on many assumptions, you may want to plan to expend some resources on the decision process involved in weighing, making, and documenting these assumptions. This process may require subject matter expertise, which doesn’t come cheap.

Similarly, estimates often involve complex calculations that can output a range of possible estimates based on numerous values and assumptions you input. You may need to plan for software, personnel, and other resources that can help you understand your estimation options and make the required calculations.

AUDIENCE LEVEL OF KNOWLEDGE

A related consideration in scoping your indicators is how easily they can be understood by your audience. Some indicators can be measured using relatively simple methods. For instance, you might be interested in customer satisfaction with the timeliness, relevance, and aesthetic quality of your products and services. You could conduct a survey with a reasonably sized, representative sample of your customer base that includes three questions about their satisfaction with each of those factors. You could then tally up the percentage of the surveyed customers who rated their satisfaction as Satisfied, Very satisfied, or Extremely satisfied, and report that percentage as an indicator of customer satisfaction.

However, some indicators may not be as simple to measure. These indicators may require some complex mental gymnastics to understand how they were calculated and how exactly they represent the outcome you’re measuring. For instance, you might want to measure the risk of a high-consequence,
low-probability event such as a nuclear disaster. Because the consequences of a nuclear disaster could be so far-reaching, in terms of both time and geography, you will have to build in a lot of assumptions to whatever estimate you use as an indicator of the event’s potential consequences. Likewise, because these kinds of events occur so infrequently, your probability estimate for these events may also rest on many assumptions.

In all of these assumptions, there’s a lot of room for other program stakeholders to disagree with the values you use as a part of your estimate. To address this disagreement, it’s important to understand the assumptions you made and how they impact your overall estimate for that indicator or outcome. Usefully, many estimation programs will allow you to enter a range of assumed values, rather than a single value. You can then look more closely at the sensitivity of your estimate to changes in particular assumptions or variables.66

Let’s say, for instance, you talked to all the right subject matter experts and read all the right literature to try to determine the potential consequences of a nuclear disaster. Of course, none of them agreed on an exact number – instead, you ended up with a range of estimated costs somewhere in the order of $1 trillion to $10 trillion. This is a big range, and so it seems unwise to pick a single number – even the average, $5.5 trillion – for your overall risk estimate.

What a sensitivity analysis does is allow you to plug in the full range of possible costs – from $1 trillion to $10 trillion – and then look at how changes to this value would influence your overall estimate of risk. If, for instance, you’ve weighted consequences as a relatively small component of the overall risk estimate for a nuclear disaster, then even large changes to this consequences value may not impact your overall risk estimate very much. If, however, you’ve weighted consequences heavily in the overall risk estimate, even very small changes to this value could drastically impact your overall estimate of risk.

As you’ll note, then, sensitivity analysis can be helpful for answering stakeholders’ questions about the inputs to your overall estimate – the values of different variables and how they impact your final estimate. However, you still need to have a good understanding of the math behind your estimate – that is, how you chose to weight and combine each of the key variables that comprise your estimate. You may notice the effects of this math as

---

you are conducting a sensitivity analysis with your stakeholders – but the sensitivity analysis itself isn't designed to tell you how good your math is. If your underlying model for generating the estimate is flawed, no amount of sensitivity analysis will correct it (although it might help you figure out where some of the obvious problems are).

So what does all this have to do with selecting indicator measures and the sophistication of your audience? When you’re deciding what indicators to measure, and how to measure them, it’s important to think about how receptive your audience will be to the complexities involved in calculating some indicator measures. Some indicators can be measured in any number of ways, allowing you to tailor the complexity of your measures or estimates to your audience. Other indicators leave you with fewer options, virtually requiring that you use complex calculations to generate sound and relevant data on the indicator.

If you must use complex methods to generate your indicator data, be prepared to walk your stakeholders through your process and assumptions, and possibly to speak to the results of a sensitivity analysis that captures the uncertainty inherent in your final estimate. Visual aids and interactive tools can be helpful for making these complex indicator calculations and assumptions more concrete, particularly when your stakeholders are likely to be interested or are already measurement-savvy.

However, if your stakeholders don’t have the time or the interest to sit through a detailed description of all your mathematical maneuvers, you should be prepared to field any measurement-related questions your audience may have and briefly issue any important caveats about your data. If you know you’ll be facing a particularly tough audience, you may even reconsider using such a complex indicator in the first place or consider using a simple indicator with a more complex indicator as an ancillary measure to build confidence.

In the DHS context, however, these complex indicators will often be your best or only option for measuring your program’s key outcomes, particularly those that are hidden and/or fuzzy (see chapter 5). As always, you don’t want to cut costs by oversimplifying your performance measurement system, particularly if it would sacrifice the quality or relevance of your data to the outcomes you’re trying to measure. Using complex measures is possible, and although they often involve some degree of uncertainty, having some high-quality, appropriately caveated data on your program’s most important outcomes is always preferable to guesswork.
Primary and Secondary Data: Some Key Tradeoffs

The above sections provided an overview of some scoping considerations for choosing the indicators you want to measure in order to assess your key program outcomes. As we discussed earlier in this chapter, you will often have much more flexibility at the indicator level (rather than the outcome level) to pick and choose which ones to measure, and how to go about measuring them. These decisions can affect the overall costs and benefits of your measurement system.

In this section, we’ll cover some key costs and benefits associated with how you go about measuring your indicators. Understanding the tradeoffs involved in different measurement approaches and techniques can help you plan more effectively to increase the value you’re getting out of your measurement system.

In the literature on measurement, two major types of data are commonly discussed: primary data and secondary data. The table below provides definitions, examples, and an overview of the costs and benefits of primary and secondary data.67

<table>
<thead>
<tr>
<th></th>
<th>Primary Data Sources</th>
<th>Secondary Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>Data collected by you</td>
<td>Data collected by others</td>
</tr>
<tr>
<td>Examples</td>
<td>Surveys</td>
<td>Archival records</td>
</tr>
<tr>
<td></td>
<td>Interviews</td>
<td>Open-source or social media</td>
</tr>
<tr>
<td></td>
<td>Focus Groups</td>
<td>Transcripts</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>Databases</td>
</tr>
<tr>
<td>Benefits</td>
<td>Greater control over data sources, collection methods, and overall quality</td>
<td>Someone else did all (or most of) the work, so quicker to “collect” data</td>
</tr>
<tr>
<td></td>
<td>Measures can directly assess the outcomes/indicators of interest to you</td>
<td>Often free or relatively inexpensive to obtain data</td>
</tr>
<tr>
<td>Costs</td>
<td>Can be labor-intensive, time-consuming, and thus expensive to collect and collate data</td>
<td>May not directly assess the outcomes/indicators of interest to you</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No control over quality of data</td>
</tr>
</tbody>
</table>

Figure 12. Differences between primary and secondary data.

As you can see, some trade-offs are involved in the decision to use primary or secondary data sources to measure particular indicators. Depending on your performance measurement needs, you can use your knowledge of these

---

tradeoffs to decide what kinds of data to use for different indicators in your measurement system.

For instance, if you need greater control over what you’re measuring and the quality of your data – say, if reliable and relevant secondary data sources are unavailable – primary data are likely best. If, however, sound data already exist on the indicators of interest to you, secondary data sources will often be the cheaper option. If you use secondary data, it’s very important to ensure that the data are actually assessing the outcomes and indicators of interest to you. Otherwise, the data will help you measure convenient things – but the wrong things, nonetheless. (This is another example of the streetlight effect – see chapter 4 for an additional example of this effect.)

Along these lines, you have a couple of options in using secondary data for performance measures, assuming that the data are of sound quality. One option is to use the secondary data in their raw form as a measure of key program outcomes. This option is advisable when the data are directly relevant to your outcomes and indicators. If, however, the secondary data are less immediately relevant to your outcomes and indicators, you may be able to transform the raw data into an estimate, aggregate, or some other form that more directly speaks to your program’s key outcomes. Often these kinds of indirect measures are known as proxy variables and can be very useful for measuring hidden things (see chapter 5 for a discussion on measurement challenges). For example:

- A manager might measure the hours/days spent at work or tardiness rates as a proxy for employee performance.
- Education level may sometimes be used as a proxy for intelligence.
- In conventional warfare, the quantity of an opponent’s tanks, ships, airplanes, and troops has been used as a proxy for military strength.

Note how the above proxy measures are different from indicators of fuzzy constructs. Whereas indicators of fuzzy constructs are encompassed by the construct itself, proxy variables are a step or two removed from the construct you want to measure. Let’s look at each example above to see how these proxy measures differ from indicators of fuzzy constructs.

- Proxy measures for employee performance might include hours/days at work or tardiness. Indicators of employee performance might include product outputs or customer satisfaction.

---

**Proxy Variable**

A variable that is used to measure an unobservable quantity of interest.

Education level might be used as a proxy measure for intelligence. An indicator of intelligence might include an IQ or standardized test score.

In the DHS context, proxy measures for illegal border crossings might include indirect evidence of illegal border crossings, such as footprints, trails, trash and other discarded items.

You can use a combination of primary and secondary data collection methods in your performance measurement system, depending on what tradeoffs are acceptable for each indicator of interest. To potentially reduce long-term costs, we advise dedicating some resources at the outset of the performance measurement effort to identifying and evaluating any relevant secondary sources of outcome data. If secondary data are available, relevant, and high quality, it’s generally more affordable than primary data collection.

If, however, secondary data sources on your outcomes and indicators are unavailable, low quality, or only marginally relevant to what you’re trying to measure, you can always plan for primary data collection instead. Although cutting costs is important, it’s not worth sacrificing the quality or relevance of your data to your program’s key outcomes. This kind of cost cutting could lead to many problems when you report and use your performance data to make important decisions. In scoping and measuring your indicators, try to strike a balance between the relevance and quality of your data on the one hand and the need to work within your resource constraints on the other.
6. Narrowing the Field: Deciding Which Outcomes and Indicators to Measure

ADDITIONAL RESOURCES

SELECTING INDICATORS


PRIMARY AND SECONDARY DATA


7. Special Challenges in Measurement

Taken together, all the chapters thus far in Part II have provided an approach for identifying, understanding, scoping, and ultimately measuring your program’s results. These chapters covered key concepts and considerations at each stage in this results-focused process. However, as we’ve mentioned throughout this resource, the DHS context poses some interesting challenges for program managers as they seek to measure their key program outcomes. This chapter tackles some of these challenges, so that you can better anticipate and plan for them.

As the final chapter of this resource, this chapter focuses on some of the most advanced and challenging situations and issues you may encounter as they seek to measure your program’s results. These challenges go beyond the fuzzy and hidden measurement issues discussed in chapter 5, and have much broader implications throughout the performance measurement process. Fortunately, plentiful guidance exists on these issues, as well as a variety of tools and techniques that can help you address them.

Although this chapter cannot capture the entirety of the literature on these advanced challenges and considerations, we can offer readers a conceptual-level understanding of these challenges and their implications in the DHS context. We also suggest some additional resources that offer methods for addressing or overcoming these challenges. Having a basic understanding of these issues can help you to anticipate and account for them, and to defend and acknowledge any limitations of your performance measures with key program stakeholders.
One major challenge for effective program performance measurement is the “causality issue.” That is, how do you know that your program is actually what’s producing – or at least contributing to – the outputs or outcomes it’s supposed to produce? Without knowing this, performance measures – no matter how well designed – may not tell you very much about how your program is actually doing at achieving the desired outputs/outcomes. In some cases, these causal links between a program and its outputs or outcomes will be very clear. For instance, let’s say that your program is supposed to produce a target number of widgets each month, and yours is the only widget shop in town. If performance measures show that the target number of widgets was reached for this month, then it’s not much of a leap to say that your program is what produced those widgets.

In the DHS context, however, it is often much more difficult to attribute any trends in measured outcomes or outputs to your program. For one, there are many external factors, or influences outside of the program, that may also affect the intended outputs or outcomes of the program. These external factors may include, for example, economic trends and resource limitations, weather patterns and natural disasters, or public opinion and political dynamics. The specific external factors that impact the intended outputs or outcomes of a program will vary according to the program’s environment. It is therefore important to assess the environment to anticipate these factors and their potential effects on programs’ intended outputs or outcomes.

Consider, for example, the “If You See Something, Say Something™” public awareness campaign, which was originally launched by New York’s Metropolitan Transportation Authority and was later adopted by DHS and many of its partners. One major purpose of this program is “to emphasize the importance of reporting suspicious activity to the proper local law enforcement authorities.” Accordingly, one output measure that may be useful to examine is the number of suspicious activity reports made to local law enforcement authorities in localities where the program is implemented. Let’s say that across a number of localities in the United States, suspicious activity reports in a given locality tend to increase after the “If You See Something, Say Something™” campaign is initiated in that locality. So can you...
say that the program actually caused the spike in suspicious activity reports? Not necessarily. It could be, for instance, that localities tend to adopt the campaign after a terrorist incident (attack, plot, etc.) has occurred, and that the spike in suspicious activity reports is actually due to heightened public vigilance following the incident, rather than the campaign. This is known as the third variable problem, in which a third, external variable (a terrorist incident) actually causes both the output/outcome (spike in suspicious activity reports) and the proposed “cause” (the program). Without further analysis to try to disentangle the effects of the program from those of other causes, you may encounter skepticism when reporting performance data.

In addition to external factors, it can also be difficult to disentangle the effects (if any) of your program from the effects of other programs that may influence the same outputs or outcomes. For instance, let’s say your program is intended to ensure effective emergency response to disasters. One key outcome measure might be the public’s perception of effective emergency response to specific disasters. However, your program is one of many programs in both the public and private sectors that contributed to emergency response during these disasters. If the performance rating is good, how do you know whether – and how much – your program contributed to the effective response? Was the response effective because of your program, or despite your program? And conversely, if the public gave poor ratings on response effectiveness, how much of the blame and responsibility – if any – should your program shoulder? Are modifications to your program necessary, or would closer scrutiny amount to wasted dollars? Overlap in different programs’ outputs and outcomes can make it difficult for you to tell whether your program actually contributed to the observed performance levels – and if so, how much did it contribute, relative to other programs?

---

To explain this point, let’s discuss the role of the Federal Air Marshal Service (FAMS) as a layer of security in preventing against an airline attack. The primary role of FAMS is “to detect, deter, and defeat hostile acts targeting U.S. air carriers, airports, passengers, and crews” within the commercial aviation system. The FAMS program is one of 20 “layers of security” identified by TSA designed to prevent an airline attack. Knowing this, you can count the number of layers of security in place to prevent airline attacks and then estimate the relative impact of the FAMS program versus the other layers of security. Because the FAMS program is one of 20 layers of security, the FAMS program could be estimated to account for 5 percent of the total deterrent impact of all 20 layers. Of course, if you had more information about the relative impacts of each security layer, or knowledge of the extent to which some programs overlap in their impacts or activities, you could weight your estimates for each layer or set of similar layers accordingly.

These causality considerations are among the most challenging performance measurement issues. Fortunately, many methods are available to help tackle these issues. Perhaps the simplest method is awareness. Simply being cognizant of and straightforward about any causality-related limitations of your performance data can help you to more effectively interpret, discuss, and field questions about it, particularly with leadership and other program stakeholders.

Beyond simple awareness, there are also many methods that can provide you with greater confidence in the causal impacts of your program. These methods are often discussed as part of program evaluation, which provides a much more granular look at specific program elements and their relationships, often seeking to find root causes of program outcomes. A detailed discussion of these methods can be found in the additional resources at the end of this chapter. In the call-out box, we provide a few key principles that underlie these methods and may be helpful for disentangling the effects of a program from the effects of other factors.

To illustrate these principles in action, let’s return to the example of DHS’s “If You See Something, Say Something™” campaign. The program manager in this situation needs to show that the campaign – and not the terrorist incident that drove its implementation – was what produced a spike in suspicious activity reports. To help establish this, she might compare the trends in suspicious activity reports across different localities that implemented the

---

74 Ibid., 49.
campaign: some that experienced a terrorist incident around the same time as the campaign was implemented, and some that did not. If a spike in suspicious activity reports is only present in localities where a terrorist incident had occurred, the program manager may be a bit more skeptical about the campaign’s supposed effects. However, if there was a spike in suspicious activity reports in both types of localities, the program manager can be more confident that the campaign actually contributed to this result.

Of course, all the usual principles of measurement apply here. For instance, the program manager would need to think carefully about how she’s defining a “locality that experienced a terrorist incident,” because terrorist incidents’ effects on public vigilance may reach far beyond the limited locality in which they occurred.

Numerous other methods are available for teasing out the effects of a specific program from the effects of other possible causes. Some of these methods are fairly difficult to implement, but many others are relatively simple and cost effective. Regardless of your ability to use any of these methods, the fact remains: simply being aware of potential limitations on your performance data can go a long way in interpreting, reporting, and defending it with key stakeholders.
Unintended Consequences

In all the chapters thus far, we’ve talked about strategies for identifying, defining, and measuring the desired or intended outcomes of DHS programs. This is all very useful for understanding whether your program is achieving its goals and fulfilling its purpose. But what about the *unintended* outcomes of your program – the consequences you may not expect or be looking for? These unintended outcomes aren’t necessarily all bad – they could be positive or negative, or effectively neutral. So is it important – or perhaps, *when* is it important – to consider and measure them?

Considering the unintended consequences of your program – whether you can measure these consequences or not – is useful for a number of reasons. For one, negative unintended consequences are often the things that come back to haunt program managers in program reviews, audits, and the public eye. For example, a commonly cited unintended consequence of the Social Security program is that American workers now save less for their retirement than they did in the past.  

75 Martin S. Feldstein, “Social Security, Induced Retirement, and Aggregate Capital Accumulation,” *Journal of Political Economy* 82, no. 5 (September-October 1974): 905-926. Notably, this unintended consequence has been the subject of much scrutiny and debate by later researchers.  


This sort of controversy over unintended consequences can happen fairly often. They are difficult to establish for the same reason that any program impacts – intended or unintended – are difficult to establish: causality issues (see previous section for more on this). Despite this difficulty, it is still important and can be highly beneficial to consider unintended consequences when measuring your program’s performance. By thinking through the potential unintended impacts of your program, you can anticipate and perhaps fend off those “Whoops, we didn’t mean to …” moments, rather than waiting for hindsight to reveal the problems.

For instance, consider a 2011 report from the Homeland Security Advisory Council’s Task Force on Secure Communities.  

which the FBI automatically sends fingerprints of anyone arrested or booked by police for a state or local criminal offense to DHS to check against its immigration and enforcement records so that ICE can determine whether that person is a criminal alien or falls under ICE’s civil immigration enforcement priorities.”78 The task force’s report describes some possible unintended consequences of the Secure Communities program, including, for instance, potential chilling effects on crime reports by members of immigrant communities. The report further notes that although the Secure Communities program was designed to minimize these chilling effects by obtaining information only on persons arrested and fingerprinted, DHS may consider establishing and enforcing systematic mechanisms to provide protections for victims and witnesses of crime and domestic violence.79

It can also be useful to anticipate and measure the positive unintended consequences of your program. In some cases, these consequences may suggest additional value propositions for your program that will help you argue for its continued existence and perhaps make a case for greater resources. One example of a positive unintended consequence could be a farmer who lives next to a beekeeper and benefits from the bees that pollinate his crop, thus boosting his crop production.80

Considering and measuring the unanticipated consequences of your program can also provide valuable data to other program managers, whose programs may actually be intended to impact those outcomes. Recall the discussion in the causality section about teasing apart the impact of your program from those of other programs. Just as you may face this issue, other program managers may also wrestle with it. By measuring the unintended consequences of your program on the outcomes that matter to other program managers, you can provide a valuable data point that will help them to disentangle their programs’ effects from yours.

So how can you go about identifying these unintended consequences of your program? Strange as it may seem, some of the best sources for identifying unintended program outcomes – particularly the negative ones – may be the program’s biggest critics. If controversy and debate is already buzzing around your program, it is likely that one of the opposing parties has already identified some possible negative impacts of your program.

SOURCES OF UNINTENDED CONSEQUENCES

In 1936, Robert Merton became one of the first scholars to discuss unintended consequences. In his article, he proposed five major sources of unintended consequences:

- **Ignorance**: uncertainty, unknowns, and unknown unknowns (didn’t see it coming)
- **Error**: faulty assumptions and oversights (should have seen it coming)
- **Immediacy of interest**: failure to consider unintended consequences, due to intense interest in achieving the intended outcomes (didn’t want to see it coming)
- **Basic values**: paradoxical consequences of values-driven actions that actually undermine the values and intended purpose of the actions (doesn’t matter if I saw it coming – I had to do it anyway)
- **Self-defeating predictions**: unanticipated consequences stemming from human reactions and responses to expected consequences (I said I saw it coming, and other people kept it at bay)

Thinking critically through these possibilities can help you and other program stakeholders to identify the unanticipated consequences you may have overlooked.

— Robert K. Merton, “The Unanticipated Consequences of Purposive Social Action,” American Sociological Review 1, no. 6 (December 1936), 894-904.

---

These critics’ arguments may reveal unintended outcomes that are worth further consideration.

Another useful resource for identifying your program’s unintended consequences is you – and your program’s key stakeholders. What keeps you up at night? What issues or second-order problems do you worry your program might accidentally unleash? If you can envision the outcome happening, it likely qualifies as a potential unintended consequence. If your stakeholders see danger looming over the horizon, it is likely a scenario worth considering as a potential unintended consequence. It’s important here to consider the concerns of a range of stakeholders, including those with a more tactical, on-the-ground perspective as well as those with a higher-level, strategic perspective. This will help you to capture unintended consequences stemming from both the day-to-day grind and the broader context of your program.

Once you’ve identified and can better anticipate the potential unintended consequences of your program, the next step is to decide whether to measure these outcomes. Many of the same factors that constrain the scope of your measures (discussed in chapter 6) can also constrain the number of unintended consequences you measure. However, you may have more of an uphill battle to fight if you want to measure unintended consequences, because these are often unrelated to the program’s purpose.

For instance, resource limitations will play an important role in determining whether and how many unintended consequences you can feasibly measure. If you’re already forced to scope down the measures of outcomes your program is supposed to impact, you probably don’t have the resources to measure its unintended effects. In these cases, however, you can still seek to identify and understand what these unintended consequences might be, so that you can anticipate and better prepare for the potential fallout. It’s also advisable to work closely with leadership and other key stakeholders in deciding whether to measure unintended program consequences – and if so, which ones – so that external factors (such as political optics) can be considered, and nobody is blindsided by data that reveals negative unintended consequences.

**Shared Program Ownership**

In the DHS context, it is not uncommon for programs to be “owned” by multiple stakeholders. These program owners may include other program managers, funders, customers, employees or volunteers, political leadership
from multiple branches of government, or – thinking broadly – the American people. Additional program owners may include:

- other DHS program managers from other offices or components;
- other federal departments and agencies;
- state, local, tribal, or territorial governments;
- foreign governments and other international partners;
- private sector organizations, such as NGOs, universities, or community groups; or
- private citizens.

Consider, for instance, DHS information-sharing programs. Almost by definition, DHS information sharing is a complex, multi-stakeholder endeavor that involves all levels of government, the private sector, private citizens, and even foreign partners. This endeavor may also involve multiple specific programs or efforts that are each owned by multiple stakeholders. For example, fusion centers – which facilitate sharing of threat information with federal, state, local, tribal, territorial, and private-sector partners – are owned and operated at the state and local levels. However, fusion centers are governed at the federal level by the Fusion Center Subcommittee of the Information Sharing and Access Interagency Policy Committee (ISA IPC). This ISA IPC Fusion Center Subcommittee is jointly led by DHS and the FBI, and other members include “the U.S. Department of Defense, the Joint Staff, the U.S. Department of Justice, the Office of the Director of National Intelligence, the PM-ISE, the U.S. Drug Enforcement Administration, the White House Office of National Drug Control Policy, the U.S. Department of Health and Human Services, and the Bureau of Alcohol, Tobacco, Firearms and Explosives.”

As you might imagine, multi-owner programs can raise some interesting issues for those seeking to understand and measure their programs’ outcomes. For one, the outcomes that matter to you as a DHS program manager may be very different from those that matter to program owners outside the federal government. It’s important to work with other program owners and stakeholders whenever possible to ensure that you’ve identified and understand the full range of your program’s potential outcomes.

Working with other program owners is also useful when you’re determining which of these many outcomes to measure. Some program outcomes may be relevant to more than one program owner, and could feasibly be...
measured by multiple parties. In these cases, it will be important to clarify roles and responsibilities for measurement so that none of the important program outcomes fall through the cracks. Who will be keeping track of each important program outcome? How are they going to go about measuring it? Is the measure designed in a way that will satisfy other program owners’ reporting requirements or other needs for information on this outcome?

When deciding who will measure which outcomes, a couple of situations are possible. In some cases, multiple program owners may be required to measure the same outcome, perhaps due to congressional requests, departmental or programmatic requirements, and so forth. In these cases, redundancy of measurement across different program owners may be unavoidable. In other situations, program owners may have more flexibility in what they choose to measure, and may decide to divvy up the outcome measurement responsibilities to avoid redundancy. In the latter case, it’s important to ensure that all program owners understand and agree on the measures that are relevant to them, so that they are getting the information they need on the outcomes that matter.

If agreement can’t be reached, or more than one measure of the outcome is necessary to ensure all the desired information is obtained, another option is for different program owners to measure the same outcome in slightly different ways. In fact, this sort of triangulation is a common method used to bolster – or at least check – your outcome measures’ robustness; if multiple measures of the same outcome yield converging evidence of a trend or pattern, you can be a bit more confident in your data and conclusions. As always, the limiting factor here is the amount of resources you have available for measurement.

---

THREE QUESTIONS FOR MULTI-OWNER PROGRAMS

In multi-owner programs, the program funding and outcomes are often shared among owners. A 2003 White House report recommends using your performance information in these cases to answer three key questions:

- Is the overall effort working?
- Is the federal contribution making a difference?
- How is the funding effort shared between federal and non-federal partners? Does the distribution of funding compare to the distribution of need or benefit?

---

TRIANGULATION

“Use of more than one approach to the investigation of a research question in order to enhance confidence in the ensuing findings.”

---

ADDITIONAL RESOURCES

GENERAL RESOURCES ON ADVANCED TOPICS AND CHALLENGES


RESOURCES ON SELECT ADVANCED TOPICS AND CHALLENGES

Causality and Impact Analysis


Unintended Consequences


Shared Goals and Program Ownership


8. Conclusion

In this resource, we have covered a broad range of concepts and principles that you as a DHS program manager can apply in measuring your program’s performance. Throughout, the common thread has been a focus on the importance of measuring for results – that is, identifying and measuring the outcomes that matter for your program and its stakeholders. Although you can and often should measure many elements of your program – inputs, processes, outputs, and outcomes – a focus on outcomes will help you assess whether your program is making the difference it was intended to make.

Measuring outcomes is not without its challenges, particularly in the DHS context. We’ve covered some of the common problems and pitfalls you may encounter as you seek to measure their programs’ results. We’ve also explored several strategies for tackling these problems and provided additional resources to help you enact these strategies as needed in your work.

As you’re thinking through the set of outcomes and indicators that represent the most important program results, you may find it useful to think about your program performance measures as a system. This system will likely involve multiple outcomes that span different time horizons and organizational or strategic levels, and could include any number of indicators that represent each of these outcomes.84

Beyond outcomes, which have been the primary focus of this resource, one could also incorporate the other elements of the conceptual framework into the program performance measurement system. Recall the elements of the conceptual framework from Part I.

---

84 Other sources may refer to performance measurement systems as suites of measures. See, for instance, Metrics for Measuring.
One could create a program performance measurement system that assesses each of these elements for any given program. Let’s consider a specific example in the DHS context. In a 2012 HSSAI report, Fleming and Goldstein proposed numerous metrics for measuring the efficacy of critical-infrastructure-centric cybersecurity information-sharing efforts. These metrics all assumed that:

- “information-sharing should be goal-directed;
- information should be shared with entities who can effect or affect achievement of the goal – but not with those who cannot;
- shared information should be used for purposes that can effect or affect achievement of the goal – but not for purposes that cannot;
- shared information should be fit for the purpose of reducing uncertainty (i.e., comprising data + meaning, relevant, timely, accessible, accurate); and
- information sharing cannot reduce all uncertainty (in some cases, it may increase it).”

The recommended suite or system of measures that Fleming and Goldstein proposed for cybersecurity information sharing is depicted in Figure 17, per our conceptual framework.

Note the critical link between the program purpose and outcome measures. In this example, the purpose of any given cybersecurity information-sharing program may vary, depending on the specific context for the information sharing program or effort. Therefore, the contents of the outcome measure – whether the goal of information sharing was achieved – will also vary depending on what that goal is.

Taken together, this suite of measures could be viewed as a system, a complex whole made up of connected parts or elements. You can examine many aspects of such a measurement system to arrive at useful conclusions that might drive improvements to your program's performance. For instance, does the program’s purpose drive the inputs, which feed into the processes, which

---

**Figure 16. A conceptual framework for program performance measurement.**
yield the outputs, which result in outcomes that are congruent or aligned with the program’s purpose? If data are available for all the metrics on all the program elements, do the data support the linkages or relationships between the inputs, processes, outputs, and outcomes? Does each element in turn predict or correlate with the following element?

And at a more basic level, do the metrics in your measurement system provide a comprehensive picture of the key elements of your program? Are you capturing all the important inputs, processes, outputs, and outcomes of your program? Are any of your measures redundant? Do you have the right metrics in the right “buckets” – for instance, are your outcome metrics actually measuring outcomes, or some other program element? By viewing your program performance measures as a system, it is easier to spot the alignments and misalignments, the strengths and gaps. The conceptual framework we’ve presented, along with our guidance on how to measure for results, can inform your measurement and improvement of your program’s performance on the outcomes that matter.

As noted in the preface, our hope is that this resource can serve as a valuable resource and reference for program managers – and other stakeholders – as
they work toward measuring the performance of individual DHS programs, but also of DHS divisions, offices, components, directorates, and the homeland security enterprise in its broadest sense. Measuring for results helps reduce uncertainty about the value of DHS programs and efforts at all levels and reveals the difference DHS is making. Good performance measurement can and does inform decisions that make the nation’s homeland security enterprise more effective and efficient. It is our hope that this resource provides you with a foundation of knowledge and tools to help you measure for results.
Activity: See process.

Balanced Scorecard: a common method for identifying the most important things to measure in an organization or program. The balanced scorecard approach advocates that you consider and measure key performance indicators (KPIs) from four different perspectives or domains: financial/stewardship, customer/stakeholder, internal business processes and, organizational capacity.87

Construct: a concept, typically one that is relatively abstract, intangible, and often multi faceted.88

Construct validity: the extent to which your measures are actually assessing the construct you want to measure. Construct validity includes many elements or components that could be assessed individually or together to establish construct validity. These elements include, for instance, face validity, content validity, concurrent and predictive validity, and convergent and discriminant validity.

Conceptual framework: an adapted program logic model to help illustrate the basic elements of a program and the relationships between them. It can be used to conduct results-based program performance measurement.


Content validity: the extent to which you’ve identified and measured all the major facets of a construct, such as threat, vulnerability, and consequence for the risk construct.\(^9\)

Data: “facts and statistics collected together for reference or analysis.”\(^90\)

Denominator management: When using a ratio management method for measuring performance, the tendency to adjust the denominator to give the impression of improvement in overall performance.\(^91\)

External factors: factors outside of a program that may impact the program’s inputs, processes, outputs, or outcomes.\(^92\)

Face validity: the extent to which a measure seems “on its face” to measure the intended construct.\(^93\)

Fuzzy concepts: concepts that “do not easily lend themselves to quantification.”\(^94\)

Goals: “a target level of performance expressed as a tangible, measurable objective, against which actual achievement can be compared.”\(^95\)

Hidden concepts: concepts “for which relevant information … is not available.”\(^96\)

Horizontal dimension of measurement: the time horizon when you would expect your program’s outcomes to materialize. Some outcomes may be near term, others may be mid term, and still others may be long term.

Indicators: the observable variables used to represent, or provide evidence of, an underlying construct.\(^97\)

---


**Inputs:** the resources used to undertake activities, produce outputs, and influence outcomes

**Key performance indicators (KPIs):** the measures that provide the most important performance information

**Legacy measures:** inherited measures that already exist for some specific purpose

**Line of sight:** captures the alignments between your program's lowest-level, most immediate impacts and its higher-level, more distant impacts. Also used to capture alignments between program elements, including the program's purpose, inputs, processes, outputs, and outcomes.

**Logic Model:** "A picture of how your organization does its work – the theory and assumptions underlying the program. A program logic model links outcomes (both short- and long-term) with program activities/processes and the theoretical assumptions/principles of the program."

**Measures:** the quantitative and qualitative information collected on key aspects of a program

**Measuring for Results:** Monitoring the vital indicators of a program's outcomes in a way that enables decision makers to be aware of and improve their results. It enables program managers to assess whether the program's actual results are aligned with its intended results, and make any necessary changes to improve this alignment.

**Measurement Fatigue:** When you attempt to measure too many things, the people providing the data become tired, bored, and/or exasperated.

**Metric:** "a standard or unit of measure"

**Missions:** comprehensive statements that cover the major functions and operations of the agency, each typically including multiple lower-level goals and objectives

---


Objectives: a lower-level statement of intent, generally under a goal, See Goal.

Outcomes: “the desired results of a program”104 They are also interpreted as the impact of a program.105

Output: he “amount of products or services delivered”106

Performance management: “the use of performance information to affect programs, policies, or any other organization actions aimed at maximizing the benefits of public services”107

Performance measurement: the “ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals,” where a program is defined as “any activity, project, function, or policy that has an identifiable purpose or set of objectives”108

Primary data sources: data collected by you. Examples include surveys, interviews, focus groups, observations.109

Process: a program activity, workload, or other effort110

Program: a set of inputs, processes and, outputs used to achieve a desired mission, goal, or objective

Program manager: anyone responsible for a program

Proxy variable: “a variable that is used to measure an unobservable quantity of interest”111

Purpose: a statement about what the program intends to accomplish or its desired impact112

---

105 Metrics for Measuring, 18.
Results: See Outcome.

Risk: “the potential for an adverse outcome assessed as a function of threats, vulnerabilities, and consequences associated with an incident, event, or occurrence”

Secondary data sources: data collected by others. Examples include archival records, open-source or social media, transcripts, and databases.

Stakeholders: individuals interested in performance measurement, such as DHS leadership or those responsible for creating a performance measurement approach for an organization

Streetlight Effect: the tendency to use measures that are readily available or easy to obtain

System: “a set of connected things or parts forming a complex whole, in particular”

Triangulation: “use of more than one approach to the investigation of a research question in order to enhance confidence in the ensuing findings”

Tornado Plot: ranks the inputs to a model or estimate according to their impact on the model outputs

Variable: any factor that can take on more than one value. Resource levels, activity levels, products generated, and public opinion are all variables, because they can have more than one value (e.g., positive, negative, or neutral).

Vertical dimension of measurement: the levels at which you are defining your outcomes. These levels could be defined in terms of the many layers of the organization: your division, office, and component, all the way up to the departmental, national, or even global level. They could also reflect the different levels at which a program’s purpose is defined: the objectives, goals, missions, or other formal or informal strategic statements of purpose.

---

Several authorities and guiding documents mandate performance measurement and specific characteristics of good measurement systems for federal programs. The most relevant government authorities are the Government Performance and Results Act of 1993 (GPRA) and its update, the GPRA Modernization Act of 2010 (GPRAMA). In addition, the Office of Management and Budget (OMB) has published Circular No. A-11, which provides instructions for how agencies should implement GPRAMA. Other authorities that contain performance measurement language include the Homeland Security Act of 2002, as well as Executive Orders 13450 and 13576.

The following list provides a sampling of the mandates and recommendations found in this guidance:

- GPRAMA states that agencies must “establish a balanced set of performance indicators to be used in measuring or assessing progress toward each performance goal.”\(^\text{118}\)
- OMB says that agencies should report progress on their performance goals more frequently than annually.\(^\text{119}\)
- OMB requires that performance information must be accurate and reliable for its intended use, and requires the use of verification and validation techniques to ensure this.\(^\text{120}\)
- Subtitle D of the Homeland Security Act of 2002 instructs the Department of Justice Office of Science and Technology to establish and maintain performance standards in accordance with the National Technology

\(^{119}\)OMB Circular No. A-11, Part 6, Section 260.2
\(^{120}\)OMB Circular No. A-11, Part 6, Section 260.9
Transfer and Advancement Act of 1995, and Title VII of the same act requires DHS to identify and track performance measures.\(^{121}\)

- Executive Order 13450, “Improving Government Program Performance,” makes it the official policy of the federal government to spend taxpayer dollars efficiently and effectively.\(^{122}\)
- Executive Order 13576, “Delivering an Efficient, Effective, and Accountable Government,” established committees and requires studies to facilitate GPRA mandates by streamlining and cutting waste from government procurement.\(^{123}\)

In addition to the many motivators for establishing and maintaining sound performance measurement practices, a number of recent GAO and DHS OIG reports have identified important gaps in DHS program measurement. These reports have increased the profile and priority of performance measurement in DHS. For instance, in an April 2013 report to congressional requesters, GAO concluded that DHS needed to develop outcome-oriented measures if it intended to gauge cybersecurity effectiveness in the future.\(^{124}\) This criticism is not confined to the realm of cybersecurity, unfortunately. The following bullets provide some additional examples of GAO and DHS OIG reports that identify various shortcomings in DHS performance measurement practices.

### DHS OIG Reports

- The DHS OIG, in an audit of the CBP, recommended that CBP establish “updated, formal guidance for monitoring and benchmarking performance.”\(^{125}\)
- The DHS OIG, in an audit of TSA’s Screening of Passengers by Observation Technique (SPOT) program, stated that “TSA cannot accurately assess the effectiveness or evaluate the progress of the SPOT program. The program does not have a finalized strategic plan that identifies the mission, goals, and objectives needed to develop a system of performance

---

measures.” Specifically, the OIG pointed out that “TSA had not developed performance measures for the SPOT program.”


A 2013 DHS OIG report found that “CBP is challenged in its ability to measure its performance and effectiveness, its technological capability to detect illicit cross-border tunnels, and in its use of air and marine assets.”


GAO Reports

- The GAO, in a report on DHS's Chemical Security Program, recommended that DHS develop performance measures so that it would be better equipped to identify any gaps between actual and expected results and take corrective action.


- A GAO report on measuring performance in ICE's Mexican Interior Repatriation Program (MIRP) found that ICE “has not established performance measures to evaluate its progress in meeting its objective consistent with internal control standards. Thus, ICE does not know the effectiveness of its efforts related to MIRP at deterring individuals from illegally returning to the United States.”


- In relation to acquisition management, a 2013 GAO report found that DHS needs to show measurable progress in developing performance metrics for their biggest programs.


- GAO has also found that DHS has not created outcome-based performance measures for four FEMA grant programs, making it difficult to assess their effectiveness.

131 Additionally, a 2013 DHS OIG report found
that FEMA did not create “good” measures of how effective preparedness grants were at increasing overall preparedness.\textsuperscript{132}

- GAO found that when developing the Arizona Border Surveillance Technology Plan DHS did not develop performance metrics and therefore could not assess the effectiveness of the plan’s implementation.\textsuperscript{133}
  
  Overall, GAO found that due to a lack of performance information, DHS was unable to enhance its performance across all missions.\textsuperscript{134}

- A 2014 GAO report found that “DHS has collected some data on contracts awarded to small businesses under strategic sourcing initiatives, but it … generally did not have baseline data and performances measures to determine how small businesses were affected by strategic sourcing.”\textsuperscript{135}

  The lack of data makes it difficult to determine the effectiveness of these strategic sourcing initiatives on small businesses.\textsuperscript{136} Therefore, GAO recommended that the Secretary of DHS collect the necessary baseline data and create performance measures to determine the effectiveness of strategic sourcing initiatives for small businesses.\textsuperscript{137}

Although these are some of the major authorities and reports on federal and DHS performance measurement, others may also be useful. For additional resources that outline parameters, variables, and considerations for DHS program performance measurement, see chapter 3.


\textsuperscript{133} Berrick, Department of Homeland Security: Progress Made and Work Remaining, 7.

\textsuperscript{134} Ibid., 13.


\textsuperscript{136} Ibid.

\textsuperscript{137} Ibid., 20.
APPENDIX C: The Program Logic Model

In Part I of this resource, we introduced a conceptual framework that was based on concepts from the program logic model. This appendix provides further information and guidance on the logic model. Alternative frameworks could also be used to help you assess your program’s performance; however, we adapted the logic model because of its intuitive, start-to-finish approach.

According to the W.K. Kellogg Foundation, “a logic model is a systematic and visual way to present and share your understanding of the relationships among the resources you have to operate your program, the activities you plan, and the changes or results you hope to achieve.”

As the Kellogg Logic Model Development Guide notes, “The purpose of a logic model is to provide stakeholders with a road map describing the sequence of related events connecting the need for the planned program with the program’s desired results.” It is an effective way to aid in a program’s success because it helps to systematically organize the plans, management, and functions of the program.

We suggest that the reader review the Kellogg Foundation’s Logic Model Development Guide for a thorough discussion of the purpose, use, and design of program logic models. The organization’s website offers a free PDF of the guide for download. It can be accessed at the W.K. Kellogg Foundation website: http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide.

139 Ibid., 3.
140 Ibid., 5.
Appendix C: The Program Logic Model

Figure 18. How to read a logic model. Taken, with the permission of the W.K. Kellogg Foundation, from the W.K. Kellogg Foundation Logic Model Development Guide.

The Kellogg Guide presents a depiction of a basic program logic model and provides instruction on how to read one. We reproduce this diagram here for the reader’s instruction, with the permission of the W.K. Kellogg Foundation.

In addition to this general guidance on logic models, “how-to” guidance is also available for building a logic model that is tailored to the specifics of the program. One example of a step-by-step process for building a program logic model is provided in the margin.

1. **Collecting information:** The program manager, or a designated work group, must gather as much information as possible about the program. This information can be gained by studying existing program documentation, interviewing key stakeholders, researching related or historical efforts, and reviewing applicable laws, regulations, or guidance.141

2. **Describing the goal of the program and influencing factors:** For program managers in the public sector, the goal or purpose of the program, as well as associated objectives, are often provided. In these cases, the program manager should be careful to accurately articulate

---

the program goal and objectives and be able to explain it to program stakeholders. In those cases where the program manager has some latitude to define the goal of the program, they should ensure that the goal seeks to address the key problem(s) identified as a result of the information collected in step 1. In either case, the program manager should also be able to identify any external factors that may also influence the program and its outputs or outcomes. Accounting for their influence will allow the program manager to plan for their impact on the program inputs, processes, outputs, or outcomes. For more on identifying a program’s purpose and relevant external factors, see chapters 3 and 7.

3. **Defining the components of the logic model:** All of the information collected in steps 1 and 2 of this process can be categorized into different bins that correspond to the components of the logic model: inputs, processes, outputs, and outcomes. As these bins are filled and the linkages between them made explicit, the program manager should continually check with key stakeholders to ensure accuracy. Importantly, the logic model can be used to describe either the as-is state of the program or the desired (to-be) program logic. If the program is in the planning stages, the model will be entirely theoretical, a vision of how the program should work if executed well. If the program already exists, it may be more useful to define the inputs and processes of the program as is but define the outputs and outcomes of the program as intended or desired. Then evaluate whether the program’s inputs and activities are actually producing the intended outputs and outcomes.

4. **Drawing the logic model:** The logic model is typically presented as a series of boxes underneath headings which correspond to the elements of the model. The boxes are meant to be read from left to right, and arrows show the linkages between each. These linkages could be drawn simply with the whole set of program inputs leading to the whole set of program activities or processes, and so forth. They could also be drawn to show more complex relationships between specific subsets of program inputs and activities, activities and outputs, and outputs and outcomes. For instance, maybe only one program activity produces a particular output, whereas all of the program activities feed together into another output. The level of complexity a program manager should draw in the model depends on its intended use – for instance, succinctly describing a program to senior leadership versus comprehensively...
documenting the logic for a program evaluation – and different versions of the model could be created for different purposes.

5. **Verifying the logic model:** As stated in step 3, the program manager should continually check the logic model for accuracy throughout the process of developing it. For the final check, the program manager and key stakeholders can walk through four questions:145

- **Is the level of detail sufficient?** Does the model provide a sufficient level of granularity on specific inputs, processes, outputs, and outcomes? Or is it too granular? As noted above, the use case and intended audience for the model will often determine the appropriate level of detail to draw in the model.

- **Are the key elements of the model addressed?** Does the model capture all of the relevant inputs, processes, outputs, and outcomes? Conversely, are there any extraneous factors that, upon closer examination, aren’t actually relevant to the program?

- **Is the model logically sound?** Do the linkages or relationships between elements of the program model make sense? The outputs and outcomes should clearly follow from the inputs and activities linked to them, for instance. If they don’t, there is likely a problem with either the theoretical program model or the actual program’s strategy and operations, depending on whether the model is as-is or to-be.

- **Have all of the external factors been identified and incorporated?** Think about the program’s environment – the broader system in which it’s embedded. Politics, society, economics, the landscape of other programs across the public and private sectors – all of these and more may be relevant to the program and its intended outputs or outcomes. The program manager and other stakeholders know this environment best. Has the landscape been covered?

For more on program logic models, see the resources cited in this section, as well as the additional resources provided at the end of chapter 2.

145 Ibid., 70.
APPENDIX D: Bibliography


Peng, Chao-Ying Joanne and Daniel J. Mueller. “Construct Validity.” In Sage Encyclopedia of Social Science Research Methods, edited by Michael S. Lewis-


umass.edu/oapa/oapa/publications/online_handbooks/program_based_assessment/ch2.php.


