Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade

Desk Reference

Christopher Paul, Jessica Yeats, Colin P. Clarke, Miriam Matthews

*Cover image: A service member and interpreter talk with local residents about malaria prevention in
Ethiopia (U.S. Navy photo by Petty Officer 2nd Class Nathan Laird).*

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

www.rand.org

Preface

This desk reference compiles and presents a wide range of observations about and recommendations for improving assessment of U.S. Department of Defense (DoD) efforts to inform, influence, and persuade (IIP). It was developed as part of the project "Laying the Foundation for the Assessment of Inform, Influence, and Persuade Efforts," which sought to identify and recommend selected best practices in assessment and evaluation drawn from existing practice in DoD, academic evaluation research, public relations, public diplomacy, and public communication, including social marketing.

The contents are part advice to policymakers, part advice to assessment practitioners, and part reference guide on the subject. While the core audience consists of stakeholders and practitioners involved in conducting or evaluating DoD IIP efforts (through both information operations and the various information-related capabilities), the assessment principles extolled here should be applicable across a wide range of defense undertakings.

An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*, distills the best practices, lessons, and recommendations presented here in a quick-reference format tailored specifically to personnel who are responsible for planning, executing, and assessing DoD IIP efforts.[1]

This research was jointly sponsored by the Rapid Reaction Technology Office in the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics and the Information Operations Directorate in the Office of the Under Secretary of Defense for Policy. The research was conducted within the International Security and Defense Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community, under contract number W91WAW-12-C-0030.

---

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

For more information on the International Security and Defense Policy Center, see http://www.rand.org/nsrd/ndri/centers/isdp.html or contact the director (contact information is provided on the web page).

Contents

Figures, Tables, and Boxes

Figures

Tables

Boxes

Summary

The U.S. Department of Defense (DoD) spends more than $250 million per year on information operations (IO) and information-related capabilities (IRCs) for influence efforts at the strategic and operational levels. How effective are those efforts? Are they well executed? How well do they support military objectives? Are they efficient (cost-effective)? Are some efforts better than others in terms of execution, effectiveness, or efficiency? Could some of them be improved? If so, how? Unfortunately, generating assessments of efforts to inform, influence, and persuade (IIP) has proven to be challenging across the government and DoD. Challenges include difficulties associated with observing changes in behavior and attitudes, lengthy timelines to achieve impact, causal ambiguity, and struggles to present results in ways that are useful to stakeholders and decisionmakers.

This desk reference addresses these challenges by reviewing and compiling existing advice and examples of strong practices in the defense sector, industry (including commercial marketing and public communication), and academia (evaluation research), drawn from a comprehensive literature review and more than 100 interviews with subject-matter experts across sectors. It then distills and synthesizes insights and advice for improving the assessment of DoD IIP efforts and programs.

An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*, covers many of the topics addressed here and is tailored specifically to personnel who are responsible for planning, executing, and assessing DoD IIP efforts.[1]

Methods and Approach

This research relied primarily on literature review and subject-matter expert (SME) interviews. The project team interviewed more than 100 experts with a range of roles in

---

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

government, industry, and academia. (Table S.1 provides a list of SMEs by sector.) In addition to SME interviews, a copious and wide-ranging literature review considered hundreds of documents (e.g., reports, assessments, doctrine, guidance, strategy papers, articles, white papers, textbooks) across the same sectors.

Once we compiled the practices, principles, advice, guidance, and recommendations, we distilled and synthesized this material for application within DoD. We grouped observations and insights topically, and this approach guided the structure of this report.

Table S.1
Number of Interviews Conducted, by Sector

Sector	Description	SMEs Interviewed
Industry		
Marketing/ public relations	Professionals in the marketing, advertising, or public relations fields in the for-profit sector	18
Public communication	Practitioners in public communication (including social marketing) or public communication evaluation in the nonprofit sector	26
Academia		
Evaluation research	Academics specializing in evaluation research (not necessarily IIP)	10
IIP evaluation	Academics specializing in influence or persuasion, with relevant expertise in IIP measurement, assessment, or evaluation	22
Media evaluation	Academics specializing in media evaluation	11
Defense		
Practitioners	Uniformed military, civilian, or contractor personnel with experience conducting or assessing defense IIP efforts	33
Academics/ think tanks	Academics or scholars who have conducted research on IIP or IIP assessment in the defense context	8
Other government representatives		
Practitioners	Personnel from elsewhere in government (beyond DoD) with experience assessing government IIP efforts	8
Congressional staff	Former or current congressional staff interviewed for stakeholder perspectives	5

Good Assessment Practices Across Sectors

Across all the sectors reviewed in our study (industry, academia, and government), certain headline principles appeared again and again. We collected and distilled the most central (and most applicable to the defense IIP context).

Effective Assessment Requires Clear, Realistic, and Measurable Goals

How can you determine whether an effort has achieved its desired outcomes if the desired outcomes are not clear? How can you develop and design activities to accomplish desired goals if the desired goals have not yet been articulated? How can you evaluate a process if it is not clear what the process is supposed to accomplish? While the importance of setting clear goals may appear to be self-evident, too often, this obvious requirement is not met. Good assessment demands not just goals but clear, realistic, specific, and measurable goals.

Effective Assessment Starts in the Planning Phase

Assessment personnel need to be involved in IIP program planning to be able to point out when objectives are not specified in a way that can be measured and to make sure that data collection is part of the plan. Likewise, planners need to be involved in assessment design to ensure that assessments will provide useful information and that they will have stakeholder buy-in. Building assessment into an IIP effort from the very beginning also allows the impact of the effort to be tracked over time and failures to be detected early on, when adaptations can still be made.

Effective Assessment Requires a Theory of Change or Explicit Logic of the Effort Connecting Activities to Objectives

Implicit in many examples of effective assessment and explicit in much of the work by scholars of evaluation is the importance of a *theory of change*. A theory of change, or the logic of the effort, is the underlying logic for how planners think elements of an activity, line of effort, or operation will lead to desired results. Simply put, it is a statement of how you believe the things you are planning to do will lead to the objectives you seek. When a program does not produce all the expected outcomes and you want to determine why, a logic model (or other articulation of a theory of change) really shines.

Evaluating Change Requires a Baseline

While both the need for a baseline against which to evaluate change and the importance of taking a baseline measurement before change-causing activities begin seem self-evident, these principles are often not adhered to in practice. Without a baseline it is difficult to determine whether an IIP effort has had its desired impact—or any impact at all. You cannot evaluate change without a starting point.

Assessment over Time Requires Continuity and Consistency

Continuity and consistency are essential to the assessment of DoD IIP efforts. Behaviors and attitudes can change slowly over long periods, and data must be collected over the long term to provide an accurate picture of an effort's impact and to determine whether that impact was attributable to the effort itself or to some change in the context of the effort. If the data or the way they are collected were to change during that time, it would become harder to tell whether observed changes are due to changes in the behaviors or attitudes of interest or simply to changes in how these behaviors or attitudes are being measured. All military activities face a challenge in this area due to individual, unit, and command rotations, and IIP efforts are no exception.

Assessment Is Iterative

Assessment is an inherently iterative process, not something planned and executed once. It is unusual for an IIP effort to remain static for long, particularly in a complex environment. The context of an IIP effort can change over time, as can an effort's objectives or the priorities of commanders and funders. Assessment must be able to adapt to these changes to help IIP efforts make course corrections, and it must be able to evolve with the efforts themselves.

Assessment Requires Resources

Organizations that routinely conduct successful and strong evaluation have a respect for research and evaluation ingrained in their organizational cultures, and they dedicate substantial resources to evaluation. Unfortunately, assessment of DoD IIP efforts has been perennially underfunded. That said, some assessment (done well) is better than no assessment. Even if the scope is narrow and the assessment effort is underfunded and understaffed, any assessment that reduces the uncertainty under which future decisions are made adds value. And not all assessment needs to be at the same level of depth or quality. Where assessment resources are scarce, they need to be prioritized.

Challenges to Good Assessment and Successful IIP Efforts

Making Causal Connections

Because of the many actions and voices affecting the information environment, it is often difficult to tell whether a certain behavioral change was actually *caused* by defense IIP efforts. Where effectiveness is paramount, causation does not matter, and correlation is sufficient; if the target audience does what you want, you may not care exactly why. However, for accountability purposes, causation does matter. Being able to claim that a certain program or capability caused a certain effect or outcome increases the likelihood that the capability will continue to be valued (and funded).

While attributing causation in the information environment can be challenging, it is never impossible. If assessments need to demonstrate causal connections, thoughtful assessment design at the outset of the process can allow them to do so.

Building a Shared Understanding of DoD IIP Efforts

Our interviews with congressional staffers revealed a challenge that is inherent to IIP efforts relative to conventional kinetic military capabilities: a lack of shared understanding about, or intuition for, what IIP capabilities do and how they actually work (including a limited understanding of the psychology of influence).

Military personnel and congressional staffers have good intuition when it comes to the combined-arms contributions of different military platforms and formations. They also have a shared understanding of the force-projection capabilities of a bomber wing, for example, or a destroyer, an artillery battery, or a battalion of infantry.

However, this shared understanding does not extend to most IRCs. Intuition (whether correct or not) has a profound impact on assessment and expectations for assessment. Where shared understanding is strong, heuristics and mental shortcuts allow much to be taken for granted or assumed away; where there is a lack of shared understanding about capabilities, everything has to be spelled out, because the assumptions are not already agreed upon.

Where shared understanding is lacking, assessment design must be more thoughtful. The dots must be connected, with documentation to policymakers and other stakeholders explicitly spelling out what might be assumed away in other contexts. Greater detail and granularity become necessary, as do deliberate efforts to build shared understanding. Despite the potential burden of the demand to provide congressional stakeholders with more information about IIP efforts and capabilities to support their decisionmaking and fulfill oversight requirements, there are significant potential benefits for future IIP efforts. Greater shared understanding can not only potentially improve advocacy for these efforts but also strengthen the efforts themselves by encouraging more-rigorous assessments.

Confronting Constraints, Barriers, Disruptors, and Unintended Consequences

If potential barriers to successful execution or disruptors of the intended logical sequence of an effort are considered as part of the planning process, they can also be included in the measurement and data collection plan. Collecting information in a way that takes into account potential points of failure can both facilitate adjustments to the effort and help ensure that assessment captures the effort's progress as accurately as possible. If the effort is found to be unsuccessful, it may be that there was not, in fact, a problem with the objectives or the underlying theory but that the effort has just been temporarily derailed by outside circumstances.

In a complex environment, IIP efforts face obstacles that can also challenge good assessment practices. For this reason, it is particularly important for DoD IIP assess-

ment to incorporate the principles of good assessment articulated earlier and to ensure that an effort can adapt to changes in context.

Learning from Failure

DoD requires IIP assessment for accountability purposes, of course, but it also depends on assessment to support a host of critical planning, funding, and process requirements. Consequently, it is vitally important to determine as early as possible whether certain activities are failing or have failed, so they can be corrected or abandoned. The unique challenge facing IIP planners is that they must do so without suggesting that IO overall is a failure.

Assessment can directly support learning from failure, midcourse correction, and planning improvements.[2] In military circles, there is a tendency to be overoptimistic about the likely success of an effort and a reluctance to abandon pursuits that are not achieving desired results. For this reason, we address failure—strategies to prevent it and strategies to learn from it—throughout this report.

After-action review is a familiar and widely used form of evaluation that is dedicated to learning from both success and failure. It has a major shortcoming, however: It is retrospective and timed in a way that makes it difficult for campaigns that are going to fail to do so quickly. The principles of good assessment articulated earlier can help prevent program failure, but they can also detect imminent failure early on, saving precious time and resources.

Topics Addressed and Key Insights

Identifying Best Practices and Methods for Assessment

In Chapter One, we begin with a brief overview of current DoD assessment practices and guidance on assessment, along with the framework for fitting best practices for assessment, drawn from a range of sectors, to the DoD IIP context—specifically via operational design and the joint operation planning process—and this is a theme we revisit throughout this report. The chapter also introduces our research objectives and approach and reveals that the best analogy for DoD IIP efforts is best practice in public communication (including social marketing). The best work in public communication leverages the best insights from the academic evaluation research and industry but moves away from the profit-based metrics that frequently appear in business marketing (and are poor analogs for DoD). The chapter concludes by explaining how DoD

---

[2]   These three aims were emphasized, respectively, in an interview with Mary Elizabeth Germaine, March 2013; Marla C. Haims, Melinda Moore, Harold D. Green, and Cynthia Clapp-Wincek, *Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects*, Santa Monica, Calif.: RAND Corporation, TR-784-OSD, 2011, p. 2; and author interview with LTC Scott Nelson, October 10, 2013.

IIP efforts can learn from both success and failure, another key point that we revisit throughout the report.

Why Evaluate? An Overview of Assessment and Its Utility

Chapter Two explores the motives for assessment and evaluation, beginning with the simple question, "Why evaluate?" Myriad reasons for assessment connect to three core motives: to support planning, improve effectiveness and efficiency, and enforce accountability. These three motives correspond roughly to the three types, or stages, of evaluation: formative, process, and summative. One key insight is that assessment should always support decisionmaking, and assessment that does not is suspect. Furthermore, our research suggests that DoD requires IIP assessment to support planning, improvement, and accountability, and we explore some of the unique challenges facing IIP efforts when it comes to meeting these requirements.

Applying Assessment and Evaluation Principles to IIP Efforts

Chapter Three offers a comprehensive overview of the IIP assessment best practices drawn from all the sectors reviewed (and presented at the beginning of this summary). We also describe how objectives can be nested, or broken into several subordinate, intermediate, or incremental steps. This approach facilitates assessment, particularly in the case of long-term effort, which may not produce results within the time frame demanded by stakeholders.

Challenges to Organizing for Assessment and Ways to Overcome Them

Chapter Four addresses the important matter of how to organize for assessment. The research shows that organizations that conduct assessment well usually have an organizational culture that values assessment, as well as leadership that is willing to learn from (and make changes based on) assessment. Here, we reiterate the point that assessment requires resources; experts suggest that roughly 5 percent of total program resources should be dedicated to evaluation. A culture of assessment can facilitate the success of IIP efforts and the implementation of the processes described in subsequent chapters.

Determining What's Worth Measuring: Objectives, Theories of Change, and Logic Models

Chapter Five revisits the principles of good assessment presented in Chapter Two and the assessment approaches described in Chapter Three as a way to identify the desirable properties of objectives and theories of change. Good objectives are SMART: specific, measurable, achievable, relevant, and time-bound. Good IIP objectives specify both the target audience and desired behaviors. Theories of change allow planners and assessors to express assumptions as hypotheses, identify possible disruptors that can interfere with the generation of desired effects, and, most important, determine where an effort is going awry if it is not achieving its objectives (and provide guidance on how

to fix it). A fully explicit theory of change is particularly important in IIP assessment because—unlike kinetic operations—IIP efforts lack commonly held (and validated) assumptions.

From Logic Models to Measures: Developing Measures for IIP Efforts

In Chapter Six, we address the processes and principles that govern the development of valid, reliable, feasible, and useful measures that can be used to assess the effectiveness of IIP activities and campaigns. We review two general processes: deciding which constructs are essential to measure and operationally defining the measures. Good measures should consider as many of the confounding and environmental factors that shape the outcome of interest as possible. Feasibility and utility can be in tension, however: Something may be easy to measure, but that does not mean it is useful to measure.

Assessment Design and Stages of Evaluation

Chapter Seven addresses the design of evaluation and assessment, specifying criteria to help select the appropriate design. The single most important property of assessment design is that it specifies the way in which the results will (or will not) enable causal inference regarding the outputs, outcomes, or impacts of the effort. The best designs are valid, generalizable, practical, and useful. However, there are tensions and trade-offs inherent in pursuing each of those objectives. Rigor and resources are the two conflicting forces in designing assessment. These two forces must be balanced with utility, but assessment design must always be tailored to the needs of stakeholders and end users.

Formative and Qualitative Methods for IIP Efforts

Chapter Eight reviews formative evaluation and qualitative data collection methods. Input from the SMEs interviewed for this study strongly suggests that DoD should invest more in qualitative and quantitative formative research to improve understanding of the mechanisms by which IIP activities achieve behavioral change and other desired outcomes. Initial investment in this area would pay off in the long run by reducing the chances of failure, identifying cost inefficiencies, and decreasing the resource requirements for summative evaluation.

Research Methods and Data Sources for Evaluating IIP Outputs, Outcomes, and Impacts

Chapter Nine describes methods and data sources for assessing outputs, outcomes, and impacts—those specific to IIP efforts and those related to process and summative evaluation. Even the most complicated analytical tools cannot overcome bad data. Furthermore, contrary to prevailing wisdom, *good data* is not synonymous with *quantita-*

tive data. Whether qualitative or quantitative, data should be validated using data from other collection methods whenever possible.

Surveys and Sampling in IIP Assessments: Best Practices and Challenges

Chapter Ten reviews the role of surveys and sampling in IIP assessment. Despite known limitations, surveys are likely to remain one of the most prominent and promising tools in this area. Survey sample size and sampling methods must be carefully considered and matched to both the target audience and analytic requirements. The chapter describes a litany of potential challenges and offers useful advice for addressing them.

Presenting and Using Assessments

Chapter Eleven addresses the presentation of assessments and ways to maximize their utility and ability to support decisionmaking. The main insight is that it is important to tailor the presentation of assessment results to the needs of stakeholders. Presentation must strike the right balance between offering detailed data and analyses (so that results are convincing) and supporting stakeholder decisions in a way that avoids overwhelming stakeholders with data. Some of the most effective presentations mix quantitative and qualitative data, allowing the qualitative data to provide context and nuance. Summary narratives can be an effective way to synthesize and aggregate information across programs, efforts, and activities to inform efforts at the operational or campaign level.

Technical Appendixes

This report is supported by four appendixes that offer readers much more detail on a selection of key topics. Appendix A includes a metaevaluation checklist for summative evaluations or for summative evaluations with a process evaluation component. The checklist addresses SMART objectives, theories of change, measurement, and so on to allow IIP assessment practitioners to test their assessment designs. Appendix B supplements the discussion of surveys and sampling in Chapter Ten with a review of sampling models and survey management, oversight, collaboration, and transparency. Appendix C highlights key examples and resources to guide the assessment of DoD IIP efforts, drawn from all the sectors addressed in this research. Finally, Appendix D briefly reviews several major theories of influence or persuasion, again drawn from the range of sectors that informed this research.

Recommendations

This report contains insights that are particularly useful for those charged with planning and conducting assessment, but there is also an abundance of information that is relevant to other stakeholders, including those who make decisions based on assess-

ments and those responsible for setting priorities and allocating resources for assessment and evaluation. Because assessment design, data collection, and the analysis and presentation of assessment results are all driven by the intended uses and users of the information produced, our recommendations are organized by stakeholder audience:

- DoD IIP assessment practitioners
- the broader DoD IIP community
- those responsible for congressional oversight
- those who manage DoD IO assessment reporting to Congress.

Although the recommendations presented here are targeted toward specific types of stakeholders, a recurring theme in our discussions of assessment challenges and practice improvement is the need for shared understanding across stakeholder groups. Therefore, points drawn from the experiences of one particular group are likely to prove informative for the others.

Recommendations for DoD IIP Assessment Practitioners

Our recommendations for assessment practitioners echo some of the most important practical insights described in the conclusions:

- Practitioners should *demand specific, measurable, achievable, relevant, and time-bound (SMART) objectives*. Where program and activity managers cannot provide assessable objectives, assessment practitioners should infer or create their own.
- Practitioners should *be explicit about theories of change*. A theory of change or logic of the effort ideally comes from commanders or program designers, but if theories of change are not made explicit, assessment practitioners should elicit or develop them in support of assessment.
- Practitioners should be provided with *resources for assessment*. Assessment is not free, and if its benefits are to be realized, it must be resourced.
- Practitioners must take care to *match the design, rigor, and presentation of assessment results to the intended uses and users*. Assessment supports decisionmaking, and providing the best decision support possible should remain at the forefront of practitioners' minds.

An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*, focuses more specifically on these and other recommendations for practitioners.[3]

---

[3]   Paul, Yeats, Clarke, Matthews, and Skrabala, 2015.

Recommendations for the Broader DoD IIP Community

Our recommendations for the broader DoD IIP community (by which we mean the stakeholders, proponents, and capability managers for IO, public affairs, military information support operations, and all other IRCs) emphasize how advocacy and a few specific practices will improve the quality of assessment across the community, but such efforts cannot be accomplished by assessment practitioners alone.

- DoD leadership needs to *provide greater advocacy, better doctrine and training, and improved access to expertise (in both influence and assessment)* for DoD IIP efforts. Assessment is important for both accountability and improvement, and it needs to be treated as such.
- DoD doctrine needs to *establish common assessment standards*. There is a large range of possible approaches to assessment, with a similarly large range of possible assessment rigor and quality. The routine and standardized employment of something like the assessment metaevaluation checklist in this report (described in Chapter Eleven and presented in Appendix A) would help ensure that all assessments meet a target minimum threshold.
- DoD leadership and guidance need to recognize that *not every assessment must be conducted to the highest standard*. Sometimes, good enough really is good enough, and significant assessment expenditures cannot be justified for some efforts, either because of the low overall cost of the effort or because of its relatively modest goals.
- DoD should *conduct more formative research*. IIP efforts and programs will be made better, and assessment will be made easier. Specifically,
  - Conduct target-audience analysis with greater frequency and intensity, and improve capabilities in this area.
  - Conduct more pilot testing, more small-scale experiments, and more early efforts to validate a specific theory of change in a new cultural context.
  - Try different things on small scales to learn from them (i.e., fail fast).
- DoD leaders need to *explicitly incorporate assessment into orders*. If assessment is in the operation order—or maybe in the execute order or even a fragmentary order—then it is clearly a requirement and will be more likely to occur, with requests for resources or assistance less likely to be resisted.
- DoD leaders should support the *development of a clearinghouse of validated (and rejected) IIP measures*. When it comes to assessment, the devil is in the details. Even when assessment principles are adhered to, some measures just do not work out, either because they prove hard to collect or because they end up being poor proxies for the construct of interest. Assessment practitioners should not have to develop measures in a vacuum. A clearinghouse of measures tried (with both success and failures) would be an extremely useful resource.

Recommendations for Congressional Overseers

To date, iterations of IO reporting to Congress have not been wholly satisfactory to either side (members of Congress and their staffers or DoD representatives). To foster continued improvement in this area, we offer recommendations for both, beginning with recommendations for congressional overseers.

- Congressional stakeholders should *continue to demand accountability in assessment*. It is important for DoD to conduct assessments of IIP efforts so that those that are not effective can be improved or eliminated and so that scarce resources are allocated to the most important and effective efforts.
- Congressional demands for accountability in assessment must be *clearer about what is required and expected*.
- When refining requirements, DoD representatives must *balance expectations*. Assessment in this area is certainly possible and should be conducted, but assessment should not be expected to fill in for a lack of shared understanding about the psychosocial processes of influence. (Understanding is much more fully shared for kinetic capabilities, such as naval vessels or infantry formations, making accountability for those capabilities much more straightforward.)

Recommendations for Those Who Manage DoD Reporting to Congress

To those who manage congressional reporting on the DoD side, we make the following recommendations.

- DoD reporting should *strive to meet the congressional desire for standardization, transition from output- to outcome-focused assessments, and retrospective comparison of what has and has not worked*. While these improvements are not trivial or simple, they are possible, and they are part of the congressional requirement that has been made clear.
- DoD reporting *must acknowledge that congressional calls for accountability follow two lines of inquiry and must show how assessment meets them both*. Congress wants to see justification for spending and evidence of the efficacy (traditional accountability), but it also wants proof that IIP activities are appropriate military undertakings. IIP efforts that can be shown (not just claimed) to be contributing to approved military objectives will go a long way toward satisfying both lines of inquiry.

Acknowledgments

This report owes a great debt to the many experts and colleagues whose good ideas and insightful observations are captured herein. We are particularly grateful for the advice and support of personnel in the sponsors' offices: COL Dan Ermer and Paula Trimble in the Rapid Reaction Technology Office in the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics; LTC Albert Armonda in the Office of the Under Secretary of Defense for Policy, Information Operations Directorate (OUSD[P] IO); and Austin Branch, former director of OUSD(P) IO. Though not a formal sponsor, the Joint Information Operations Warfare Center has been an engaged stakeholder throughout the process. We thank the following personnel there, who provided thoughts and input during the project: Rick Grimes, Charlie Chenoweth, Rick Josten, and Tom Lorenzen.

We also wish to acknowledge the subject-matter experts from the fields considered in this research. Under the terms of our interviews, insights and contributions from many of these experts, particularly in the defense sector, are reported on a not-for-attribution basis. We particularly thank these anonymous contributors: You know who you are, and your contribution is valued!

Because we sought to give credit where credit is due, as many of our interviews as possible were conducted on the record and for attribution. It is with deepest gratitude that we thank the following individuals who were able to share with us in this way, in alphabetical order: Susan Abbott, Joie Acosta, Sean Aday, Rebecca Andersen, Amelia Arsenault, Tarek Azzam, Elizabeth Ballard, Tom Beall, Paul Bell, Johanna Blakely, Steve Booth-Butterfield, Katherine Brown, Vincent Bruzzese, Julia Coffman, Charlotte Cole, Becky Collins, Steve Corman, John Croll, Nick Cull, Brian Cullin, Heidi D'Agostino, James Deane, Emmanuel de Dinechin, Kavita Abraham Dowsing, Kim Andrew Elliot, Marcus Gault, Mary Elizabeth Germaine, Joshua Gryniewicz, Jenn Gusikoff, Simon Haselock, Craig Hayden, Mark Helmke, Angela Jeffrey, Pamela Jull, Beau Kilmer, Sheri Klein, Marie-Louise Mares, Steven Martino, Lisa Meredith, David Michaelson, Devra Moehler, Chris Nelson, LTC Scott Nelson, James Pamment, Andrew Parker, Geeta Patel, Marc Patry, Julianne Paunescu, Gerry Power, Anthony Pratkanis, Monroe Price, Ronald Rice, David Rockland, Victoria Romero, Jon Schroden, Chris Scully, Phil Seib, William Shadel, Amanda Snyder, Amy Stolnis,

Abbreviations

ACSOR	Afghan Center for Socio-Economic and Opinion Research
ANOVA	analysis of variance
ANQAR	Afghan Nationwide Quarterly Assessment Research
BBC	British Broadcasting Corporation
BBG	Broadcasting Board of Governors
CARVER	criticality, accessibility, recuperability, vulnerability, effect, and recognizability
COA	course of action
DoD	U.S. Department of Defense
ECB	evaluation capacity building
EPSEM	equal probability of selection method
FM	field manual
GTO	Getting To Outcomes
HQ	headquarters
HUMINT	human intelligence
IE	information environment
IIA	inform and influence activities
IIP	inform, influence, and persuade
IO	information operations
IOTF	Information Operations Task Force
IRC	information-related capability

ISAF	International Security Assistance Force
ISR	intelligence, surveillance, and reconnaissance
JALLC	Joint Analysis and Lessons Learned Centre
JOPP	joint operation planning process
JP	joint publication
KAP	knowledge, attitudes, and practices
KPI	key performance indicator
M&E	monitoring and evaluation
MINDSPACE	messenger, incentives, norms, defaults, salience, priming, affect, commitment, and ego
MISO	military information support operations
MOE	measure of effectiveness
MOP	measure of performance
NATO	North Atlantic Treaty Organization
NGO	nongovernmental organization
OIF	Operation Iraqi Freedom
ORSA	operations research and systems analysis
PSYOP	psychological operations
ROI	return on investment
SEM	structural equation modeling
SIGACT	significant activity
SMART	specific, measurable, achievable, relevant, and time-bound
SME	subject-matter expert
TAA	target-audience analysis
USAID	U.S. Agency for International Development
USNORTHCOM	U.S. Northern Command

Identifying Best Practices and Methods for Assessment

Achieving key U.S. national security objectives demands that the U.S. government and the U.S. Department of Defense (DoD) effectively and credibly communicate with and influence a broad range of foreign audiences. To meet this objective, it is important to measure the performance and effectiveness of activities aimed at informing, influencing, and persuading. Thorough and accurate assessments of these efforts guide their refinement, ensure that finite resources are allocated efficiently, and inform accurate reporting of progress toward DoD's goals. Such efforts represent a significant investment for the U.S. government: DoD spends more than $250 million per year on information operations (IO) and information-related capabilities (IRCs) for influence efforts at the strategic and operational levels. How effective are those efforts? Are they well executed? How well do they support military objectives? Are they efficient (cost-effective)? Are some efforts better than others in terms of execution, effectiveness, or efficiency? Could some of them be improved? If so, how?

Unfortunately, generating assessments of such activities has been a challenge across the government and DoD. Inform, influence, and persuade (IIP) efforts often target the human cognitive dimension, attempting to effect changes in attitudes and opinions. These changes can be quite difficult to observe or measure accurately.

Even when activities seek to influence *behavior* (more easily observable and thus more measurable), causal conflation is a constant challenge. Did the influence activity generate this behavior, or is it a product of other exogenous factors? For example, many Iraqi soldiers surrendered at the outset of Operation Iraqi Freedom (OIF). Was that because of psychological operations (PSYOP) leaflets, demonstrations of coalition military might, dissatisfaction with the Saddam Hussein regime, some combination thereof, or something else entirely? Causal conflation is often compounded by the lengthy timelines of IIP activities; if a program seeks to change attitudes among a selected subpopulation over the course of a year, how can program personnel tell whether they are making good progress after three months, and how can they be certain that observed changes are due to their efforts rather than other influences in the information environment (IE)? Even where satisfactory assessment is conducted at the program or activity level, it remains a challenge to meaningfully aggregate different

forms and types of assessments to compare the relative merits of activities or to create a composite picture at the campaign level.

While these are difficult challenges in any domain, both the marketing sector and academic evaluation researchers have a long history of grappling with such issues and have achieved many successes that could provide useful insights in the defense context. U.S. businesses spend more than $30 billion annually on advertising and public relations activities to promote products and enhance corporate reputations. Corporate executives must justify these large sums to shareholders, so significant resources are dedicated to measuring both the execution (e.g., measures of performance) and effects (e.g., measures of effectiveness [MOEs]) of advertising and public relations initiatives. Closely related to corporate communication assessment is the academic discipline of evaluation research. Evaluation research employs a wide range of research methods to assess various programs and initiatives—among them, thoughtful frameworks for matching appropriate types of assessment with decisional needs and multivariate statistical techniques that can help disaggregate seemingly confounding sets of variables.

Current DoD Assessment Practice

Across DoD, assessment and evaluation vary widely in practice, not just for IIP efforts but also for a wide range of military undertakings. Pockets of strong practice exist, and we have sought to learn from those instances where possible.

A common misperception about assessment within DoD is that it is something pursued after the fact and that the primary uses of assessment results are after-action reporting and periodic funding justification. But as we discuss later, accountability is just one of the possible uses of assessment. As those who conduct assessments know, gauging progress or determining the impact of an effort post hoc is difficult and unrewarding if assessment was not included in plans at the outset. Including assessment as part of initial plans would have ensured that an effort was structured in a way that was amenable to assessment and that needed data could be collected over time. We explore these and other principles of effective assessment in Chapter Three.

A point that should not be overlooked in the planning, conduct, and assessment of activities that fall under the umbrella of IO is the relationship between these activities and kinetic operations and the unique challenges that stem directly from tensions between them. Chapter Two touches on challenges related to a lack of shared understanding of the goals, utility, timeline, and impact of IIP efforts; Chapter Four shows how kinetic and IIP efforts follow similar planning and decisionmaking paths and how they can work in concert in support of broader campaign goals. DoD has taken steps in recent years to acknowledge and leverage the roles of kinetic and information operations, both individually and collectively, in the joint environment.

Current DoD Assessment Guidance

As of this writing, there have been numerous developments that have raised the profile of assessment within DoD. Few of these initiatives are specific to the assessment of IIP activities, but all represent an attempt to encourage better assessment practice in DoD and to provide the needed foundation and guidance for doing so. The following are among the efforts currently under way; when complete, they should be of use to future users of this report:

- the development of an Air Land Sea Application Center manual of multiservice tactics, techniques, and procedures for assessment
- a planned joint doctrine note on operations assessments
- the Joint Test and Evaluation Program's *Joint Assessments Doctrine Evaluation Quick Reaction Test*, which will support the two efforts above and provide additional rigor to the integration of assessment guidance in future editions of Joint Publication (JP) 3-0, *Joint Operations*, and JP 5-0, *Joint Operation Planning*
- A new chapter on assessments in JP 3-13, *Information Operations*, to be incorporated into the planned update of that publication.

The remainder of this section briefly describes some of the existing doctrinal guidance relevant to the assessment of IO activities. Although they have been criticized for being overly vague, DoD doctrinal publications describe and provide definitions of critical components of operational assessments.[1] They offer helpful background on the reasons for assessment and encourage something of a common vocabulary for assessment that can be particularly useful in joint efforts or in aggregating individual efforts in support of broader campaigns, points discussed in greater detail in Chapter Two. There is room for improvement, but even in their current format, they provide some useful insights. For example, a fundamental contribution that the publications discussed here have made to the practice of good assessment is their emphasis on continuous evaluation throughout a given effort.

Field Manual 3-53: Military Information Support Operations

Field Manual (FM) 3-53 provides guidance for U.S. Army military information support operations (MISO) activities.[2] Part of this guidance focuses on assessment, which is considered one of the core components of a MISO program.[3] Specifically, plans for a MISO program should identify target audiences for these operations, key themes

---

[1] Jonathan Schroden, "Why Operations Assessments Fail: It's Not Just the Metrics," *Naval War College Review*, Vol. 64, No. 4, Fall 2011.

[2] Formerly known in doctrine as PSYOP.

[3] Headquarters, U.S. Department of the Army, *Military Information Support Operations*, Field Manual 3-53, Washington, D.C., January 2013b.

to promote and avoid, channels for dissemination, concepts that outline operational goals, paths to achieving the goals, and appropriate assessment approaches.

As described in FM 3-53, assessment is "the continuous monitoring and evaluation of the current situation, particularly the enemy, and the progress of an operation." Continuous assessment involves MISO planners working with commanders to determine operational goals and establish informative and useful MOEs. This communication and the overall process are informed by current knowledge of target audiences, adversary influence on these audiences, and past and current data collection efforts.

Field Manual 3-13: Inform and Influence Activities

MISO serves as just one line of support for inform and influence activities (IIA).[4] Where FM 3-53 focuses on MISO organization and implementation, FM 3-13 specifically focuses on IIA. Although FM 3-53 and FM 3-13 describe overlapping aspects of assessments, FM 3-13 provides more-detailed guidance on the assessment of IIA, including methodologies for selecting high-value entities on which to focus efforts (i.e., targeting).

Joint Publication 5-0: Joint Operation Planning

JP 5-0 provides joint-level guidance regarding assessment, describing it as "the continuous monitoring and evaluation of the current situation and progress of a joint operation toward mission accomplishment."[5] As with the Army's field manuals described here, it addresses the necessity of ongoing assessment. It also emphasizes the use of assessment to determine current operational effectiveness in comparison with planned operational goals—a comparison that should inform subsequent adjustments to operations.

Integrating Best Practices into Future DoD IIP Assessment Efforts: Operational Design and the Joint Operation Planning Process as Touchstones

The third in our list of doctrinal publications addressing assessment, JP 5-0, addresses both operational design and the joint operation planning process (JOPP). While both are clearly aimed at a command staff during advance planning, they are sufficiently flexible to support a wide range of planning processes. Because JP 5-0 guidance is so broadly applicable and widely familiar to DoD personnel, we use *operational design* and *JOPP* throughout this report as touchstones to illustrate how and where the various

---

[4]   Headquarters, U.S. Department of the Army, *Inform and Influence Activities*, FM 3-13, Washington, D.C., January 2013a. The manual uses *inform and influence activities* to refer to a particular component of IO, and those activities would fall under our general definition of *IIP*.

[5]   U.S. Joint Chiefs of Staff, *Joint Operation Planning*, Joint Publication 5-0, Washington, D.C., August 11, 2011a.

assessment practices we recommend can be integrated into existing military processes. For those unfamiliar with operational design and JOPP, we briefly review both here.

Operational Design

As described in JP 5-0, operational art is about describing the military end state that must be achieved (ends), the sequence of actions that are likely to lead to those objectives (ways), and the resources required (means). This specification of *ends*, *ways*, and *means* sounds very much like the articulation of a theory of change (described in Chapters Three and Five).

Operational design is the part of operational art that combines an understanding of the current state of affairs, the military problem, and the desired end state to develop the operational approach. These are the four steps in operational design:

1. understand the strategic direction
2. understand the operational environment
3. define the problem
4. use the results of steps 1–3 to develop a solution—i.e., the operational approach.

Joint Operation Planning Process

Operational design and JOPP are related in that operational design provides an iterative process that can be applied within the confines of JOPP. JOPP formally has seven steps:

1. planning initiation
2. mission analysis
3. course-of-action (COA) development
4. COA analysis and war-gaming
5. COA comparison
6. COA approval
7. plan or order development.

For practical purposes, mission analysis should be disaggregated so that it begins with a subprocess related to operational art—problem framing and visualization—and incorporates a full iteration of operational design. In our discussion of JOPP, we treat those two subprocesses as part of step 2, mission analysis. Those who would like further detail on either organizational design or JOPP are referred to JP 5-0.

What RAND Was Asked to Do

This project's sponsors in the Office of the Secretary of Defense asked RAND to identify effective principles and best practices for the assessment of IIP efforts from

across sectors and distill them for future application in DoD. As part of this effort, the RAND project team was asked to review existing DoD IIP assessment practices (and broader DoD assessment practices), identify IIP assessment practices in industry (commercial marketing, public relations, and public communication), and review guidance and practices from the academic evaluation research community. Specific project tasks included a review of existing approaches to assessment, identifying relevant state-of-the-art practices, and synthesizing what was discovered for application to DoD IIP assessment.

Methods and Approach

To complete these tasks and provide DoD with a structured set of insights, principles, and practices applicable to the assessment and evaluation of IIP efforts, we conducted a comprehensive literature review and more than 100 interviews with subject-matter experts (SMEs) who held a range of roles in government, industry, and academia. The literature reviewed was copious and wide-ranging, encompassing hundreds of documents; we compiled the most informative and useful of those resources into an annotated bibliography and reading list, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*.[6] Interviews and documents are cited throughout this report as well. Many of our SME interviews were conducted on a for-attribution basis, so we are able to provide direct quotes and give credit where credit is due for good ideas.

Once we compiled the practices, principles, advice, guidance, and recommendations, we distilled and synthesized all the material for application to DoD. This portion of the effort was at least as much art as science. We grouped observations and insights topically, identifying substantive areas for discussion, with a corresponding chapter devoted to each. Practices or principles emphasized across all (or many) sectors were prioritized by virtue of that consensus. Where certain practices appeared in only one sector, we considered their applicability to defense IIP contexts and used our judgment as to whether or not they should be included here. Where practices appeared to conflict or disagree within or across sectors, we present both sides of the debate, list possible pros and cons, or, through the application of logic and our understanding of the defense IIP context, offer only the most applicable advice.

To further extend the utility of the findings and best practices presented here, we have developed a companion handbook for practitioners of IIP assessment. That volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and*

---

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

Persuade: Handbook for Practitioners, distills key points and practices in a user-friendly quick-reference format.[7]

Different Sectors Considered

Originally, we sought insights from three broadly defined sectors: industry, academia, and government. As we explored and gained experience with these sectors, we found that a number of more nuanced characterizations and descriptions were appropriate. Table 1.1 lists the sectors that best capture the breadth of our sources, provides a description of the sector, and indicates the number of SMEs interviewed in that sector.

Table 1.1
Number of Interviews Conducted, by Sector

Sector	Description	SMEs Interviewed
Industry		
Marketing/ public relations	Professionals in the marketing, advertising, or public relations fields in the for-profit sector	18
Public communication	Practitioners in public communication (including social marketing) or public communication evaluation in the nonprofit sector	26
Academia		
Evaluation research	Academics specializing in evaluation research (not necessarily IIP)	10
IIP evaluation	Academics specializing in influence or persuasion, with relevant expertise in IIP measurement, assessment, or evaluation	22
Media evaluation	Academics specializing in media evaluation	11
Defense		
Practitioners	Uniformed military, civilian, or contractor personnel with experience conducting or assessing defense IIP efforts	33
Academics/ think tanks	Academics or scholars who have conducted research on IIP or IIP assessment in the defense context	8
Other government representatives		
Practitioners	Personnel from elsewhere in government (beyond DoD) with experience assessing government IIP efforts	8
Congressional staff	Former or current congressional staff interviewed for stakeholder perspectives	5

---

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

The Most-Informative Results for DoD IIP Efforts Were at the Intersection of Academic Evaluation Research and Public Communication

While usable and useful lessons came from all the sectors reviewed, the best insights came from the intersection of public communication (particularly social marketing) and academia. When we say *best*, we mean best in terms of applicability to defense IIP assessment, methodological rigor, and being novel to defense assessment. Public communication provided the best analogy for defense IIP. In the for-profit sector, many assessment efforts and measures connected to sales, earnings, return on investment (ROI), or something else that is explicitly monetized, which tends to break analogy with defense. In public communication, however, behavior or attitudinal change is sought (as in defense IIP efforts), and often from at-risk, hard-to-reach, or other challenging audiences (again, as is often the case in defense IIP). Where public communication has been conducted according to the best practices of evaluation research, it has achieved a very compelling combination of effective, thoughtful assessment and methodological rigor. This combination is rare in existing defense IIP assessment practice, but we believe that the core principles and best practices from top-quality assessment efforts in public communication provide an excellent template.

DoD IIP Efforts Can Learn from Both Success and Failure

DoD requires IIP assessment for accountability purposes, of course, but it also depends on assessment to support a host of critical planning, funding, and process requirements. Many IIP efforts involve uncertainty. When trying to influence a population to do something new and different in a new context, there are many unknowns that might slow, diminish, or disrupt an effort. Under such circumstances, one way to figure out what works and what does not is to try something and monitor the results. The guiding principle here should be to *fail fast*. If early and frequent assessment reveals that it is not working, you can adjust, correct, or try something else entirely.

Assessment can directly support learning from failure, midcourse correction, and planning improvements.[8] In military circles, there is a tendency to be overoptimistic about the likely success of an effort and a reluctance to abandon pursuits that are not achieving desired results. For this reason, we address failure—strategies to prevent it and strategies to learn from it—throughout this report. More to the point: Building an organizational culture that values assessment requires getting over the fear of the results.

After-action review is a familiar and widely used form of evaluation that is dedicated to learning from both success and failure. It has a major shortcoming, however: It

---

[8] These three aims were emphasized, respectively, in an author interview with Mary Elizabeth Germaine, March 2013; Marla C. Haims, Melinda Moore, Harold D. Green, and Cynthia Clapp-Wincek, *Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects*, Santa Monica, Calif.: RAND Corporation, TR-784-OSD, 2011, p. 2; and an author interview with LTC Scott Nelson, October 10, 2013.

is retrospective and timed in a way that makes it difficult for campaigns that are going to fail to do so quickly. In contrast, implicit in many examples of effective assessment and explicit in much of the work by scholars of evaluation is the importance of a *theory of change*.[9] Simply put, the theory of change or logic of the effort is a statement of how you believe the things you are planning on doing are going to lead to the objectives you seek. The main benefit of articulating the logic of the effort in the assessment context is that it allows assumptions of any kind to be turned into *hypotheses*. Assessment along an effort's chain of logic (testing the hypotheses) enables process improvement, makes it possible to test assumptions, and can tell evaluators why and how an unsuccessful effort is failing.

JP 5-0 describes operational design as an iterative process. Iteration should occur not just during initial planning but also during operations as assumptions and plans are forced to change in response to constraints, barriers, disruptors, and unintended consequences. Operational design also advocates continuous learning and adaptation, and well-structured assessment supports that process.

Outline of This Report

The remainder of this report is organized as follows. Chapter Two explores the motives for assessment and evaluation, beginning with the simple question, "Why evaluate?" Myriad reasons for assessment connect to three core motives: to support planning, improve effectiveness and efficiency, and enforce accountability. These three motives correspond roughly to the three types, or stages, of evaluation: formative, process, and summative. Chapter Three offers a comprehensive overview of the IIP assessment best practices drawn from all the sectors reviewed (and presented at the beginning of this summary). Chapter Four addresses the important matter of how to organize for assessment. The research shows that organizations that conduct assessment well usually have an organizational culture that values assessment, as well as leadership that is willing to learn from (and make changes based on) assessment. A culture of assessment can facilitate the success of IIP efforts and the implementation of the processes described in subsequent chapters.

Chapter Five revisits the principles of good assessment presented in Chapter Two and the assessment approaches described in Chapter Three as ways to identify the desirable properties of objectives and theories of change. Theories of change allow planners and assessors to express assumptions as hypotheses, identify possible disruptors that can interfere with the generation of desired effects, and, most important, determine

---

[9]   In presentations of early results, we noticed that some uniformed stakeholders were uncomfortable with the phrase *theory of change*, suggesting that *theory* sounds too theoretical, too abstract, and impractical. While used in the academic literature and throughout this report, where the phrase *theory of change* is at risk of alienating a group of stakeholders, we include an alternative term of art, *logic of the effort*.

where an effort is going awry if it is not achieving its objectives (and provide guidance on how to fix it). In Chapter Six, we address the processes and principles that govern the development of valid, reliable, feasible, and useful measures that can be used to assess the effectiveness of IIP activities and campaigns. Chapter Seven addresses the design of evaluation and assessment, specifying criteria with which to help select the appropriate design.

Turning to the topic of research and data sources to support assessment, Chapter Eight reviews formative evaluation and qualitative data collection methods. Chapter Nine describes methods and data sources for assessing outputs, outcomes, and impacts—those specific to IIP efforts and those related to process and summative evaluation. Chapter Ten reviews the role of surveys and sampling in IIP assessment. Despite known limitations, surveys are likely to remain one of the most prominent and promising tools in this area.

Chapter Eleven brings the discussion back to the overriding motivation for assessment introduced in Chapter Two: the uses and users of assessment results. It discusses the presentation of assessments and ways to maximize their utility and ability to support decisionmaking.

Chapter Twelve revisits some key insights offered throughout this report, synthesizing them and offering recommendations for DoD IIP assessment practitioners, the broader DoD IIP community, congressional overseers, and those who manage DoD reporting to Congress.

This report is supported by four appendixes: Appendix A includes a metaevaluation checklist for summative evaluations or for summative evaluations with a process evaluation component, intended to guide IIP assessment practitioners in testing their assessment designs. Appendix B supplements the discussion of surveys and sampling in Chapter Ten with a review of sampling models and survey management, oversight, collaboration, and transparency. Appendix C highlights key examples and resources to guide the assessment of DoD IIP efforts, drawn from all the sectors addressed in this research. Finally, Appendix D briefly reviews several major theories of influence or persuasion, again drawn from the range of sectors that informed this research.

Why Evaluate? An Overview of Assessment and Its Utility

This chapter lays a foundation for the discussion of assessment and evaluation that follows by describing the motives for assessment in different sectors. We begin by identifying the core reasons for assessment, as well as some arguably illegitimate motives for evaluation. We then address the specific arguments for improved assessment of DoD IIP efforts, clarifying both the requirement for assessment and its utility and benefits.

The Language of *Assessment*

One factor that varies across government, defense, industry, and academia is *how* assessment is discussed. Different sectors use different terms of art to describe things that are similar, if not entirely overlapping. In government and defense, the term of choice is *assessment*, while academic evaluation researchers (unsurprisingly) talk about *evaluation*. In commercial marketing, the conversation is usually about *metrics* or just *measurement*. Others have written about *monitoring*, and many of the people we interviewed used more than one of these terms, sometimes as synonyms and sometimes to denote slightly different things. As one of these SMEs noted, "There are as many different definitions of *assessment* as there are people doing it."[1]

Here, we use *assessment* and *evaluation* interchangeably and synonymously, with our choice of the two terms driven by the source of the discussion: When the sources we are citing discussed *evaluation*, we use *evaluation*, and vice versa. When in doubt, or when the same topic was discussed by experts in multiple fields using different terminology, we lean toward *assessment* because it is the preferred term of art in the defense community. Where we use other terms (such as *measurement*, *measures of effectiveness*, or *formative evaluation*), we do so intentionally and specifically, and we make clear what we mean by those terms.

---

[1]   Author interview on a not-for-attribution basis, December 5, 2012.

Three Motivations for Evaluation and Assessment: Planning, Improvement, and Accountability

Assessment or evaluation is fundamentally a judgment of merit against criteria or standards.[2] But for what purpose? To what end do we make these judgments of merit? This report draws on examples from government and military campaigns, industry (both commercial marketing and public communication), and academia, collected through more than 100 interviews and a rigorous literature review to inform its findings. Across these sectors, all motivations or proposals for assessment or evaluation aligned comfortably with one (or more) of three broad goals: to improve planning, to improve effectiveness and efficiency, and to enforce accountability.

Within these categories, assessment efforts have many—and more-specific—goals. The following is merely a sampling of the motivations for assessment that we encountered in the course of our study. To improve planning, assessment efforts sought to

- force the setting of objectives[3]
- plan for future programs[4]
- refine plans[5]
- assist in developing a new program[6]
- monitor assumptions[7]
- reveal best practices[8]
- generate knowledge.[9]

To improve effectiveness and efficiency, assessment efforts sought to

- determine how well a program worked (if it did)[10]

---

[2]   Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach*, 7th ed., Thousand Oaks, Calif.: Sage Publications, 2004.

[3]   Author interview with Thomas Valente, June 18, 2013.

[4]   Author interview with Thomas Valente, June 18, 2013.

[5]   North Atlantic Treaty Organization (NATO), *NATO Operations Assessment Handbook*, interim version 1.0, January 29, 2011. Not available to the general public.

[6]   Barbara Schneider and Nicole Cheslock, *Measuring Results: Gaining Insight on Behavior Change Strategies and Evaluation Methods from Environmental Education, Museum, Health, and Social Marketing Programs*, San Francisco, Calif.: Coevolution Institute, April 2003.

[7]   UK Ministry of Defence, *Assessment*, Joint Doctrine Note 2/12, Shriveham, UK, February 2012.

[8]   Robert Banks, *A Resource Guide to Public Diplomacy Evaluation*, Los Angeles, Calif.: Figueroa Press, November 2011.

[9]   Rossi, Lipsey, and Freeman, 2004.

[10]   Author interview with Thomas Valente, June 18, 2013.

- measure progress[11]
- support resource management decisions[12]
- monitor a current program[13]
- estimate the effects of a program on different populations[14]
- estimate the cost-effectiveness of a program[15]
- monitor implementation[16]
- inform the improved allocation of resources.[17]

Finally, for the purposes of enforcing accountability, assessment efforts sought to

- determine whether the program met its objectives[18]
- ensure that the program met federal accountability requirements[19]
- measure results[20]
- identify a better available program[21]
- justify budget requests.[22]

Assessment can service any or all of these goals and more.

Three Types of Evaluation: Formative, Process, and Summative

The three broad motivations for assessment (improve planning, improve effectiveness and efficiency, and support accountability) roughly correspond to three primary types of evaluation. These concepts are drawn from the academic literature, so we use the term *evaluation* in this discussion; however, the implication is the same regardless of context.

---

[11] NATO, 2011.

[12] NATO, 2011.

[13] Schneider and Cheslock, 2003.

[14] Schneider and Cheslock, 2003.

[15] Schneider and Cheslock, 2003.

[16] UK Ministry of Defence, 2012.

[17] Banks, 2011.

[18] Author interview with Thomas Valente, June 18, 2013.

[19] Author interview with Thomas Valente, June 18, 2013.

[20] NATO, 2011.

[21] Schneider and Cheslock, 2003.

[22] Banks, 2011.

The three types or stages of evaluation are formative evaluation, process evaluation, and summative evaluation:

- *Formative evaluation* occurs primarily during (or even prior to) the planning stage, prior to the execution of IIP activities, and includes efforts designed to develop and test messages, determine baseline values, analyze audience and network characteristics, and specify the logic by which program activities are designed to generate influence, including barriers to behavioral change.
- *Process evaluation* determines whether the program has been or is being implemented as designed, assesses output measures (such as reach and exposure), and provides feedback to program implementers to inform course adjustments.
- *Summative evaluation*, including outcome and impact evaluation, is the post-intervention analysis to determine whether the program achieved its desired outcomes or impact.

These types of evaluation can be characterized as *stages*, because they can be undertaken one after the other in an inherently linked way and can be conceptually integrated as part of a full range of evaluation activities over the duration of a program or campaign. Thomas Valente, a professor at the University of Southern California's Keck School of Medicine and a highly respected expert on evaluation methods and network analysis for health communication campaigns, has noted synergies between phases of campaigns, with good formative and process evaluation making summative evaluation easier.[23] Julia Coffman, director of the Center for Innovation in Evaluation, a senior consultant at the Harvard Family Research Project, and author of the 2002 study *Evaluating Public Communication Campaigns*, suggests timing data collection in evaluation so that one phase is continually informing the others.[24]

For example, imagine planning and conducting an IIP effort to promote democracy in a country by encouraging participation in national elections, not unlike efforts that have occurred in Iraq and Afghanistan as part of OIF and Operation Enduring Freedom. The formative stage could include a range of activities. One might begin by examining the records of programs that promote election participation in other countries or previous efforts in the current country. The formative stage is a good time to identify a baseline; in this case, voter turnout in previous elections would be a good baseline, supplemented by information about regional variation or variation by different demographic characteristics, if possible. If a baseline is not available (perhaps it is the first election under a new democratic scheme, or perhaps data were not recorded during previous elections), formative research could include preliminary surveys of intention to vote. Based on existing data or data collected as part of formative research,

---

[23] Author interview with Thomas Valente, June 18, 2013.

[24] Author interview with Julia Coffman, May 7, 2013.

you could identify groups least likely to participate and try to identify ways to increase their participation. Formative research could include focus groups with representatives from populations of interest to identify barriers to participation in elections. Draft election-promotion materials could be presented and tested in other focus groups, with feedback contributing to their revision. Formative research could include limited pilot testing of materials with real audiences, provided there is some mechanism in place to see how well they are working (such as observations, a small survey, or quick interviews after exposure to the materials).

With as much planning and preparation as possible informed by the formative research, the delivery of the effort (what would be called the *intervention* in the academic literature) can begin. At this point, process evaluation can also begin. Process evaluation includes the collection of measures of performance (MOPs), and might measure whether the planned amount of material has been printed and distributed or broadcast and whether it has been viewed.

An important part of process evaluation is making sure that the things that are supposed to happen are happening—and in the way envisioned. Are contractors delivering on their contracts? Are program personnel executing tasks, and are those tasks taking the amount of time and effort planned for them? Are audiences actually receiving materials as planned? Process evaluation is not just about recording these inputs, activities, and outputs; it is also about identifying problems in delivery, the reasons for those problems, and how they might be fixed. If, for example, a television commercial promoting election participation is being broadcast but no one reports seeing it, process evaluation turns back toward the methods of formative evaluation to find out why. Perhaps the commercial is airing on one channel in a time slot when the vast majority of the potential audience tunes in to a very popular program on a different channel. Note that while additional assessment activities begin when delivery begins, formative research need not stop. In this example, monitoring the early results of the election-promotion program's delivery may provide new information that informs adjustments to the plan in progress.

For election-participation promotion, the core of summative evaluation takes place at the end: Was voter turnout increased by the desired amount or not? There is more to it than that, however. Even getting the answer to that simple question requires earlier thought and planning. If there is no baseline against which to compare voter turnout (either from a previous election or through some kind of projection), then change in turnout cannot be calculated. If objectives did not specify the desired increase in turnout, an absolute value of turnout or change in turnout could be calculated but it would be difficult to know whether that is sufficient. Furthermore, those responsible for oversight of the effort might want to know how much of the change in turnout is attributable to the effort. This is a question about causation—often a particularly challenging one in the IIP context—and would also be part of summative evaluation. If such a question is to be answered in the summative phase, it has to be consid-

ered from the outset: Some form of quasi-experimental design would need to have been planned and executed, perhaps a design in which one or more areas were excluded from program delivery (either for a time or entirely), with differences in planned or actual voting behavior between areas that were exposed to the program and areas that were not (controlling for differences between the areas, perhaps statistically). This process would indicate the portion of the change in voter turnout that is due to the program.

Although the stages of evaluation seem sequential, being listed one after the other, they overlap and feed back into each other, and all require some planning from the outset to execute properly. The stages of evaluation, along with additional examples, are discussed in greater detail in Chapter Seven.

Nesting: The Hierarchy of Evaluation

The nested relationship among the three stages of evaluation offers a slightly different conceptual scheme for thinking about evaluation. "The hierarchy of evaluation" as developed by evaluation researchers Peter Rossi, Mark Lipsey, and Howard Freeman is presented in Figure 2.1.[25] The hierarchy divides potential evaluations and assessments into five nested levels. They are nested such that each higher level is predicated on success at a lower level. For example, positive results for cost-effectiveness (the highest level) are possible only if supported by positive results at all lower levels.

Figure 2.1
The Hierarchy of Evaluation
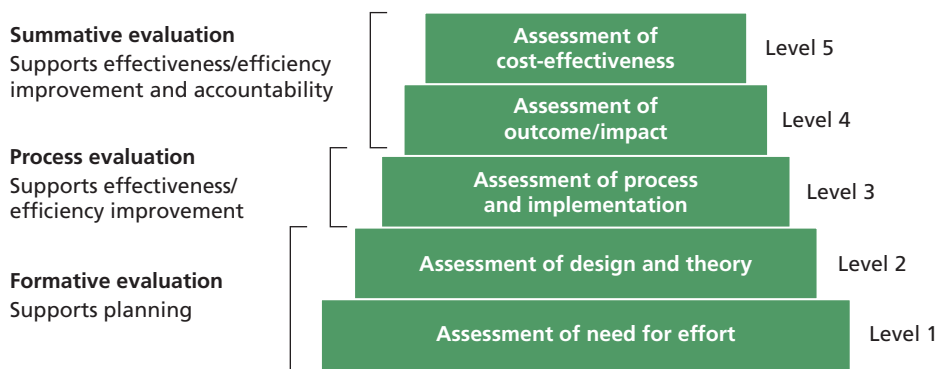
SOURCE: Adapted from Christopher Paul, Harry J. Thie, Elaine Reardon, Deanna Weber Prine, and Laurence Smallman, *Implementing and Evaluating an Innovative Approach to Simulation Training Acquisitions*, Santa Monica, Calif.: RAND Corporation, MG-442-OSD, 2006, Figure 7.1.
RAND *RR809/1-2.1*

---

[25] Rossi, Lipsey, and Freeman, 2004.

These five levels roughly correspond to the three motives and three stages of evaluation already described. Working from the bottom of the hierarchy, needs assessment and assessment of design and theory both support planning and are part of formative evaluation. Assessment of process and implementation directly corresponds to process evaluation and contributes to improving effectiveness and efficiency. Assessment of outcome/impact and assessment of cost-benefit effectiveness are part of summative evaluation and can be applied to efforts to improve efficiency and effectiveness and efforts to enforce accountability.

As noted earlier, this framework is described as a hierarchy because the levels nest with each other; solutions to problems observed at higher levels of assessment often lie at levels below. If the desired outcomes (level 4) are achieved at the desired levels of cost-effectiveness (level 5), then lower levels of evaluation are irrelevant. But what about when they are not?

When desired high-level outcomes are not achieved, information from the lower levels of evaluation needs to be available and examined. For example, if an effort is not realizing its target outcomes, is that because the process is not being executed as designed (level 3) or because the theory of change/assumed logic of the effort is incorrect (level 2)?[26] Evaluators encounter problems when an assessment scheme does not include evaluations at a sufficiently low level to inform effective policy decisions and diagnose problems. When the lowest levels of evaluation have been "assumed away," skipping lower-level evaluation steps is acceptable only if those assumptions prove correct. By then, it could prove exceptionally difficult and costly to revisit those levels.

Assessment to Support Decisionmaking

While assessment can have a range of uses and users and serve a number of different specific purposes, it should always support decisionmaking of some kind. This foundational view is represented—if not always emphasized—in the best practices across all the sectors we investigated. The *Commander's Handbook for Assessment Planning and Execution* clearly states, "The purpose of assessment is to support the commander's decisionmaking."[27]

The handbook, developed by the Joint Chiefs of Staff to fill the gap in the doctrinal guidance on planning and executing assessments, draws on "extensive lessons learned and best practices gained throughout the joint environment" and expands on the concepts articulated in the prevailing joint doctrine, *Joint Operations* (JP 3-0),

---

[26] This is a distinction between program failure and theory failure and is discussed in greater detail in Chapter Five in the section "Program Failure Versus Theory Failure."

[27] U.S. Joint Chiefs of Staff, *Commander's Handbook for Assessment Planning and Execution*, version 1.0, Suffolk, Va.: Joint and Coalition Warfighting, September 9, 2011c, p vii.

Joint Operation Planning (JP 5-0), and *Joint Intelligence* (JP 2-0).[28] It is designed to complement and connect assessment activities at all levels and integrate with formal military planning and decisionmaking processes, such as JP 5-0's JOPP or the attendant discussion of operational art and operational design. We discuss the relationship between JOPP and the broader concept of assessment in Chapter Three, and we periodically point out where in the operational planning or design process a noted assessment best practice is most salient. For example, formative evaluation can support decisionmaking as part of the planning process. Formative research could be foundational to the portion of operational design concerned with understanding the operational environment, and it could also play an important role in validating the assumptions necessary to propose solutions and develop an operational approach. JOPP is, itself, a form of formative research, with the activities of mission analysis, COA development, COA analysis and war-gaming, and COA comparison all fitting within the rubric of formative research—with the ultimate decision, COA approval, supported by those formative evaluations. Because assessment should support decisionmaking, it always has a potential role in operational design, planning, and execution.

According to Maureen Taylor, a professor and media evaluation specialist at the University of Oklahoma, getting assessment results into a form that is useful to the people who need them to make decisions is one of the biggest challenges of assessment. Supporting this view, evaluation researcher Charlotte Cole has noted that methodologically rigorous assessments that fail to inform the decisionmaker before a decision is made are simply useless.[29] At the intersection of academia and marketing, Douglas Hubbard has noted, "If a measurement matters at all, it is because it must have some conceivable effect on decisions and behavior."[30] While this principle seems obvious, it is not always adhered to. Rossi, Lipsey, and Freeman have found that, unfortunately, some sponsors commission evaluation research with little intention of using the results.[31] Poorly motivated assessments include those done simply for the purpose of saying that assessment has taken place, those done to justify decisions already made, and those done to satisfy curiosity without any connection to decisions of any kind.[32] For example, if the commander asks for assessment to justify his or her chosen COA *after* it has been selected rather than before (during COA development or during COA analysis and war-gaming), then it is not really an assessment.

---

[28] U.S. Joint Chiefs of Staff, *Joint Operations*, Joint Publication 3-0, Washington, D.C., August 11, 2011b; U.S. Joint Chiefs of Staff, 2011a; U.S. Joint Chiefs of Staff, *Joint Intelligence*, Joint Publication 2-0, Washington, D.C., October 22, 2013.

[29] Author interview with Maureen Taylor, April 4, 2013; interview with Charlotte Cole, May 29, 2013.

[30] Douglas W. Hubbard, *How to Measure Anything: Finding the Value of "Intangibles" in Business*, 2nd ed., Hoboken, N.J.: John Wiley and Sons, 2010, p. 47.

[31] Rossi, Lipsey, and Freeman, 2004.

[32] Author interviews on a not-for-attribution basis, October 30 and February 20, 2013.

Users of Evaluation

The three motives for assessment (improving planning, improving effectiveness and efficiency, and enforcing accountability) can be categorized even more narrowly. Assessments are primarily either up- and out-focused (accountability to an external stakeholder) or down- and in-focused (supporting planning or improvement internally). This categorization focuses on the users of the assessments.

If assessment is to support decisionmaking, it must be tailored in its design and presentation to its intended uses and users. Doing so involves clearly understanding both the assessment users (stakeholders, other assessment audiences) and how assessment results will be used (the purposes served and the specific decisions to be supported). Monroe Price, the director of the Center for Global Communication Studies at the University of Pennsylvania's Annenberg School for Communication, has said that the core question governing evaluation design is *who* and *what* decisions the evaluation is informing. Field commanders, for example, will have a different set of questions than congressional leaders.[33]

The context of uses and users should be considered as part of evaluation design and considered again when presenting evaluation results. Chapter Seven discusses uses and users in greater detail (including instructions for building a uses-users matrix); Chapter Eleven expands on ways to match the presentation of assessment results to user needs.

Requirements for the Assessment of DoD Efforts to Inform, Influence, and Persuade

This discussion about uses and users of assessment connects nicely to the central topic of this report, the DoD requirement to assess IIP efforts. There is considerable pressure on and within DoD for improved assessment in this area.

The main driver of the evolving assessment requirement is congressional scrutiny; several of the annual National Defense Authorization Acts of the past few years have included language specifying reports or reporting requirements having to do with the assessment of IO. The result has been a flurry of activity within DoD to meet congressional demands. DoD conducted several internal studies (the 2009–2010 Joint IO Force Optimization Study and the 2010 Secretary of Defense's Front-End Assessment of Strategic Communication and Information Operations are two of the best-known examples), which led to the reorganization and refocusing of the department's

---

[33] Author interview with Monroe Price, July 19, 2013.

IO structures.[34] Assessment has been a primary focus of these changes, with the standing requirement that "DoD IO programs and activities will incorporate an explicit means of assessing the results of operations in relation to expectations."[35] Congress now receives quarterly assessment reports on DoD IO activities. These reports do not yet fully satisfy congressional interests, however, and are viewed as part of an evolving process in need of further improvement and refinement. What does Congress need to consider its requirement met for assessment supporting accountability for DoD IIP efforts?

In the following sections, we explore the congressional mandate for accountability for DoD IIP efforts. We then touch on two other requirements for the assessment of these efforts, which are both congressionally motivated and in the interest of DoD leadership: to improve the effectiveness and efficiency of these efforts and to aggregate the results, lessons learned, and improvements stemming from assessments of IIP efforts into a broader campaign assessment. Chapter Four offers a more complete discussion of the challenges to merging these last two requirements and ensuring that IIP assessment meets its potential as a valuable contributor to overall campaign success—achieving stakeholder buy-in and navigating roadblocks to ensuring that assessment results reach the appropriate decisionmakers.

Requirements Regarding Congressional Interest and Accountability

To better understand congressional accountability requirements, we met with a number of congressional staffers to get the relevant stakeholder view. We also conducted interviews with DoD personnel involved in the process, including personnel in the Office of the Secretary of Defense with knowledge of the preparation and delivery of the final quarterly reports and personnel at the Joint Information Operations Warfare Center who provide subject-matter expertise to personnel at the geographic combatant commands and the assessment support and execution components.

Congressional interest is almost exclusively about accountability-oriented processes and summative assessment: How much did DoD spend on IO, what was done with that money, and what was accomplished? The decisions to be supported by these assessments concern funding and authority: Which, if any, IO programs should be funded? What legislative and policy constraints should be placed on the conduct of IO? What future oversight and reporting will be required?

Congressional interest connotes, in part, a very real threat to DoD IIP efforts; some in Congress are highly skeptical of the general efficacy of DoD's IIP efforts and would consider substantially curtailing such efforts and diminishing related capabili-

---

[34] Robert Gates, Secretary of Defense, "Strategic Communication and Information Operations in the DoD," memorandum, Washington, D.C., January 25, 2011.

[35] U.S. Department of Defense Directive 3600.01, *Information Operations (IO)*, Washington, D.C., May 2, 2013, p. 2.

ties.[36] As an outside observer, retired UK Royal Navy Commander Steve Tatham, the Ministry of Defence's longest-serving IO expert, has noted, "The U.S. has a very real 'baby with the bathwater' danger" regarding IO.[37] He elaborated that the existential threat to IO in the United States is real, and, to save it, there is a pressing need to admit that certain activities have been failures without showing that IO overall is a failure.

Some congressional stakeholders are more sanguine about the utility of IIP efforts but have a different question (and decision) in mind when they look to assessment: Who—that is, which government departments or organizations—should conduct IIP efforts and under what circumstances? This is much more of a policy question, and a political question, about the division of labor among government departments than it is about the assessment of DoD IO efforts.

Our conversations with congressional staffers provided several useful insights about the congressional requirement for accountability in DoD IIP efforts. In the following sections, we discuss the major themes.

Continue to Improve

Congressional staffers with whom we spoke acknowledged that there are challenges associated with assessment and that meeting congressional requirements is not an area in which assessment has traditionally been done or done well. As long as the quality of assessments continues to improve and draws closer to meeting congressional requirements—and as long as DoD demonstrates a good-faith effort toward this end—congressional stakeholders are willing to be patient. For some, however, patience is beginning to wear thin. As one congressional staffer told us, "There is an understanding that these things are hard, but they can't possibly be as hard as we've been told."[38] The consensus at the time of our interviews in mid-2013 was that Congress expects DoD to make continued progress in this area.

Progress from Outputs to Outcomes

Staffers told us that IO reporting to Congress remained too output focused rather than *outcome* focused. Reporting now effectively connects money to activities and programs; it indicates what those programs produce but stops short of connecting the results of activities with broader objectives. Staffers indicated that they would like to see assessments connect to strategy, to the outcomes of efforts. Mused one, "Could we get 'extent to which they accomplish [theater security cooperation plan] goals'?"[39]

---

[36] Author interview on a not-for-attribution basis, May 7, 2013.

[37] Author interview with UK Royal Navy CDR (ret.) Steve Tatham, March 29, 2013.

[38] Author interview on a not-for-attribution basis, May 7, 2013.

[39] Author interview on a not-for-attribution basis, May 7, 2013.

Standardize

Another theme in congressional staffers' comments regarding the requirement for IO assessments was that they be *standardized*. Several staffers used this term and expressed an interest in standardization at several levels: within the process that produces the assessments (standardized and routinized, so that different commands are asked the same questions in the same way) and in reporting across activities ("compare apples to apples").[40] The desire for standardization clearly connects to oversight decisions. Congressional stakeholders want to understand why some programs receive more resources than others, and they want to see which programs are particularly effective (or cost-effective) to inform resource allocation decisions.

Justify as DoD Activities

Another theme in our interviews with congressional staffers was the need for assessments to justify IO activities as appropriate pursuits for DoD. An underlying current in many recent congressional inquiries can be captured by the question, "Shouldn't the State Department be doing that?"[41] Congressional oversight extends beyond decisions about DoD resource allocation (and choosing one defense activity over another); Congress must make decisions about resource and authority allocation *across* departments as well—including assigning responsibility for conducting congressionally mandated activities. Although decisions of that kind are more a matter of policy and politics than accountability or improvement, they are still decisions that assessment can support, and in supporting such decisions, assessment can help ensure the continuity important to DoD IIP efforts (see Box 2.1).

As one congressional staffer suggested, "You've got to help the Hill get a better handle on why this is a military activity,"[42] adding that there is a requirement not only to make a logical argument that IIP efforts support military objectives but also to show, through assessment, that the efforts are effectively servicing those objectives. Good assessment, then, can meet multiple stakeholder needs by demonstrating that an IIP effort is effective and also by explicitly measuring its contribution to broader defense objectives. Congressional staffers indicated that it is much more compelling to measure the contribution of an effort to legitimate defense objectives than to simply argue that it contributes.[43]

Perspectives on Congressional Reporting from DoD Personnel

Many DoD personnel with whom we spoke about congressional IO assessment requirements shared the views of congressional staff. There was general acceptance

---

[40] Author interview on a not-for-attribution basis, May 7, 2013.

[41] Author interview on a not-for-attribution basis, May 7, 2013.

[42] Author interview on a not-for-attribution basis, May 29, 2013.

[43] Author interview on a not-for-attribution basis, May 29, 2013.

Box 2.1
Challenge: Lack of Shared Understanding

In our interviews, congressional staffers touched on a challenge that is inherent to IIP efforts relative to conventional kinetic military capabilities: a lack of shared understanding about, or intuition for, what IIP capabilities do and how they actually work (including a limited understanding of the psychology of influence).

Military personnel and congressional staffers have good intuition when it comes to the combined-arms contributions of different military platforms and formations. They also have a shared understanding of the force-projection capabilities of a bomber wing, a destroyer, an artillery battery, or a battalion of infantry. While congressional staffers may not know the exact tonnage of bombs or shells required to destroy a bridge, they certainly understand that bombs and shells can be used to destroy bridges.

This shared understanding does not extend to most IRCs. Congressional stakeholders (and, to be fair, many military personnel) do not necessarily have a shared understanding of the value of a leaflet drop, a radio call-in program, or a MISO detachment with a loudspeaker truck.

Intuition (whether correct or not) has a profound impact on assessment and expectations for assessment. Where shared understanding is strong, heuristics and mental shortcuts allow much to be taken for granted or assumed away. In its absence, everything has to be spelled out.

Consider the issue of standardization. In the realm of kinetic capabilities, about which there is strong shared understanding, no one asks that the capabilities be assessed against standardized benchmarks. Everyone understands that there is no standardized comparison measure for both aircraft carriers and infantry battalions; the two do not equate, and if trade-offs are sought, the balance has little to do with the relative merits of each and much more to do with the need to hedge against different global security threats. In contrast, for IIP capabilities, the lack of shared understanding reinforces the desire to standardize, in part as a substitute for intuition.

Consider the value of a capability—its ROI. As one of the military officers we interviewed remarked, "No one ever asks what the ROI was for a carrier strike group."[a] Many of the benefits of such naval forces are easy to comprehend but hard to quantify. There is, however, a shared understanding of the benefits (e.g., strike, deterrence, mobility, security, sometimes in a nebulous sense) and an appreciation for their complexity. There is also recognition of the time-conditional value of such capabilities: A carrier strike group has little ROI in port but a great deal of value during a contingency.

The story is slightly different when it comes to the ROI of IO investments and capabilities. Our interviews and literature review reinforced the conclusion that this is due to a general lack of shared understanding of the benefits of these efforts and the fact that many of these efforts are transitory (i.e., a contracted information campaign). For these reasons, there may be greater pressure to demonstrate the value of IO efforts. As one IO officer lamented, "We're held to different standards." This appears to be true.

Where shared understanding is lacking, assessments must be more thoughtful. The dots must be connected, with documentation to policymakers and other stakeholders spelling out explicitly what might be assumed away in other contexts. Greater detail and granularity become necessary, as do deliberate efforts to build shared understanding. Despite the burden of providing congressional stakeholders with more information about IIP efforts and capabilities to support their decisionmaking and fulfill oversight requirements, there are significant potential benefits for future IIP efforts. Greater shared understanding can not only help improve advocacy for these efforts but also strengthen the efforts themselves by encouraging more-rigorous assessments.

[a] Author interview on a not-for-attribution basis, August 1, 2013.

[b] Author interview on a not-for-attribution basis, October 28, 2013. See Chapter Five for a more detailed discussion of how this dynamic comes into play setting objectives for kinetic military efforts versus IIP efforts (in the section "How IIP Objectives Differ from Kinetic Objectives").

of the overall importance of assessment for accountability in support of the congressional oversight role: "Congress gave you this much money. What did you achieve?"[44] There was also recognition that IO-related reporting to Congress has improved since 2009,[45] though the consensus was that there is still room for improvement.[46] Many DoD respondents also agreed with the need for standardization to satisfy the need for accountability, both at the congressional level and within DoD chains of command.[47]

There was some divergence in perspectives on the matter of expectations, and this was a source of frustration for DoD respondents. Specifically, they viewed congressional reporting as a moving target, lacking clear articulation of what was actually desired or involving frequently changing questions.[48] We believe there is some truth to this complaint, in that it is clear that congressional staffers do not know exactly what the assessments will look like, and they may want data and answers that are simply not available (see Box 2.1 for a more detailed discussion of the dynamic at play). Further contributing to this perception is the way in which (often vague) congressional requests are translated through several layers of DoD bureaucracy and layers of command before reaching the level at which IIP activities are conducted and data are collected.

Another way in which the views of some defense personnel differed from congressional perspectives concerned the utility of the assessments produced. Although DoD respondents understood the need for assessment to meet congressional demand for accountability, they did not perceive this type of assessment as useful at the level of IO planning and execution: "At the end of the day—me, as a [theater special operations command] planner—I don't care why."[49] This view certainly has some merit. Accountability is also a priority within DoD, but it has a different character closer to the level of execution. This begs the question: What (if any) are DoD's IIP assessment requirements beyond congressional accountability?

In our discussion of the requirements for DoD IIP assessment, we have focused on congressional mandates, particularly those intended to enforce accountability and justify federal funding. Next, we pause briefly to preview the two primary roles assessment plays within DoD. There is certainly overlap between these two requirements and those articulated by Congress, but to a much greater degree, these requirements directly benefit IIP efforts and guide assessment in service of broader DoD goals: improving the effectiveness and efficiency of programs and ensuring the continuity

---

[44] Author interview on a not-for-attribution basis, August 1, 2013.

[45] Author interview on a not-for-attribution basis, July 31, 2013.

[46] William F. Wechsler, Deputy Assistant Secretary of Defense for Special Operations and Combating Terrorism, "Information Operations (IO) 2nd Quarter Reporting Requirement," memorandum, undated.

[47] Author interview on a not-for-attribution basis, July 30, 2013.

[48] Author interviews on a not-for-attribution basis, July 31 and August 1, 2013.

[49] Author interview on a not-for-attribution basis, October 30, 2013.

and value of efforts as part of larger campaigns. We revisit these themes repeatedly in examples of successful assessment efforts later in this report.

Requirement to Improve Effectiveness and Efficiency

In addition to the importance of assessment for meeting congressional accountability demands, DoD relies on assessment to improve the effectiveness and efficiency of all its programs. The current era of fiscal austerity has put pressure on budgets across DoD, and budgets for IIP efforts are no exception. Opportunities to increase the effectiveness, and cost-effectiveness, of such efforts cannot be missed. For example, suppose a recruiting campaign for partner-nation police targets three different groups with three different messages, but assessment reveals that no recruits are coming from one of those groups. Further assessment should help explain why and help program managers decide how to alter the messages or their delivery for that group. Formative assessment (perhaps focus groups, interviews, or small surveys) might help determine why products are not working and suggest potential changes. Iterative assessment, coupled with iterative changes to messages or delivery, can help managers find the best approach. Similarly, assessment can help monitor the performance of processes. Suppose two different contractors are delivering recruitment posters and flyers, but assessment reveals that one costs twice as much per unit volume delivered. Managers can take action based on this information, either finding an explanation (perhaps one contractor is delivering materials over a much more geographically dispersed area) or making a more informed decision when the contract is next competed (perhaps one contractor is just performing much better than the other).

Assessment supports learning from failure,[50] midcourse correction,[51] and planning improvements.[52] DoD requires IIP assessment for accountability purposes, of course, but it also depends on assessment to support a host of critical planning, funding, and process requirements.

Requirement to Aggregate IIP Assessments with Campaign Assessments

The final noteworthy requirement for DoD IIP assessment concerns the aggregation of assessments of individual IIP activities with larger campaign goals. The challenge here is twofold. First, the assessment of individual activities and programs does not necessarily connect to the assessment of overall campaigns or operations. It is a familiar dilemma in campaign planning and execution: *You can win the battles but still lose the war; the operation can be a success, but the patient can still die. The whole is sometimes greater than the sum of its parts.* This implies a requirement for assessment at multiple

---

[50] Author interview with Mary Elizabeth Germaine, March 2013.

[51] Haims et al., 2011, p. 2.

[52] Author interview with LTC Scott Nelson, October 10, 2013.

levels—at the level of the individual programs and activities, to be sure, but also at the level of contribution to overall campaigns.

Second, assessments of IIP efforts need to be aggregated with other military lines of operation as parts of whole campaigns. This is necessary not only to assess the contribution of IIP efforts to broader campaigns but also to better integrate such efforts into routine military planning and into the overall military assessment process, a process from which IO have often been excluded, historically.[53]

Summary

This chapter provided a general introduction to the critical role of assessment in terms of meeting congressional requirements, serving larger DoD goals, and supporting the refinement and improvement of IIP efforts themselves. We also reviewed the primary motivations for conducting assessment and evaluation and provided an introduction to the prevailing types of assessment that can serve the needs of DoD IIP efforts in meeting requirements at multiple levels. Key takeaways include the following:

- Formative, process, and summative evaluations have nested and connected relationships in which unexpected results at higher levels can be explained by thoughtful assessment at lower levels. This is captured in the hierarchy of evaluation.
- Good assessment supports and informs decisionmaking.
- There are a range of different uses for and users of assessment. As we discuss in greater detail in Chapter Eleven, assessments need to be tailored to the needs of users in both design and their presentation.
- Assessment of IIP efforts for accountability purposes is complicated by a lack of shared understanding or intuition. Everyone can intuit the value of kinetic military capabilities (an aircraft carrier or infantry battalion, for example), but this is not necessarily true for IIP. A result is greater uncertainty about the basic value of IIP efforts and an increased need for granularity and specificity in IIP assessment.
- In addition to accountability, the DoD assessment requirement supports the greater effectiveness and efficiency of IIP efforts. Some good efforts can undoubtedly be better, and some weaker efforts could be made better through evaluation and assessment.
- You can win the battles but still lose the war; the operation can be a success, but the patient can still die. DoD IIP assessment must address many needs simultaneously: those of the individual efforts, those of broader campaigns, and the contribution of the former to the latter.

---

[53] Author interview on a not-for-attribution basis, August 1, 2013.

Applying Assessment and Evaluation Principles to IIP Efforts

Across all the sectors reviewed in our study (industry, academia, and government), certain headline principles appeared again and again. We collected and distilled the most central (and most applicable to the defense IIP context) and present them here:

- Effective assessment requires clear, realistic, and measurable goals.
- Effective assessment starts in the planning phase.
- Effective assessment requires a theory of change/logic of the effort connecting activities to objectives.
- Evaluating change requires a baseline.
- Assessment over time requires continuity and consistency.
- Assessment is iterative.
- Assessment requires resources.

We discuss each principle in greater detail in the sections that follow.

Effective Assessment Requires Clear, Realistic, and Measurable Goals

It appears to be self-evident that it is impossible to do assessment without having a clear goal in mind. Consider the three stages of evaluation, discussed in Chapter Two: How can you do summative evaluation, which seeks to determine whether an effort has achieved its desired outcomes, if the desired outcomes are not clear? How can you do formative evaluation, which supports the development and design of activities to accomplish desired goals, if the desired goals have not yet been articulated? How can you do process evaluation if it is not clear what the process is supposed to accomplish?

Assessment and evaluation advice from every sector comes with an admonition to set clear goals. In the public relations world, "the importance of goal setting and measurement" is the first of the seven "Barcelona Principles," the industry standard for rig-

orous measurement.[1] Internal guidance documents for assessment by Ketchum Global Research and Analytics note, "A clear set of goals is key to understanding what you want to achieve and hence measuring it."[2] "Begin with the end in mind" is the advice given by Sarah Bruce and Mary Tiger for social marketing campaigns.[3] UK Ministry of Defence doctrine for assessment offers four assessment principles, the first of which is "objectives led"; it notes, "The assessment should be derived from the campaign objectives (end-state), otherwise it is likely to be irrelevant."[4]

While the importance of clear goals appears to be a self-evident requirement and is repeated throughout the existing assessment advice, too often this obvious requirement is not met. According to one industry SME, a complete lack of clarity about end goals prior to launching an assessment program renders any data collected unusable, a situation she had seen many times.[5] In the words of the public communication evaluation consultant Pamela Jull, "We'll often get called to help out with a neat idea, and people cannot articulate what they're trying to achieve."[6] Such challenges are not uncommon in defense assessment efforts, either. As one DoD SME described defense IIP goals, "Too often, lofty goals that are unattainable."[7] A PSYOP officer we interviewed raised concerns about the MISO planning process, indicating that if the objectives are flawed, the whole process will be flawed, adding that he had seen such a situation occur and unfold into failure.[8]

Though it seems self-evident, when conducting (or planning) assessment, remember that "it is practically impossible to evaluate something if your goal isn't explicit."[9]

Assessment and evaluation require not just goals but clear, realistic, specific, and measurable goals. Goals must be realistic or assessment becomes unnecessary; unrealistic goals cannot be achieved, so there is no point in assessing. The prevailing advice from the evaluation research is clear: When planning a project, planners should consider what results they would like to achieve, the processes that are most likely to lead to those results, and the indicators to determine whether or not those results have been

---

[1] "Barcelona Declaration of Measurement Principles," 2nd European Summit on Measurement, International Association for Measurement and Evaluation of Communication, July 19, 2010.

[2] Ketchum Global Research and Analytics, *The Principles of PR Measurement*, undated, p. 6.

[3] Sarah Bruce and Mary Tiger, *A Review of Research Relevant to Evaluating Social Marketing Mass Media Campaigns*, Durham, N.C.: Clean Water Education Partnership, undated, p. 3.

[4] UK Ministry of Defence, 2012.

[5] Author interview with Angela Jeffrey, April 3, 2013.

[6] Author interview with Pamela Jull, August 2, 2013.

[7] Author interview on a not-for-attribution basis, July 30, 2013.

[8] Author interview on a not for attribution basis, October 28, 2013.

[9] Author interview with Gaby van den Berg, April 22, 2013.

achieved.[10] One defense SME we interviewed summed up the importance of clear, measurable objectives quite succinctly: "An effect that can't be measured isn't worth fighting for."[11]

The requirement for clear, realistic, and measurable goals also frequently goes unmet in practice. The RAND academic evaluation research expert Joie Acosta reports that clients often want their projects to accomplish more than is feasible, and that objectives change and are "moving targets."[12] A PSYOP soldier we interviewed described how mission objectives and PSYOP objectives are often expressed as aspirational rather than measurable and achievable objectives.[13] This problem is not unique to MISO; it can arise in any IRC effort or in IO more broadly.

JP 5-0's discussion of operational art and operational design highlights the importance of clear objectives while recognizing that complex or ill-defined problems or a disconnect between strategic and operational points of view can impede progress toward clear objectives. JP 5-0 notes, "Strategic guidance addressing complex problems can initially be vague, requiring the commander to interpret and filter it for the staff."[14] It goes on to note that subordinates should be aggressive in sharing their perspectives with higher echelons, working to resolve differences at the earliest opportunity. This is useful advice for assessors: If the provided objectives are too vague to assess against, try to define them more precisely and then push them back to higher levels for discussion and confirmation. In JOPP, most of the elements of operational design should take place as part of step 2, mission analysis. Mission analysis is when objectives should be articulated and refined, in concert with higher headquarters, if necessary. Clear objectives should be an input to mission analysis, but if they are not, mission analysis should provide an opportunity to seek refinement.

In our interviews, one SME suggested that the problem of inadequately specified objectives could be partially solved by articulating measurable subordinate objectives, though it can still be difficult to connect low-level measurable objectives with high-level strategic objectives, and that can cause further assessment challenges.[15] In this same vein, subordinates can "lead up" with regard to goals, not only specifying measurable subordinate goals but also adding specificity to higher-level goals and then submitting those rearticulated goals to higher command levels for review. Sending slightly modified goals back up the chain of command could produce one of two positive outcomes: either approval and acceptance of the rearticulated objectives or their rejection,

---

[10] Haims et al., 2011, p. 9.

[11] Author interview on a not-for-attribution basis, December 5, 2012.

[12] Author interview with Joie Acosta, March 20, 2013.

[13] Author interview on a not-for-attribution basis, January 23, 2013.

[14] U.S. Joint Chiefs of Staff, 2011a, p. III-3.

[15] Author interview on a not-for-attribution basis, January 23, 2013.

ideally accompanied by needed specificity from that higher echelon.[16] We discuss the importance of these concepts again in Chapter Five in the context of setting clear and measurable goals.

Effective Assessment Starts in Planning

In the words of Jonathan Schroden, who played a pivotal role in redesigning the International Security Assistance Force (ISAF) campaign assessment process while serving as CNA's field representative to the Afghan Assessments Group, "Problems in assessment stem from problems in planning."[17] As noted earlier, assessment requires clear, realistic, and measurable goals. Goal refinement and specification should be important parts of the planning process, and the need to articulate assessable goals and objectives is certainly part of what is meant when experts advise that assessment start in planning: "Assessment begins in plan initiation during mission planning and continues throughout the campaign. Approaching this from the start, the assessor can ensure the commander has well-defined, measurable and achievable effects or end-state."[18] If poorly specified or ambiguous objectives survive the planning process, both assessment and mission accomplishment will be in jeopardy.[19]

There is more to it than that, however. In addition to specifying objectives in an assessable way during planning, assessments should be designed and planned alongside the planning of activities so that the data needed to support assessment can be collected as activities are being executed. Knowing what you want to measure and assess at the outset clarifies what success should look like at the end and allows you to collect sufficient information to observe that success (or its lack).[20]

Assessment personnel need to be involved in planning to be able to point out when an objective or subordinate objective is or is not specified in a way that can be measured and to identify decisions or decision points that could be informed by assessment. Assessors should involve planners in assessment design to ensure that assessments will provide useful information, that they will be designed to collect the desired data, and that they have stakeholder buy-in.[21] For example, at the British Broadcasting Corporation's (BBC's) international development charity, BBC Media Action, evalu-

---

[16] Christopher Paul, *Strategic Communication: Origins, Concepts, and Current Debates*, Santa Barbara, Calif.: Praeger, 2011.

[17] Author interview with Jonathan Schroden, November 12, 2013.

[18] The Initiatives Group, *Information Environment Assessment Handbook*, version 2.0, Washington, D.C.: Office of the Under Secretary of Defense for Intelligence, 2013, p. 4.

[19] Author interview on a not-for-attribution basis, January 23, 2013.

[20] Author interview with Rebecca Andersen, April 24, 2013.

[21] Author interview with Gerry Power, April 10, 2013.

ators help set (and specify) program goals during the program planning process; this approach ensures greater continuity between the program's design and what the evaluation is intended to measure, and this feedback loop ensures that assessment results can directly inform program improvement.[22]

LTC Scott Nelson, who served as the chief of influence assessment at U.S. Northern Command (USNORTHCOM), went so far as to suggest that "assessment should drive the planning process."[23] He argued that military planning and decisionmaking processes are designed in a way that supports assessment-driven planning: These processes are supposed to work backward from measurable objectives in much the same way as good assessment design. The *Commander's Handbook for Assessment Planning and Execution* notes, "Planning for assessment begins during mission analysis when the commander and staff consider what to measure and how to measure it in order to determine progress toward accomplishing a task, creating an effect, or achieving an objective."[24]

There is a feedback loop here, too. Inasmuch as assessment plans should be part of activity plans, assessment results should feed back into future planning cycles—cycles in which activity (and assessment) plans may evolve as understanding of the context improves, as objectives are refined, or as additional lines of effort are added. In the words of Marine Air-Ground Task Force Training Program materials, "Assessment precedes, accompanies and follows all operations."[25]

SMEs across sectors recounted horror stories in which assessment was not considered at the outset. If stakeholders do not think about measurement until after the fact, assessment could be more difficult, if not impossible.[26] On the other hand, SMEs also reported clear examples of the successful integration of assessment into the planning process. The Navy's Pacific Fleet N5 was intimately involved in assessment and assessment planning for Pacific Partnership exercises in 2012 and 2013 and reported that integrated planning and assessment were critical and beneficial for both assessment and planning.[27]

In the JOPP framework, assessment considerations should be present at the earliest stages. Formative assessment may inform operational design during mission analysis. Preliminary assessment plans should be included in COA development and should be war-gamed along with other COA elements during COA analysis and war-gaming.

---

[22] Author interview with James Deane, May 15, 2013.

[23] Author interview with LTC Scott Nelson, October 10, 2013.

[24] U.S. Joint Chiefs of Staff, 2011c, p. IV-1.

[25] U.S. Marine Corps, *Assessment: MAGTF Staff Training Program (MSTP)*, MSTP Pamphlet 6-9, Quantico, Va.: Marine Air-Ground Task Force Staff Training Program, October 25, 2007, p. 1.

[26] Author interview with Angela Jeffrey, April 3, 2013.

[27] U.S. Pacific Fleet, "Pacific Partnership 2012 to 2013: Assessment Transition Brief," briefing, undated.

Effective Assessment Requires a Theory of Change or Logic of the Effort Connecting Activities to Objectives

Implicit in many examples of effective assessment and explicit in much of the work by scholars of evaluation is the importance of a *theory of change*.[28] The theory of change or logic of the effort for an activity, line of effort, or operation is the underlying logic for how planners think that elements of the overall activity, line of effort, or operation will lead to desired results. Simply put, a theory of change is a statement of how you believe that the things you are planning to do are going to lead to the objectives you seek. A theory of change can include logic, assumptions, beliefs, or doctrinal principles. The main benefit of articulating the logic of the effort in the assessment context is that it allows assumptions of any kind to be turned into *hypotheses*. These hypotheses can then be explicitly tested as part of the assessment process, with any failed hypotheses replaced in subsequent efforts until a validated, logical chain connects activities with objectives and objectives are met. This is exactly what is described in the *Commander's Handbook for Assessment Planning and Execution*: "Assumptions made in establishing cause and effect must be recorded explicitly and challenged periodically to ensure they are still valid."[29]

Here is an example of a theory of change/logic of the effort:

> Training and arming local security guards makes them more able and willing to resist insurgents, which will increase security in the locale. Increased security, coupled with efforts to spread information about improvements in security, will lead to increased perceptions of security, which will, coupled with the encouragement to do so, promote participation in local government, which will lead to better governance. Improved perceptions of security and better governance will lead to increased stability.

As is often the case with IIP objectives, the IIP portion (increased perceptions of security and increased participation in local government) of this theory of change is just one line of effort in an array of efforts connected to the main goal. The IIP portion is dependent on the success of other lines of effort—specifically, real increases in security.

This theory of change shows a clear, logical connection between the activities (training and arming locals, spreading information about improving security) and the desired outcomes, both intermediate (improved security, improved perceptions of security) and long-term (increased stability). The theory of change makes some assumptions, but those assumptions are clearly stated, so they can be challenged if they prove

---

[28] Much of the discussion in this section is drawn directly from Christopher Paul, "Foundations for Assessment: The Hierarchy of Evaluation and the Importance of Articulating a Theory of Change," *Small Wars Journal*, Vol. 10, No. 3, 2014.

[29] U.S. Joint Chiefs of Staff, 2011c, pp. II-10.

to be incorrect. Further, those activities and assumptions suggest things to measure: the performance of the activities (training and arming, publicizing improved security) and the ultimate outcome (change in stability), to be sure, but also elements of all the intermediate logical nodes, such as the capability and willingness of local security forces, change in security, change in perception of security, change in participation in local government, and change in governance. Evaluation researchers assert that measures often "fall out" of a theory of change.[30]

The theory of change suggests things to measure, and if one of those measurements does not report the desired result, assessors will have a fairly good idea of where in the chain the logic is breaking down (that is, which hypotheses are not substantiated). They can then make modifications to the theory of change and to the activities being conducted, reconnecting the logical pathway and continuing to push toward the objectives.

Articulated at the outset, during planning, a theory of change can help clarify goals, explicitly connect planned activities to those goals, and support the assessment process.[31] A good theory of change will also capture possible unintended consequences or provide indicators of failure, things to help you identify where links in the logical chain have been broken by faulty assumptions, inadequate execution, or factors outside your control (disruptors).[32] Identifying and articulating a theory of change (and expressing a theory of change as a logic model) is discussed in greater detail in Chapter Five.

Evaluating Change Requires a Baseline

Olivier Blanchard writes, "Regardless of your focus (macro- or micro-measurement), what you are looking for in these data sets is change. What you want to see are shifts in behavior indicating that something you are doing is having an effect."[33] To see change (delta), you need a starting point, a baseline with which to compare and from which to measure change. Further, it is best to measure the baseline before your interventions—your IIP activities—begin.[34]

---

[30] The quote is from an interview with Christopher Nelson, February 18, 2013; for the general principle, see William J. McGuire, "McGuire's Classic Input-Output Framework for Constructing Persuasive Messages," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2012.

[31] Author interview with Maureen Taylor, April 4, 2013.

[32] Author interview with Steve Booth-Butterfield, January 7, 2013.

[33] Olivier Blanchard, *Social Media ROI: Managing and Measuring Social Media Efforts in Your Organization*, Indianapolis, Ind.: Que, 2011, p. 201.

[34] Author interview with Charlotte Cole, May 29, 2013.

Box 3.1
Nested Objectives

Supporting intermediate steps in a theory of change and making it easier to get to specific and measurable objectives is the idea of nesting, as described in Chapter Two. If an overall objective can be broken into several subordinate, intermediate, or incremental steps, it will be easier to specify a theory of change, measure those nested objectives, and conduct productive assessment.

The *Commander's Handbook for Assessment Planning and Execution* provides an example of such nesting, describing how tactical objectives and missions support operational-level objectives and end states, which support theater strategic objectives and end states.[a] A contractor who trains defense personnel in IIP assessment indicated that the contracted organization teaches a corresponding approach that begins with clear, overarching objectives but then necks down to specific supporting behavioral objectives based on desired outcomes on the ground.[b] This "necking down" is nesting.

Ideally, nested goals will not just be subordinate but also be sequential and incremental, moving one step at a time along a logical pathway that culminates with the overall objective. A MISO soldier pointed out the importance of incremental goals, especially when the ultimate goal is long-term; being able to show slow-burn progress—but real, scientifically measured progress—toward stated intermediate goals is important for accountability and justifying the continuation of an effort.[c] Another MISO SME advocated moving to more-segmented supporting PSYOP objectives, breaking bigger problems into smaller, incremental segments.[d] Input from other SMEs and principles distilled from the literature across sectors endorsed this view.

In JOPP, specification for nesting objectives is part of the broader process of setting goals and identifying objectives, which should take place during mission analysis. Operational design, a primary approach to mission analysis (see Chapter One), recommends thoughtfully defining the problem and developing an operational approach that contains the solution. The design process should strive to specify both the problem and the solution in smaller, discrete, nested chunks.

The example theory of change in which the training and arming of local security guards was hypothesized to increase security illustrates nested objectives in a defense IIP context. The short version of the theory is, provide arms and training to local security forces and promote awareness of improved security and participation in government, and stability (the overall goal) will result. The long version, with nested incremental goals, includes succeeding at training and arming local forces, succeeding at improving security, succeeding at improving perceptions of security, succeeding at improving participation in local government, succeeding at improving local governance, and, finally, achieving improved stability. Spelling out the intermediate steps reveals incremental progress (perhaps training and equipping have gone well and security is improving, but perceptions of security still lag) and identifies mistaken assumptions that can be corrected (perhaps security and perceptions of security have improved, but apathy, rather than fear, kept locals from voting).

[a] U.S. Joint Chiefs of Staff, 2011c, p. I-8.

[b] Author interview with Gaby van den Berg, April 22, 2013.

[c] Author interview on a not-for-attribution basis, July 30, 2013.

[d] Author interview on a not-for-attribution basis, January 23, 2013.

While the need for a baseline against which to evaluate change and the importance of taking a baseline measurement before change-causing activities begin again seem self-evident, these principles are often not adhered to in practice. One defense SME noted that baselines were often omitted because of insufficient time and resources.[35] Another observed that, sometimes, baseline data are collected, but forces end up

---

[35] Author interview on a not-for-attribution basis, January 23, 2013.

revising the baseline, either because the objectives changed (moving target) or because the next rotation of command or authority began the assessment process anew.[36]

The election-participation campaign example from our discussion of the three types of evaluation in Chapter Two illustrates the importance of a baseline. Without a baseline measurement of some kind to inform expectations of turnout (based on previous elections, surveys of intention, or some other source), it would be impossible to say whether DoD efforts to promote participation actually had any impact. It is sometimes possible to complete post hoc baselines against which to assess, but it is best to collect baseline data at the outset. Also note that while a baseline is essential to evaluating change, it is not always imperative that baseline data be quantitative. Sometimes, qualitative baseline data (such as data from focus groups) can provide a sufficient baseline.[37]

Assessment over Time Requires Continuity and Consistency

The previous discussion touched on "moving target" problems, where either the objectives change or the baseline is redone. These challenges point to a broader assessment principle—namely, the importance of continuity and consistency. A trend line is useful only if it reports the trend in a consistently measured way and if data are collected over a long enough period to reveal a trend. Assessment of progress toward an objective is useful only if that objective is still sought. Consistent, mediocre assessments are better than great, inconsistent assessments in many contexts.[38]

A lack of continuity and consistency is noted as a problem in industry and in evaluation research,[39] but not at the same scale as in the defense sector. The major culprit in the defense context is rotation, including personnel rotation, unit rotation, and rotation at the senior command (and combatant command) levels.

The frequent turnover of analysts can threaten continuity in assessment.[40] Further, whole assessment processes are often scrapped when new units rotate in and take over operations.[41] Changes in senior leadership can result in changes in objectives or guidance or, worse, cancellation of existing objectives or guidance without imme-

---

[36] Author interview on a not-for-attribution basis, September 8, 2013.

[37] Author interview with Kavita Abraham Dowsing, May 23, 2013. Baselines are discussed in greater detail in Chapter Eight.

[38] Author interview on a not-for-attribution basis, August 1, 2013.

[39] Rossi, Lipsey, and Freeman, 2004.

[40] P. T. Eles, E. Vincent, B. Vasiliev, and K. M. Banko, *Opinion Polling in Support of the Canadian Mission in Kandahar: A Final Report for the Kandahar Province Opinion Polling Program, Including Program Overview, Lessons, and Recommendations*, Ottawa, Ont.: Defence R&D Canada, Centre for Operational Research and Analysis, DRDC CORA TR 2012-160U, September 2012.

[41] Author interview on a not-for-attribution basis, August 1, 2013.

diate replacements.[42] Especially in a military context, objectives—even long-term objectives—will change periodically. Still, as one defense assessment SME noted, "We don't need a new revision of the [campaign plan] every eight months. That is ridiculous."[43] Significantly changing objectives can damage the design of the assessment processes marking progress toward those objectives, leaving "assessment widows" (assessments of progress toward outdated objectives) or forcing assessment processes to continually restart with new objectives (and new baselines).

Thoughtful nested or subordinate objectives can help mitigate against changing objectives at the highest level, provided existing subordinate objectives remain constant and still nest within new capstone objectives. Loss of continuity when rotating units abandon existing assessment frameworks might be avoidable if assessment practice improved in general, and if the leaders of the subsequent unit were more willing to accept existing "good enough" assessment rather than starting fresh every time.[44]

Assessment Is Iterative

Many of the SMEs we interviewed observed that, in many ways, assessment must be an iterative process, not something planned and executed once. First, efforts to track trends over time or to track incremental progress toward an objective require repeated, iterative measurement. Second, assessment needs to be planned and conducted iteratively, as things change over time; objectives can change, available data (or the ease of collecting those data) can change, or other factors can change, and assessment must change with them. A public relations expert reminded us to expect the unexpected, adding that things can happen over the course of a campaign or assessment process that can affect outcome but that you cannot control.[45]

Third, and related, IIP efforts involve numerous dynamic processes and thus require dynamic evaluation. Context changes, understanding of the context changes, theories of change change, and activities change based on revisions to theories of change; assessments need to adapt to reflect all of these changes. As IIP activities change, measures must be recalibrated and corrected, iteratively, along the way.[46]

Fourth, as activities expand, assessment needs to change and expand with them. Stakeholders from the Cure Violence community violence–prevention campaign described for us the progress of their program, from initial success to refinement and

---

[42] Author interview on a not-for-attribution basis, April 3, 2013.

[43] Author interview with John-Paul Gravelines, June 13, 2013.

[44] Author interview on a not-for-attribution basis, April 3, 2013.

[45] Author interview with Rebecca Andersen, April 24, 2013.

[46] Author interview with David Michaelson, April 1, 2013.

expansion.[47] They needed to know which aspects of their program were most successful and how to match components of their efforts to new contexts, and they needed to measure progress and effectiveness in their new, expanded service areas. Evolving iterations of assessment helped them pursue each line of inquiry.

Just about any assessment effort that contains unknowns or potentially unvalidated assumptions, or that intends to affect a dynamic context, might require iteration and change. Only an effort that has stable objectives and processes, and that is functioning effectively, is likely to have stable assessments, but periodic repetition and iteration will still be needed to make sure that everything is on track. Any nascent effort should expect iteration in both design and measurement. In the example of a DoD effort to encourage participation in partner-nation elections, presented in Chapter Two, there were several instances of iteration. In the formative stage, there might be repeated focus groups or small surveys, with each iteration being slightly revised and different, and testing constantly evolving hypotheses about how best to promote election participation. Once implementation begins, materials may be disseminated as planned, but assessments might indicate that these materials are not reaching target audiences, necessitating a new iteration of design for dissemination. Perhaps the planned means of measuring the receipt of messages by the audiences fails to collect sufficient (or sufficiently accurate) measures. This could indicate a need to revisit that portion of the assessment design. Moving toward the summative phase, early measures of intention to vote may not indicate as much improvement as desired, pushing the focus back on formative assessment to develop additional materials or efforts to achieve the desired effect. Finally, even when the election is over and efforts have succeeded or failed, viewing the program as one iteration in a possible series of election support programs that DoD might conduct encourages the collection of lessons learned for both the execution and assessment of such programs in the future.

Assessment Requires Resources

An emphasis in the literature and in our interviews was that assessment requires prioritization and a commitment of resources if it is going to succeed.[48] Organizations that routinely conduct successful evaluations have a respect for research and evaluation ingrained in their organizational cultures, and they dedicate substantial resources to evaluation.[49]

---

[47] Author interview with Joshua Gryniewicz, August 23, 2013.

[48] For example, in our interview with John Croll, April 10, 2013.

[49] Author interview with James Deane, May 15, 2013.

Unfortunately, assessment of DoD IIP efforts has been perennially underfunded, as these quotes from defense SMEs indicate:

- "How can I get the best assessment at no cost, or very low cost?"[50]
- "I wish I had time and resources to not only do better assessments but plan better assessments."[51]
- "Part of the problem is resources. They are working on a shoestring budget. There is so much ambiguity in assessment because they can't fund it properly."[52]
- "We are not funded, manned, trained, or equipped to do assessments, period."[53]

Numerous defense SMEs advocated increased investment in IIP assessment in DoD, in terms of overall funding, personnel allocations, and training and expertise.[54]

The statement that assessment requires resources warrants a caveat. Especially for small-scale IIP efforts, assessment investment has to be reasonable relative to overall program costs. One cannot and should not spend more on assessment than on the activities being assessed! Evaluators must be able to work with what their budget allows,[55] and there has to be a budget balance between assessment and activity.

With that in mind, our reviews and interviews suggested two further subordinate principles. First, some assessment (done well) is better than no assessment. Even if the scope is narrow and the assessment effort is underfunded and understaffed, any assessment that reduces the uncertainty under which future decisions are made adds value. Second, not all assessment needs to be at the same level of depth or quality. Where assessment resources are scarce, they need to be prioritized.

We identified two resource-saving priority areas for the assessment of DoD IIP efforts. First, emphasize just a sample of very similar efforts. For example, rather than assessing four similar efforts at the same (inadequate) level, it might be better to pursue a high-quality assessment of just one of those efforts, seeking to validate (or improve) the theory of change and discern the most-effective processes in that single effort. Based on those findings and, perhaps, minimal process assessment (collecting MOPs) for the other efforts, similar levels of success could be assumed for those other efforts (if the MOPs match). Perhaps which of the four efforts received high-intensity evaluation could be periodically rotated.

---

[50] Author interview on a not-for-attribution basis, December 5, 2012.

[51] Author interview on a not-for-attribution basis, December 5, 2012.

[52] Author interview on a not-for-attribution basis, July 30, 2013.

[53] Author interview on a not-for-attribution basis, July 31, 2013.

[54] For example, in our interviews with LTC Scott Nelson, October 10, 2013, and Steve Booth-Butterfield, January 7, 2013.

[55] Author interview with Sam Huxley, May 9, 2013.

Second, deemphasize the assessment of efforts that have very modest objectives or expenditures. Some efforts are not particularly extensive or ambitious, and progress toward those modest objectives could be assessed informally, based on the expert opinions of those conducting the activities. Consider, for example, that DoD efforts to inform, influence, or persuade certain target audiences in certain countries are nascent, with the current level of engagement focusing on getting a foot in the door, opening lines of communication, or identifying channels through which to conduct future IIP efforts. In such cases, *any* engagement is a success. With certain military-to-military engagements, engaging at all is a step in the right direction. In other places (and for other audiences), the relationship is much more mature, and IIP objectives have progressed beyond initial engagement and connection. The former require minimal assessment effort and expense, while the latter certainly merit more-substantial evaluation.

Not all efforts merit the same level of assessment investment; the trick, then, is in recognizing which require substantial assessment and which do not. In our example, at some point after a sufficient number of successful foot-in-the-door engagements, the effort will presumably be ready to make progress toward the next incremental objective, and measuring progress toward that objective may well require more-substantial assessment. (But, again, there is little point in measuring progress toward a later objective that actual IIP efforts are not yet trying to achieve.)

Summary

This chapter reviewed the core principles revealed in our research that are applicable to the assessment and evaluation of defense IIP efforts. Key takeaways echo the principles themselves:

- Effective assessment requires clear, realistic, and measurable goals. As one DoD respondent aptly noted, "An effect that can't be measured isn't worth fighting for."[56] Nor is one that cannot be achieved.
- Assessment must start in planning, for two reasons. First, assessment should be integrated into the plan, ensuring that assessment data collection and analysis are part of the plan (rather than something done, possibly inadequately, after the fact). Second, assessment requires assessable goals, and those goals need to be established as part of planning.
- Assessment requires an explicit theory of change, a stated logic for how the activities conducted are meant to lead to the results desired. Assessment along an effort's chain of logic enables process improvement, makes it possible to test assumptions,

---

[56] Author interview on a not-for-attribution basis, December 5, 2012.

and can tell evaluators why and how (that is, where on the logic chain) an unsuccessful effort is failing.

- Assessment should have nested and connected levels and layers. This can be the nesting of different types of evaluation (formative, process, and summative) in the same activity or the nesting of objectives and subordinate/supporting objectives and their assessment. Objectives should be broken into progressive, sequential, incremental chunks and assessed in those nested layers.
- To evaluate change, a baseline of some kind is required. While it is sometimes possible to construct a post hoc baseline, it is best to have baseline data before the activities to be assessed have begun.
- Assessment over time requires continuity and consistency in both objectives and assessment approaches. Consistent mediocre assessments are more useful than great, inconsistent assessments.
- The biggest threat to continuity and consistency in the defense context is rotation. Setbacks occur when new commanders change objectives and when new units change subordinate objectives and start new assessment processes.
- Assessment is iterative. Rarely does anything work exactly as intended, and contextual conditions change. Iterative assessment can show incremental progress toward objectives and help plans, processes, procedures, and understanding evolve.
- Assessment is not free; it requires resources. However, some assessment is better than no assessment, and not every activity merits assessment at the same level.

Challenges to Organizing for Assessment and Ways to Overcome Them

To this point, this report has focused predominantly on the principles of good assessment and how to apply them to DoD IIP efforts. We have touched on some organizational and contextual constraints, but the discussion so far has emphasized the characteristics of good assessment, with less attention to the organizations that conduct assessments and the various challenges of bureaucracy, business processes, and organizational structures and cultures. This chapter provides some insights from organizations that have succeeded at assessment and tips to inform thinking about organizing for assessment in the DoD IIP context. This chapter elaborates on the following key lessons for funders and other stakeholders, DoD leadership, and practitioners involved in designing and assessing IIP efforts:

- Organizations that do assessment well usually have cultures that value assessment.
- Assessment requires resources (as a rule of thumb, roughly 5 percent of program resources should be dedicated to assessment).
- Successful assessment depends on the willingness of leadership to learn from the results. (This echoes the admonition in Chapter Three for leaders to promote and embrace constant change, learning, and adaptation, as discussed in JP 5-0.)
- Assessment requires data to populate measures—and intelligence is potentially a good data source.
- IIP efforts should be broadly integrated into DoD processes, and IIP assessment should be integrated with broader DoD assessment efforts. The *Commander's Handbook for Assessment Planning and Execution* aims to fill the gap in doctrinal focus on assessment with guidance that complements existing service- and joint-level guidance; this is why we point out throughout this report where observed strong practices would conform to JOPP guidance.
- Assessment needs advocacy, improved doctrine and training, more trained personnel, and greater access to assessment and influence expertise to break the current "failure cycle" for assessment in DoD.
- Independent assessment and formal devil's advocacy are valuable tools in promoting a culture of assessment, especially in avoiding rose-tinted glasses in under-

standing the operational environment. These approaches could be incorporated into JOPP during COA analysis and war-gaming, but they should also be included in the iterative cycle of operational design.

- Assessment starts in planning and continues through execution. Overlaying the JOPP steps, this means that assessment begins with mission analysis (step 2) and continues through to step 7, plan or order development.

Building Organizations That Value Research

When it comes to the successful conduct of assessment, one point cannot be overstated: Organizations that do assessment well usually have a culture that values assessment. Without an understanding and appreciation for what assessment can accomplish, it is much easier to dismiss assessment as an afterthought. A critical component to conducting assessment—albeit a component that is often underappreciated—is building organizations that value *research*.

Building an Assessment Culture: Education, Resources, and Leadership Commitment

Introducing new concepts and initiating change in an organization is typically met with resistance. Organizations, and the individuals they comprise, can be reticent to anything other than "business as usual." Creating an atmosphere in which assessment is understood and appreciated takes time, especially where such a culture never existed before. Successful cultural change depends on a strong commitment from leadership. Leaders who value assessment, make decisions supported by assessment output, and are willing to allocate resources to assessment can make a huge difference. A significant part of creating this climate is fostering an appreciation for research. Too often, "creative" people (planners) are skeptical of research and its uses, which is understandable when one considers that research can threaten power. To overcome this resistance, leadership is paramount.

An example of balancing the creative tension between program creators and researchers or evaluators can be found in the BBC's international development charity, BBC Media Action, which made a strategic decision almost a decade ago to shift more resources toward research. Rather than a technical decision or augmentation, this was an investment priority, and a difficult one at that, since this wing of the company commanded only a small "core" budget. Nevertheless, the result has been an organization with research "ingrained in its DNA," according to James Deane, the director of policy and learning at BBC Media Action.[1] Kavita Abraham Dowsing, the director

---

[1]   Author interview with James Deane, May 15, 2013. As the director of policy and learning at BBC Media Action, Deane is responsible for three subdivisions (research, policy, and advisory).

of research, concurs, observing that there is a "cradle to grave" mode of research at her organization that is absent in most others in this sector. She estimates that roughly 98 percent of the work starts with research and ends with research.[2]

Ronald Rice, a leading expert in public communication evaluation and the editor of *Public Communication Campaigns*, supports the concept of "management by evidence" rather than "management by assertion."[3] Building this aspect of organizational culture requires demonstrating initiative or, put simply, leading by example. One way to promote the prominence of assessment in the DoD context would be to embrace assessment in all aspects of JOPP—and to make assessment a routine consideration in the planning process.

Building an assessment culture requires identifying enablers of the integration of evaluation into the organizational culture. In their work looking at the United Way of Greater Toronto, Jill Anne Chouinard and colleagues found three principal and interrelated enablers of an evaluation culture. The first was a more formal commitment from the leadership (senior management and board of directors) to developing a learning organization. This required learning from evaluation rather than seeing it only as an accountability mechanism. The second enabler was education and the development of a mind-set around evaluation, along with an attitude that signaled a willingness to learn and change. Finally, the researchers found that resources and time were critical to developing a culture of evaluation, meaning that staff had the resources and time required to figure out what evaluation meant and how it worked.[4]

Another organization that has received high marks for its appreciation of the value of research and its ability to build an assessment culture is Sesame Workshop, the nonprofit educational organization responsible for producing one of television's longest-running and most successful programs, *Sesame Street*. The process at the Sesame Workshop is research based: Everything is driven by research. "Our longevity [at Sesame] has to do with listening to what hasn't been right . . . because if you really want to improve over time, you have to address what's wrong," remarked Charlotte Cole, senior vice president of global education at Sesame Workshop, who oversees research related to the effects of Sesame's international programs on educational outcomes.[5]

---

[2]   Author interview with Kavita Abraham Dowsing, May 23, 2013.

[3]   Author interview with Ronald Rice, May 9, 2013.

[4]   Jill Anne Chouinard, J. Bradley Cousins, and Swee C. Goh, "Case 7: United Way of Greater Toronto (UWGT)," in J. Bradley Cousins and Isabelle Bourgeois, eds., *Organizational Capacity to Do and Use Evaluation*, No. 141, Spring 2014, pp. 89–91.

[5]   Author interview with Charlotte Cole, May 29, 2013.

Evaluation Capacity Building

How do organizations get better at evaluation? This report has already examined numerous impediments to an organization's ability and willingness to conduct and use assessment properly. As with most changes, there are a million and one reasons to put it off until another day, or to let the next commander deal with it. One deliberate path specifically designed to help organizations achieve needed change in this area is evaluation capacity building (ECB).

ECB is "an intentional process to increase individual motivation, knowledge, and skills to enhance a group or organization's ability to conduct or use evaluation."[6] In her article "Some Underexamined Aspects of Evaluation Capacity Building," Laura C. Leviton raises important questions, such as, "What is the value of evaluation for organizations?" and, for evaluators, "When ECB is low is that because of organizational capacity limitations or evaluator limitations in knowing how to enhance ECB in organizations?"[7] One potentially useful model in this area is Getting To Outcomes (GTO), a collaborative effort between researchers at RAND and the University of South Carolina. GTO is a results-based approach to accountability and involves asking and answering the following "accountability questions," which serve as steps in the model:[8]

- assessing needs and resources
- setting goals and desired outcomes
- selecting evidence-based (or promising) practices
- assessing fit
- assessing individual/organizational/community capacity for an innovation
- planning
- implementation and process evaluation
- outcome evaluation
- continuous quality improvement
- sustainability.

Don't Fear Bad News

Valuing assessment requires getting over the fear of the results. When individuals and organizations are anticipating bad news, natural reactions run the gamut from avoidance to postponement to deflection (especially when blame is attached to the bad

---

[6]   Abraham Wandersman, "Moving Forward with the Science and Practice of Evaluation Capacity Building (ECB): The Why, How, What, and Outcomes of ECB," *American Journal of Evaluation*, Vol. 35, No. 1, March 2014b, p. 87.

[7]   Laura C. Leviton, "Some Underexamined Aspects of Evaluation Capacity Building," *American Journal of Evaluation*, Vol. 35, No. 1, March 2014.

[8]   See, for example, Abraham Wandersman, "Getting To Outcomes: An Evaluation Capacity Building Example of Rationale, Science, and Practice," *American Journal of Evaluation*, Vol. 35, No. 1, March 2014a.

news). All organizations—even the most transparent—cringe at least a bit when their daily activities are placed under a microscope. However, part of developing an assessment culture is being more accepting of bad news and welcoming it as an opportunity to improve and learn.

Also important when building an assessment culture is learning to live with bad news or, at least, what might seem to be bad news at first blush. The simple fact is that evaluations, when properly executed, can make people uncomfortable because they find and describe failures and contrast them with successes with arguments that explain both the *how* and the *why*. At a baseline level, for improvement-oriented assessments to have value, stakeholders need to trust the assessment and believe in its value. "This means being able to stomach bad news or contrasting viewpoints," says Steve Booth-Butterfield, a recognized expert on persuasion.[9]

This brings us to yet another contributor to an assessment culture: fostering an environment in which people are held accountable when they do a poor job. This means empowering all individuals within an organization—from the leadership to subordinates at the lowest levels—to speak with candor and to do so without fearing retribution. Only by identifying failures and learning from them can organizations, and the evaluations they conduct, refine and improve while incorporating lessons learned, even (or especially) lessons learned from failure.

Promoting Top-to-Bottom Support for Assessment

As noted in Chapter Three, assessment is not free; it requires resources. One of the most serious impediments to conducting proper assessments is the need for resources. All businesses and organizations operate in constrained environments (some more than others) and, therefore, are forced to allocate resources judiciously. Assessment does not always make the cut. Changing this prioritization requires galvanizing top-to-bottom support and buy-in, engaging leadership and stakeholders, and overcoming a distrust of assessment, whether that distrust is inherent or learned over time.

Garnering top-to-bottom support for assessment and getting the necessary buy-in, especially for assessment of a communication campaign, means working to ensure that all relevant stakeholders agree on the key performance indicators (or metrics) the evaluation will assess, says Sam Huxley, the senior vice president at the public relations and marketing agency FleishmanHillard.[10] But securing bidirectional feedback is easier said than done. In some sectors, particularly in national security and homeland

---

[9]   Author interview with Steve Booth-Butterfield, January 7, 2013. See also Steve Booth-Butterfield, "Standard Model," *Persuasion Blog*, undated; this is the overview page with three linked pages that detail the Standard Model.

[10]   Author interview with Sam Huxley, May 9, 2013.

defense, the effectiveness of assessments can be difficult to track because it is dependent on stakeholders providing feedback on how they used the assessments, which they do not always do.[11]

Secure Both Top-Down and Bottom-Up Buy-In

Given how we have described the process so far, the tension is apparent: User feedback is essential to improving assessments, but it is not something analysts regularly receive. Further compounding this issue, according to one SME, is that even when stakeholders specifically request analysis, they have their own missions to fulfill and regularly fail to close the loop with analysts as to whether the information provided was useful and what decisions were made based on the analytic product. This can become an even more profound problem: When the production initiated has no guaranteed consumer, those who actually conduct the assessment never learn if what they produced is even being used at all.

Providing feedback about an analytic product is the responsibility of the end user, and there is little that analysts can do to ensure that they receive this feedback. One possible solution is for high-ranking authorities to make the feedback cycle compulsory. This would greatly help analysts know whether their products are appropriate and what they can do to make them even more useful.[12]

Encourage Participatory Evaluation and Promote Research Throughout the Organization

Another way to improve buy-in from program designers and stakeholders is through participatory evaluation. Julia Coffman encourages the use of participatory evaluation to increase buy-in and improve relationships between evaluators and program designers. This approach entails involving a program's creators in the research design process in a reasonable way and constantly examining and reexamining ways to engage stakeholders and ensure their buy-in in the process. Stakeholder input is invaluable because it can help shape the big questions framing the evaluation.

Moreover, participatory evaluation helps evaluators *as well as* planners. For example, program designers often collect data on attitudes at the beginning of a study, but these data are not always useful to the evaluator, especially if they do not capture conditions that will change after program implementation. According to Coffman, "If you involve yourself early in the program design stage, you can shape their formative data collection strategy so that it can be used as baseline data for the summa-

---

[11] Author interview with Elizabeth Ballard, April 18, 2013.

[12] Author interview with Amy Stolnis, May 1, 2013.

tive evaluation."[13] One of the primary advantages to formative research is that it helps demonstrate the value of research to the stakeholder.[14]

Engage Leadership and Stakeholders

Leadership and stakeholder support is essential for instilling a culture that supports research. The aforementioned success of the Sesame Workshop is a great example of cultural transformation in this sense. Edward L. Palmer, the former vice president of research at Children's Television Workshop and *Sesame Street*, explained that everything that was aired went through a rigorous formative evaluation.

Assessment can be personality dependent. This sentiment was echoed in our interviews by military personnel tasked with conducting assessment; they found that charismatic leaders could make a huge difference.[15] Without strong leadership support, the whole process can become diluted and easily sidetracked.[16]

Although it is important to educate leaders on the importance and value of assessment, it is equally important to realize that different leaders will have varying degrees of interest. Amplifying this challenge is assessment that is ad hoc, hasty, or "done on the fly." This speaks not only to budget constraints but also to misplaced priorities.[17]

Every assessment stakeholder will have his or her own perspective on how things should be done, as there is no standard operating procedure or widespread agreement on how to evaluate the effectiveness of communication, whether in marketing, advertising, journalism, the military, government, or academia or at the individual or group level. This has led to what Booth-Butterfield has named the "Tower of Babel" problem, in which everyone has as individual language for addressing a particular problem. He believes that even with an agreed-upon framework and process for evaluation, different expert evaluators will approach certain problems from very different perspectives.[18] Therefore, assessment frameworks should be flexible enough to adapt to the personality or needs of different leaders or commanders.[19]

Explain the Value of Research to Leaders and Stakeholders

Because of their key role in shaping organizational culture, it is important to explain the value of research to leaders and stakeholders without assuming that these key play-

---

[13] Author interview with Julia Coffman, May 7, 2013.

[14] Author interview with Julia Coffman, May 7, 2013.

[15] Author interview on a not-for-attribution basis, December 15, 2013.

[16] Author interview with LTC Scott Nelson, October 10, 2013.

[17] Author interview with Amanda Snyder, March 2013.

[18] Author interview with Steve Booth-Butterfield, January 7, 2013.

[19] Military Operations Research Society, *Assessments of Multinational Operations: From Analysis to Doctrine and Policy*, proceedings of the Military Operations Research Society Conference special meeting, MacDill Air Force Base, Tampa, Fla., November 5–8, 2012.

ers have a firm understanding of its importance. Some may lack information, while others may incorrectly intuit that specific activities (including both kinetic and non-kinetic activities) will drive desired behavior change. In reality, behaviors are not easy to change. Thomas Valente described one technique for convincing people of the value of research: pointing out recent and sensational failures. There are many examples of products and advertising campaigns being launched with insufficient research or focus groups, leading to expensive (but often amusing) mistakes. In the late 1990s, BIC, known primarily as a producer of disposable pens, razors, and lighters, twice tried its hand at products not typically associated with disposability: underwear and perfume. Both brand extensions failed quickly (though the company's website states that BIC perfume is still being manufactured and sold in a few markets, including Iran).

At BBC Media Action, researchers make a point to demonstrate the value of research—through temperature maps that allow creators to understand what is needed where.[20] For IIP efforts, strong anecdotes illustrating, for example, adversary awareness and concern about ongoing messaging efforts can be a potent demonstration of the effectiveness of a campaign. (See Chapter Eleven for a more complete discussion of the presentation and uses of assessment for decisionmaking.)

As mentioned in Chapter Three, not all assessments need to achieve the same quality and depth. Some IIP efforts are so small that formal assessment would be unreasonably costly by comparison. Where an effort has a fully validated theory of change, less assessment is necessary. Where multiple efforts are extremely similar, one effort might receive full assessment scrutiny while the others receive much less.

Finally, a balance must be struck between performing activities and assessing them; assessment must not consume all of the resources, nor should it be completely ignored. As stated in the introduction to this chapter, as a general rule of thumb, approximately 5 percent of program resources should be dedicated to assessment.

Foster a Willingness to Learn from Assessment

Leadership is an indispensable ingredient to building an assessment culture. The qualities of the right leader include intellectual curiosity, a willingness to take risks (within reason), appreciation for a team mentality, and genuine trust of subordinates. Shaping a learning organization means doing more than simply going through the motions. According to LTC Scott Nelson, who previously served as the chief of influence assessment at USNORTHCOM, building an assessment culture requires "a lot of team participation in the process, and there needs to be support and trust to do the right thing, and not micromanaging." Furthermore, he added, "leaders have to be willing to take risks and can't be scared to get out of their office."[21]

---

[20] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[21] Author interview with LTC Scott Nelson, October 10, 2013.

Preserving Integrity, Accountability, and Transparency in Assessment

Ensure greater transparency is one of those buzz phrases (like *implement the rule of law*) that is oft repeated yet seldom understood. (Or, at the very least, the difficulty of achieving this feat is rarely acknowledged in the same sentence.) With regard to assessment, a lack of transparency can inhibit accountability and collaboration where it is needed the most. When determining who should do the assessing (internal versus external evaluators), it is critical to recognize different assessment roles. The same individuals or organizational levels can play multiple roles, but the data collector and assessor roles should be separate from those of the validator, integrator, and recommender.

For organizations conducting measurement and evaluation, there is virtually no incentive for sharing—it's a business. Moreover, according to Professor Maureen Taylor, the author of the 2010 paper "Methods of Evaluating Media Interventions in Conflict Countries," "Transparency is also hampered by clients that attempt to obfuscate critical findings to insulate themselves from public critique. This is done for myriad reasons, chief among them fear that their program will get cut." She suggests a policy to make data and results public whenever possible as a way of increasing both transparency and accountability.[22]

Sometimes, if assessment occurs at all, the teams designing the message are not given access to results. In the defense sector, classification issues are sometimes responsible for this disconnect, because contractors without the proper clearances may not be able to access assessment results pertaining to their own efforts, says Victoria Romero, a senior scientist in the Cognitive Systems Division at Charles River Analytics, a firm that applies computational intelligence technologies to develop mission-relevant tools and solutions to transform data into knowledge that drives accurate assessment and robust decisionmaking.[23] Such restrictions make it difficult to implement improvements to an effort's design. But as discussed in Chapter Eleven, assessment results have greater value when their presentation is tailored to specific audiences. Perhaps uncleared contractors cannot know what went wrong or why, but they could benefit from guidance on modifications nonetheless.

Although it may seem intuitive, it is crucial for users of assessments to explain why specific data are important and what they will be used for. In other words, put a *why* with the *what*. According to Stephen Downes-Martin, a professor at the U.S. Naval War College, it is a matter of asking the right question in the right way: "If you ask for an explanation, an account, a reason, something connected to a hypothesis of a theory of change, you'll do better."[24]

---

[22] Author interview with Maureen Taylor, April 4, 2013.

[23] Author interview with Victoria Romero, June 24, 2013.

[24] Author interview with Stephen Downes-Martin, February 12, 2013.

In an ideal world, *all* evaluation data and results would be transparent and widely available, not just the data that tell a nice story. Making available the mistakes or successes of previous efforts to everyone involved in IIP planning and assessment can go a long way toward ensuring more-effective future efforts and can avoid a case of reinventing the wheel, as well as highlight the value of assessment.

In-House Versus Outsourced Assessment

Conducting in-house research and evaluation or having this analysis outsourced is a contentious topic. In *Evaluation: A Systematic Approach*, Rossi, Lipsey, and Freeman discuss the "corruptibility of indicators," which refers to the natural tendency of those whose performance is being evaluated to fudge and pad the indicator whenever possible to make their performance look better than it is. It is usually best for such information to be collected by individuals who are independent of the program or effort. If it is to be collected internally by program staff, it is especially important to carefully follow transparent procedures so that the results can be verified.[25] With in-house evaluations, reports can be colored by funding concerns or the bias of the report writers.[26]

In military circles, there is also the challenge of overoptimism. An organization can avoid overoptimism by deliberately establishing an adversarial process, using devil's advocacy. This process can identify and examine all the ways in which things can go wrong, so that strictly positive information does not dominate assessment or reporting.[27] According to a military conference attendee, "One negative consequence of staff ownership of the assessment is the reluctance of the staff to assess themselves critically and negatively." As an assessment passes through each review step, "the bad stuff gets watered down, justified or removed completely."[28]

Some argue that those conducting self-assessments are likely to be more rigorous in their approach because they have a strong incentive to improve. Since improvement requires knowing what works and what does not, so the argument goes, external assessors are unlikely to understand the complexity of the environment or, possibly, the objectives of the effort. The prevailing wisdom is that self-assessment is fine when it comes to improvement-oriented assessment, but external evaluators are likely needed for accountability-oriented assessment.[29]

---

[25] Rossi, Lipsey, and Freeman, 2004, p. 227.

[26] Author interview with Steve Booth-Butterfield, January 7, 2013; Booth-Butterfield, undated.

[27] Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken: What Is to Be Done?" *Naval War College Review*, Vol. 64, No. 4, Fall 2011.

[28] Military Operations Research Society, 2012.

[29] Author interview with Paul Bell, May 15, 2013.

Tension Between Collaboration and Independence: The Intellectual Firewall

A major challenge to internal assessment is that, at least in the military, personnel are considering their career trajectories while assessing the outcome of their efforts, which can create incentives that bias results.[30] The obvious solution is external evaluators, but, as mentioned in the previous section, outsourcing assessment runs the risk of establishing a divide that is too robust, preventing planners and evaluators from achieving a shared understanding and discouraging collaboration. There is thus a need to balance the integrity of the research process with the need for cooperation between planners and evaluators.

Part of building an organization that values research is ensuring that the evaluation and planning sides can work together with minimal friction. The private sector uses the term *market research* to describe evaluation that helps improve a product, while in the nonprofit and public sectors, this function is sometimes pejoratively labeled *auditing* or *monitoring*. One solution is to hold planners accountable for success according to the very metrics they help design, to bring planners into the evaluation and research process, and to demonstrate the value of research to internal stakeholders. Planners and program designers need to involve the research team in program design, which can facilitate built-in markers of success that can then be tracked over time. Similarly, researchers need to include planners in the design of evaluations and measures to help ensure buy-in.[31]

While bringing researchers and program planners closer together can foster greater collaboration and lead to more-rigorous and more-comprehensive metrics, it is also important to maintain the intellectual firewall or some modicum of separation between those implementing an effort and those evaluating it. This is a delicate balancing act, says Devra Moehler, an assistant professor at the University of Pennsylvania's Annenberg School for Communication, who believes that while some level of separation is necessary, "there can't be too strong of a firewall because then the high-quality research—where the two need to work together to function—won't be able to happen."[32]

Assessment Time Horizons, Continuity, and Accountability

For assessment to capture long-term effects, there need to be lengthier time horizons.[33] Matthew Warshaw, managing director of the Afghan Center for Socio-Economic and Opinion Research (ACSOR), which runs the quarterly Afghan Nationwide Quarterly

---

[30] Author interview with a former employee of a large IO evaluation contractor, February 25, 2013.

[31] Author interview with Gerry Power, April 10, 2013.

[32] Author interview with Devra Moehler, May 31, 2013.

[33] Author interview with Mark Helmke, May 6, 2013.

Assessment Research (ANQAR) survey, believes that the battle rhythm of a combat environment constrains the quality of analysis, which would benefit significantly from a longer-term outlook. Demand signals are short-term because they have short-term goals and objectives:

> We need to think long-term, but that's really difficult. Do you want to be contracted to a single firm for five years to help figure that out? But the truth is that you need to do it—because you're going to be spending that money over the five years anyway, so they need to build good long-term teaming partnerships with serious practitioners and serious modelers. People must be willing to think beyond their own personal deployment so that when they hand this off, there is something enduring here.[34]

A common challenge in DoD activities is a timeline for results that is far too short for the objective. Steve Corman of the Arizona State University Center for Strategic Communication explained that an IO campaign might take months or years to have an effect. In this case, a three-month timeline for demonstrating an effect would not be long enough to gain a full understanding of an effort's impact.[35] Albany Associates chief operating officer Simon Haselock agreed, citing the "need to set realistic goals that can be achieved within the time span of the project. . . . Often, the sorts of effects that clients want to see are greatly disproportionate to the time available to deliver those effects."[36]

Outcome and impact evaluation is focused in short-term outputs, which limits the overall focus to short-term issues. The British Council, the UK's international cultural relations and education organization, used to work on five-, ten-, and 15-year time frames. But in the last few years, there has been increasing pressure to show results in the nearer term, which means that more-frequent analysis, in the form of quarterly or annual reports, has become more important.[37] It would be useful to look back at what happened 15 years ago and assess its usefulness today, but this is both difficult and expensive.[38] There are unmet expectations for behaviors to change quickly, without a realization that behavior change requires time. Targeted interventions over a longer period are more influential when an effort is trying to achieve a long-term behavior change, but this method is costly and time-consuming, and therefore could be prohibitive for DoD.[39]

---

[34] Author interview with Matthew Warshaw, February 25, 2013.

[35] Author interview with Steve Corman, March 2013.

[36] Author interview with Simon Haselock, June 2013.

[37] Author interview with James Pamment, May 24, 2013.

[38] Author interview with James Pamment, May 24, 2013.

[39] Author interview with Joie Acosta, March 20, 2013.

To the extent possible, program implementers and stakeholders (including funders and other users of assessment data) must be patient and not expect to see immediate results with vast social changes.[40] The lengthy timelines of IIP activities can confound efforts to detect immediate results; if a program seeks to change attitudes among a selected subpopulation over the course of a year, how can program personnel tell whether they are making good progress after three months, and how can they be certain that observed changes are due to their efforts rather than other influences in the IE?

Challenges to Continuity: Rotations and Turnover

Rotations—of personnel and commanders—pose a challenge to all types of operations, but they can be particularly problematic for long-term IIP efforts. Changes in command and program staff can affect relationships, processes, and prioritization at all levels of a campaign. A new commander may bring new priorities and approaches, leading to cascading changes.[41] Changes affecting the conduct and assessment of IIP efforts already under way can lead to setbacks and even unanticipated failure. As a result, leadership—and others driving the design of assessments—may need to be willing to inherit assessment practices that are "good enough."[42]

Rotations can cause problems when new personnel fall in on programs or assessments they do not understand: "It can be hard to fall in and assess things you didn't start, especially if the past effort doesn't have clear indicators and the logic behind them spelled out."[43] This is also another good reason to be explicit about an effort's theory of change/logic of the effort. Short tours can also confound good assessment because there is little incentive to engage in practices that will not pay off immediately. After all, a future rotation could dismantle that process anyway.[44]

Turnover is not strictly a problem internal to DoD, either. One SME with whom we spoke was part of a military information support team that had contracted with a local university to perform a baseline assessment. Although both sides understood the importance of the baseline assessment, the timeline was such that the civilians' involvement ended before the assessment results were ready for analysis.[45]

---

[40] Author interview with Larry Bye, June 19, 2013.

[41] Author interview with Stephen Downes-Martin, February 12, 2013.

[42] Author interview on a not-for-attribution basis, December 15, 2013.

[43] Author interview on a not-for-attribution basis, December 5, 2012.

[44] Author interview on a not-for-attribution basis, January 23, 2013.

[45] Author interview on a not-for-attribution basis, January 23, 2013.

Improving Continuity: Spreading Accountability Across Rotations

Gaby van den Berg, head of SCL Training and Infrastructure, offered a range of possible solutions to the problem of continuity across rotations, including evaluation training for every rotation, external assistance with transition, making strategic communications a career option,[46] longer rotations, bringing in social scientists, devoting more time to doctrine, incorporating technology, engaging in evaluation from the outset, and establishing independent audits to validate assessments.[47]

Reachback should be a way to improve continuity, since there should not be rotation limits on reachback support. Another option might be to have personnel serve in reachback roles before or after their deployed rotations. With all the rotations affecting IIP efforts, data sources could change, individuals who were regularly sending data could rotate out and stop communicating, and questionnaires and even classifications could change.[48] Still, eschewing intermediate assessments will only make it more difficult to gauge what is working and what is not working.[49]

For evaluations to account for longer-term impacts, there need to be institutional incentives to care about the long-term. A RAND study on developing a prototype handbook for monitoring and evaluating humanitarian assistance efforts may be instructive when handing off projects, as it focuses on the importance of ensuring continuity each time a staff turns over. That study included the following suggestions:

- For ongoing projects, it is critical to provide information on the status of activities (e.g., How far along is the project?); summarize required resources to complete project assessments (e.g., How much time, manpower, or equipment will be needed to complete the assessment process?); discuss key collaborators and important community contacts (e.g., Has the project assessment successor been introduced to key contacts?); explain the project assessment plan (e.g., Have the important indicators and methods for collecting data been reviewed and explained, along with the timeline?); and hand over all indicator worksheets and other relevant project and project assessment documentation.
- For already completed projects, the goal should be to review these same points (e.g., Does the successor know exactly where to find data and reports for the completed project?) and explain what is required for the one-year follow-up and discuss the follow-up assessment plan (e.g., Does the successor know what is involved, including when the assessment should be conducted, which MOE indi-

---

[46] This is UK-specific; the United States has a PSYOP branch, but it still needs to make evaluation part of the career track.

[47] Author interview with Gaby van den Berg, April 22, 2013.

[48] Author interview with John-Paul Gravelines, June 13, 2013.

[49] Author interview on a not-for-attribution basis, July 30, 2013.

cators need to be collected and recorded, and how they can be compared with earlier measures of the same MOEs?).[50]

Longer Assessment Timelines, Continuous Measures, and Periodicity of Assessment
When conducting assessment, it can be difficult to maintain patience and perspective, especially in a culture that thrives on quick and reliable results. Effective assessment takes time. Periodicity should be based on how quickly the situation on the ground changes. It does not make sense to conduct several assessments within a given time frame when little has changed during the period in question.[51] Furthermore, "it's hard to measure impact when the programs are long-term and there are many intervening variables that might provide an explanation for an outcome." This challenge should not serve as a "cover for not doing the measurement that needs to be done," warns Katherine Brown, a former public affairs officer and an SME on media in Afghanistan and Pakistan. On the contrary, the real need is for better measures to capture long-term effects. Only by setting realistic goals can this be achieved within the time span of the project.[52]

External validation can provide a check against cheating and other quality issues by looking at data over time and how those data track with events. This is one of the reasons why it is so important to have long-term measures of impact and long-term evaluations of progress over time. Brown admires the work of the Asia Foundation in this regard, noting, "They've done one study for the past eight years. So if something was really awry one year, it would be very obvious. Like everything in this country [Afghanistan], there needs to be long-term investment and long-term attention to what is happening."[53]

Preserving Integrity, Accountability, and Transparency in Data Collection

As discussed in the section "Preserving Integrity, Accountability, and Transparency in Assessment," making data transparent and as widely available as possible can boost the integrity of an assessment and improve collaboration both within and across IIP efforts.[54]

Among marketing firms, the end user of the assessment is almost always exclusively the client. Such assessments are rarely made available to the public because they

---

[50] Haims et al., 2011, p. 57.

[51] Military Operations Research Society, 2012.

[52] Author interview with Simon Haselock, June 2013.

[53] Author interview with Katherine Brown, March 4, 2013.

[54] Author interview with Thomas Valente, June 18, 2013.

could be exploited by competitors. There are distinct parallels to IIP assessment in this approach. While there are benefits to sharing data and assessment results to guide program improvements or to inform similar efforts, there are also security risks in sharing this information too widely. That said, certain types of data collection must be conducted by local individuals or firms, and involving local populations in an IIP effort can benefit the effort itself.

Cultivating Local Research Capacity

There are myriad considerations to take into account when hiring local research firms, which can be valuable for both formative and summative evaluations. "It's essential to involve local people in the development of messaging and in the design of summative research, survey instruments, etc. I cannot emphasize that enough," says Charlotte Cole at Sesame Workshop.[55] Often, different vendors have different strengths, which could mean hiring one vendor to do data collection and another to do analysis.[56]

In Afghanistan, all staff administering surveys must be local Afghans. It can be difficult to identify locals who not only can read but can do so out loud. Between 400 and 500 field workers are involved in each ANQAR wave. Additionally, there are 40–50 keypunchers for data entry and 15–20 full-time ACSOR staff in Afghanistan, plus another ten based in Virginia. The survey takes about an hour to administer.[57] For its 2010 study of Afghan media and public perceptions, the U.S. Agency for International Development (USAID) contracted Altai Consulting, a local firm, for data collection.[58]

Similarly, the BBC hires and trains in-country local researchers to work and conduct research in its country offices; these researchers are not contracted but, rather, work for the BBC and are citizens of the countries in which the BBC operates. The country director is often a British citizen, but the research team is populated by citizens of the host nation. For example, in Nigeria, the BBC has a research director and 14 in-country researchers. Here, it is important to note that *some work is contracted*. For significant amounts of quantitative work (e.g., a nationally representative survey in Nigeria), the BBC will hire a research agency, but it has a local research team in every country in which it offers programming.[59]

---

[55] Author interview with Charlotte Cole, May 29, 2013.

[56] Author interview with Charlotte Cole, May 29, 2013.

[57] Author interview with Matthew Warshaw, February 25, 2013.

[58] Author interview with Emmanuel de Dinechin, May 16, 2013.

[59] Author interview with Kavita Abraham Dowsing, May 23, 2013; BBC Media Action conducts research in Africa, Asia, Europe and the Caucasus, and the Middle East and North Africa. Specifically, it covers nine countries in Africa (Angola, Ethiopia, Kenya, Nigeria, Sierra Leone, Somalia, South Sudan, Tanzania, and Zambia), five countries in Asia (Afghanistan, Bangladesh, India, Nepal, and Pakistan), two countries in Europe and the

Coming into any country or any situation as an external actor presents a host of difficulties for the party attempting to build local capacity. These difficulties are exacerbated when there is significant disparity in capabilities or expectations between the parties involved.

While it is essential to hire local researchers, they are often poorly trained, which threatens the integrity of the research design and the overall quality of the work. Challenges with hiring local researchers include payment issues, intercoder reliability, and differing research standards. In the words of one SME, "Our obsession with metrics doesn't always translate."[60] Often, local firms know how to talk about research but are not able to do it.[61] This means that they can win the contract but not deliver the end product. We also heard anecdotes about local researchers accepting payment from multiple customers for the same survey.[62]

In U.S.-supported efforts, there is often tension between the desire to build local research capacity and the desire for the highest-quality research standards. Cheating by local research firms or individual local researchers is an issue: About 10 percent of surveys in Afghanistan have to be redone.[63] Sometimes, field workers fill out the questionnaires themselves or have their family and friends fill them out; other times, they simply do not administer the survey.[64]

Hiring local researchers comes with a need to monitor and train them in order to build local capacity. This could mean creating research capacity in environments where it did not previously exist. Matthew Warshaw of ACSOR recalled helping to create a firm in Bosnia right after the Dayton Agreement in the mid-1990s. The idea was to start a research company that could provide services to a variety of clients, fill gaps in the market, and offer an enduring capability after the international presence in the country declined. "Bosnia was a success," said Warshaw. "The company is still there. It's not as big as it was, but it's doing a broad variety of commercial research: population surveys, commercial studies, product testing, focus groups, monitoring and evaluation work, media ratings (Nielsen sort of stuff). We try to develop as broad of a research capacity as we can. If you only do one thing and don't diversify, it can fall apart."[65]

Caucasus (Georgia and Serbia), and six countries in the Middle East and North Africa (Algeria, Egypt, Iraq, Lebanon, Palestinian Territories, and Tunisia).

[60] Author interview with Amelia Arsenault, February 14, 2013.

[61] Author interview with Emmanuel de Dinechin, May 16, 2013.

[62] Author interview with Kim Andrew Elliot, February 25, 2013.

[63] Author interview with Emmanuel de Dinechin, May 16, 2013.

[64] Author interview with Katherine Brown, March 4, 2013.

[65] Author interview with Matthew Warshaw, February 25, 2013.

BBC Media Action uses a "mentoring model" of training local researchers, and this may be an effective model for DoD and USAID efforts. A key advantage of the mentoring model is that it obviates the need for intensive monitoring. Capacity building is done in two ways: first, through an annual research workshop where local researchers receive training and, second, through country research managers. At the BBC, every in-country researcher has a dedicated point of contact in London (the country research manager), who offers training and advice and shares methodologies. The country research managers visit two to three times a year to do side-by-side training and mentorship.[66] Moreover, the BBC is looking into options for collaborating with other organizations in their capacity-building conferences and workshops. If they do open up the conferences, DoD may want to encourage local researchers to attend.[67]

There is a need for more-rigorous supervision, oversight, and training of local researchers to prevent unusable data. In this sense, the front-end investment in building research capacity is worth it because it saves the costs associated with salvaging the data when research is poorly conducted.[68]

Because operating in a foreign environment is inherently complex, the situation on the ground is not always as clear as it first seems. On paper, some local firms may seem legitimate, but ultimately they might not be qualified to conduct evaluations. Some modicum of a vetting process is necessary before making decisions about where to allocate finite resources. The amount of waste, fraud, "ghost employees," and no-show jobs in developing countries can sap an effort's momentum and distort the assessment process.

Though it should go without saying, there is a dire need to thoroughly investigate local firms before awarding contracts. The pressure to award contracts to the lowest bidder invites quality problems—creating a proliferation of communication companies with no performance history and no incentive to establish themselves or build legitimate research capacity. There is good reason to be skeptical of small research firms that appear seemingly out of nowhere; it is very difficult and takes many years to establish a research firm in Afghanistan.[69]

The Local Survey Research Marketplace
At first glance, the survey research marketplace seems a bit crowded. There are a great many organizations trying to conduct surveys in conflict environments, especially in Afghanistan. These surveys come at a high cost for program sponsors, and the quality of the survey design is often poor because it does not take into account local sensitivities, culture, and social conditions (just as the quality of a survey's administration can

---

[66] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[67] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[68] Author interview with Katherine Brown, March 4, 2013.

[69] Author interview with Matthew Warshaw, February 25, 2013.

suffer from the shortcomings of local research firms). Different cultures have different attitudes about the politeness of answering questions negatively. Survey fatigue is also a critical issue. There is a need to redirect investments toward survey quality over quantity.[70]

Among the large companies currently running surveys in Afghanistan are ACSOR/D3 Systems, EUREKA, ORCA, Gallup, and Altai, which are joined by a smattering of small research firms. ACSOR runs approximately ten surveys per year, some with quarterly data collection. ACSOR alone interviews approximately 500,000 Afghans annually.[71]

There are many disadvantages to having multiple firms and several associated advantages to consolidation, including eliminating redundancy, poor standardization, and survey fatigue. Inefficiencies and redundancy are often mentioned as concerns—in many instances the same information is being collected more than once on different surveys. There are challenges from a government client perspective, associated with having multiple surveys that are not standardized or where the standards are not being enforced. This could be improved if standards were fleshed out and better enforced.[72] "One of the major issues resulting from the large number of surveys and survey firms in conflict environments is 'survey fatigue'—respondents are surveyed too often, which adversely affects response rates and can create response bias," according to Amelia Arsenault, an assistant professor at Georgia State University who a focuses on the evaluation of media interventions in conflict environments.[73] Further, if a questionnaire grows unwieldy, there is a greater likelihood of survey fatigue on the part of the respondent and interviewer.[74]

On the flip side, there are advantages to multiple firms and disadvantages to consolidation. Warshaw does not perceive a need for consolidation among surveys and local survey research firms. First, he believes that having multiple surveys and multiple firms is good from a competition perspective because it drives down prices and improves quality: "You wouldn't want a monopoly." Second, he argues that multiple surveys are needed because each serves a different purpose, and some surveys have a different sample frame because they are concerned with a different target audience or local population: "USAID and IO organizations are looking for very different things in surveys. As another example, human terrain system surveys are looking at smaller samples—more local in orientation."[75]

---

[70] Author interview with Amelia Arsenault, February 14, 2013.

[71] Author interview with Matthew Warshaw, February 25, 2013.

[72] Author interview with Matthew Warshaw, February 25, 2013.

[73] Author interview with Amelia Arsenault, February 14, 2013.

[74] Author interview with Kim Andrew Elliot, February 25, 2013.

[75] Author interview with Matthew Warshaw, February 25, 2013.

Organizing for Assessment Within DoD

DoD IIP efforts should be broadly integrated into DoD processes, and IIP assessment should be integrated with broader DoD assessment practices.[76] As discussed in Chapter Two, the assessment of kinetic activities involves shortcuts, heuristics, and "taken-for-granteds." Assessment for IIP lacks this shared understanding, so it requires explicit steps and assumptions. In this section, we offer guidance on how to overcome organizational challenges to planning, conducting, and assessing IIP efforts in DoD.

Mission Analysis: Where a Theory of Change/Logic of the Effort Should Become Explicit

Assessment starts in planning, and the assessment process should be organizationally embedded in or connected to the planning cell. The next step in JOPP after planning initiation is mission analysis, and planning for assessment should begin during that phase, when it is determined what will be accomplished and how to measure it. In this way, assessment can help determine progress toward accomplishing a task, creating an effect, or achieving an objective.[77] At this point in the process (mission analysis), a theory of change or logic of the effort should be made explicit, and if there are competing logics, that should also be made explicit in COA development (step 3). Specifically, the assessment plan built during the mission analysis phase will identify and take into account initial desired and undesired effects. This process continues through COA development and selection.

Differences Between Information Operations and Kinetic Operations

As noted in Chapter Two, IIP efforts differ from kinetic operations. However, the planning and decisionmaking processes for DoD IIP efforts are much the same as those for kinetic efforts, and this is a good thing, as it promotes commonalities across different kinds of military operations and encourages singular standardized processes for all operations. Still, it is important to be aware of the differences, and it is important for the processes to respect those differences.

The *Marine Corps Operating Concept for Information Operations* delineates four aspects of IO that depart from the kinetic world. First, fires do not have to compete for the attention of the intended target, something that information must do. Second, unlike in kinetic operations, the target of an information operation can choose what signals to heed or ignore through filters (both social and cultural). Third, as we have seen with the ubiquity of the Internet, the second- and third-order effects of information operations can multiply well beyond the designed radius of the intended target

---

[76] See Chapter One and Appendix C for a discussion of current doctrine related to IO and the assessment of these efforts.

[77] U.S. Joint Chiefs of Staff, 2011c, p. IV.

and lead to a host of unintended consequences. Finally, while damage wrought from kinetic operations is readily apparent in most cases, the latency of information is a result of the need for interpretation by the intended target, which progresses at a different pace.[78]

Within the IE, and particularly with respect to the cognitive dimension, intended effects can be challenging to measure, evaluate, and assess. Stephen Downes-Martin sums it up as follows:

> If you are going to assess tactics, that is easy, because we have 3,000 years of statistics on that. If you took a U.S. marine and he traveled through time and joined a Roman war camp in the AD, he'd fit right in. We have 3,000 years of statistics and the foundation of physics. In that situation, it is fine to use performance as a proxy for outcomes, because we know exactly what performance leads to exactly what outcomes. This does not work at higher levels, at the operational level, or even for the tactical level, when things are more complex.[79]

Traditional battle damage assessment and the associated (directly observable) effects are not relevant when it comes to IO. As a result, properly assessing the effects of IO requires the development of other measures, including feedback loops "to gauge the effectiveness of these activities."[80] Challenges are not limited to measurement and evaluation. Because of lacking shared understanding and intuition of IRCs across the joint force, there can be unmet expectations for behaviors to change quickly, without a realization that behavior change requires time.

To have good IIP assessments, it is critical to understand the target audience for an effort and the environment in which the effort is being conducted. At a rudimentary level, some stakeholders are looking for behavioral change much too quickly.[81] A challenge in designing survey questions for IO is that influence programs need to be discreet—the intended target of influence should not be too obvious. This matters for assessment design because too many questions about the end goal could reveal the true objective of the effort. This is one reason that measuring IO is more complicated than measuring the effects of kinetic operations.[82] Another question that needs to be addressed is how to incorporate functional assessment (e.g., IO) into overall campaigns or regional assessments.[83] This is discussed further in Chapter Five.

---

[78] U.S. Marine Corps, *Marine Corps Operating Concept for Information Operations*, Washington, D.C., February 4, 2013, p. 6.

[79] Author interview with Stephen Downes-Martin, February 12, 2013.

[80] U.S. Marine Corps, 2013, p. 13.

[81] Author interview on a not-for-attribution basis, July 30, 2013.

[82] Author interview with Matthew Warshaw, February 25, 2013.

[83] Military Operations Research Society, 2012.

The Need to Standardize and Routinize Processes for IIP Planning and Assessment

Before they can be assessed, there is a dire need to integrate IIP efforts more diffusely across the government; otherwise, there will be no way to infer causality or know "what is influencing what." As Brian Cullin, former senior adviser on intergovernmental affairs to the under secretary of state for public diplomacy and public affairs, has noted, "Evaluating IO separate from broader [strategic communication] engagement is invalid and disincentivizes integration. It's possible that a whole-of-government [strategic communication] assessment could serve as the centerpiece of the coordination effort."[84]

Not only do IIP assessments need to be integrated across the government, but they also need to be integrated with DoD assessments. As one SME noted, "We [in IO] could come up with lots of different ways to improve assessment, but if we aren't integrated with broader DoD assessment, we'll be in trouble."[85] The planning processes for IRCs and kinetic capabilities are the same in DoD, and the processes that create and execute IIP assessment should be integrated with standard and routine DoD assessment processes.

Overcoming a Legacy of Poor Assessment

While there are pockets of strong assessment practice throughout DoD and many individuals have learned to value assessment, a legacy of poor assessment has created a failure cycle for assessment in many elements of DoD. To break the cycle, assessment needs advocacy, (better) doctrine and training, trained personnel, and greater access to assessment and influence expertise.

Although assessment is traditionally not a DoD strength, there is an opportunity to improve efficiency by collaborating and making better use of the data that already exist and that are being collected. DoD needs to be more collaborative in its approach to measurement and leverage the work done by other agencies, nongovernmental organizations (NGOs), and international actors. The military has much to gain by learning from people on the ground with a better understanding of the media environment.[86] Overall, fostering closer cooperation between DoD and other agencies will require overcoming an aversion to cooperation and sharing, though it would also help avoid duplication and redundancy.[87]

In his assessment of why operational assessments fail, Jonathan Schroden asserts that to assess progress in a modern military operation properly, it is necessary to gather, analyze, and fuse information on the activities of the enemy ("red"), civilians ("white"), and friendly forces ("blue"), which the U.S. military is not well structured to achieve.

---

[84] Author interview with Brian Cullin, February 25, 2013.

[85] Author interview on a not-for-attribution basis, July 30, 2013.

[86] Author interview with Maureen Taylor, April 4, 2013.

[87] Author interview with Kim Andrew Elliot, February 25, 2013.

As things stand, intelligence organizations are threat focused and only secondarily interested in information pertaining to civilian activities, but some organizations are capable of fusing the two. While some operations analysts typically gather and analyze information about blue forces, there is no entity that currently specializes in fusing and analyzing information across the entire spectrum.[88]

Assessment is not traditionally a DoD strength, but this does not mean that it will not be in the future. Audience analysis in the Middle East is limited because there are no consolidated audience data you can buy off the shelf, and there is no Nielsen-like organization investing its own resources in media research from which viewership data can be acquired. This is partly due to insufficient demand on behalf of the advertisers, which have been known to purchase media without an attendant interest in sophisticated analysis. So, according to Emmanuel de Dinechin, founder and lead partner at Altai Consulting, "if advertisers want a media snapshot, individual clients have to fund media research projects from firms like Altai. This is inefficient and hurts the quality and scope of the research because resources are spread thin."[89]

This paradigm may change very soon, given trends in emerging markets. For example, firms are starting to send their best marketers to Africa. This will create a market for a sustained Nielsen-like presence. This shift also has implications for IO assessment: Instead of sponsoring their own media share studies, these programs should soon be able to just buy data from Nielsen (or another firm that is doing the consolidated analysis), making the planning and assessment of these efforts much more cost-effective.[90]

Schroden goes on to note that the problems with operational assessment run much deeper than poor metrics and are often organizational in nature. To be sure, there is a failure cycle at work. According to this view, the key challenges that should be addressed to improve assessment include identifying an advocate for assessments; fixing DoD planning and assessment doctrine so that it provides actual guidance on *how* to assess, not just vocabulary and definitions (e.g., the difference between MOP and MOE, which is interesting but not helpful operationally); creating a military occupational specialty and formal course of instruction for operational assessment; and shifting thinking away from strictly quantitative and picture-based assessment products and toward balanced, comprehensive, analytic narratives.[91]

---

[88] Schroden, 2011, p. 98.

[89] Author interview with Emmanuel de Dinechin, May 16, 2013.

[90] Author interview with Emmanuel de Dinechin, May 16, 2013.

[91] Schroden, 2011.

To Address Deficiencies in Doctrine, Guidance, and Tools, Think Beyond Measures of Performance and Effectiveness

As mentioned earlier, due to such issues as budget limitations and misplaced priorities, much assessment is done on the fly, so to speak.[92] There is a lack of institutionalization of the ideas and difficulty getting buy-in from those who control critical assets, because they might not appreciate the value of assessment efforts.[93]

Related to this is the need for better doctrine for assessment, given shortcomings in definitions and authorities, as well as in the understanding of basic assessment principles in existing assessment doctrine.[94] Doctrine should not be overly rigid and must consider the evolution of the assessment process.[95] One SME put it thusly, "There is a difference between joint doctrine and what General [Raymond] Odierno wants. Joint doctrine does not teach down to the tactical level."[96]

Operations assessments can fall short as a result of myriad deficiencies, contradictions, and confusion in the doctrine that is supposed to guide their conduct.[97] By focusing exclusively on MOPs and MOEs, DoD is imposing limitations that can preclude effective assessment processes. One interviewee suggested that, in DoD assessment, activities (MOPs) get measured and effects (MOEs) get measured, but the connecting logical changes, the measures of impact (MOIs) get missed.[98] Again, the theory of change/logic of the effort is instructive. Perhaps it is best understood as a complex system: How do activities affect the function, behavior, or attributes of objects in the system to produce an effect?

This disconnect is not unique to DoD. In the United Kingdom, current doctrine acknowledges the requirement for assessment, but, in practice, it has often been plagued by inconsistencies in application or considered an add-on to overall campaign analyses—thought of only at the tail end of the process. A review of assessment in the UK Ministry of Defence indicates that assessment is not conducted well for various reasons, including vague campaign objectives, a lack of realistic milestones to assess short-term progress, a failure to approach assessment as an activity, frameworks that are replaced with every rotation, unrealistic data requirements imposed on subordinates, and an overly mechanistic approach that ignores the operational context.[99] Some critics

---

[92] Author interview with Amanda Snyder, March 2013.

[93] Author interview with John-Paul Gravelines, June 13, 2013.

[94] Author interview with Jonathan Schroden, November 12, 2013.

[95] Military Operations Research Society, 2012.

[96] Author interview on a not-for-attribution basis, October 30, 2013.

[97] Schroden, 2011, p. 92.

[98] Author interview on a not-for-attribution basis, February 20, 2013.

[99]   UK Ministry of Defence, 2012.

have been even harsher: "Across the whole MOD [Ministry of Defence], assessment is poor to nonexistent."[100]

In both the United States and United Kingdom, guidance and doctrine on assessment is skewed toward frameworks and conceptual pieces (e.g., definitions, explanations of why assessment matters) and light on the *how* of assessment. Doctrine on assessment should explain *how* to do it, while going beyond the traditional constructs of MOEs and MOPs. These reports and doctrine would be improved by describing the specific measures and tools that should be employed to measure various constructs, and by mapping those tools to their proper application within the assessment hierarchy. So, for a baseline, is a survey or a poll the right application at the beginning of an assessment?[101]

As we will discuss in Chapter Six, there is more to measurement than the differences between MOPs and MOEs. Prevailing doctrine (JP 3-0 and JP 5-0) is strikingly vague in its discussion of operational assessment; more instruction on how to actually conduct assessment is clearly needed. Where current doctrine contains some discussion of assessment, it is mostly at the overview level, without a great deal of specific guidance. For example, JP 5-0, *Joint Operation Planning Process*, discusses the *what* and *why* of assessment, but the details of the *how* are mostly left to practitioners. The *Commander's Handbook for Assessment Planning and Execution* offers a practical method that commanders and staffs can use as a starting point to begin thinking about the *how* in assessing operations.[102]

The Challenges of Assessment in Conflict Environments Require Being Nimble and Responsive

Assessment requires being nimble and responsive—able to adapt an effort to accommodate constraints, barriers, disruptors, and unintended consequences. This is especially critical in a conflict environment like Afghanistan, but the only way this is possible is through a free-flowing and steady trickle of information.[103] The more nonpermissive the area becomes, the more the stakeholder wants access to the information. This leads to a significant amount of bad data.[104] There are definite limitations to the use of social science methods in combat and/or tribal environments.

Applying social science methods from the commercial and marketing world to complex combat environments is fraught with pitfalls. For example, it is impossible to conduct a random-digit-dialing survey in Afghanistan because many people do not have telephones. Thus, many of the measures that would be utilized in a marketing

---

[100] Author interview with a UK officer on a not-for-attribution basis, March 29, 2013.

[101] Author interview with LTC Scott Nelson, October 10, 2013.

[102] U.S. Joint Chiefs of Staff, 2011a; U.S. Joint Chiefs of Staff, 2011c.

[103] Author interview on a not-for-attribution basis, March 2013.

[104] Author interview on a not-for-attribution basis, March 2013.

paradigm are much less feasible in the complex operational environment. In this case, expectation management is valuable. Other challenges to network analysis in local or tribal environments include definitively identifying people who use colloquial names and determining how they are connected to each other.[105]

Building Expertise and Developing a Career Field Require a Fresh Approach to Assignment Patterns and Qualifications

A fresh approach to assignment patterns and qualifications, while not an easy task, is nevertheless necessary to recruit, train, and retain individuals capable of conducting assessments. Consider the following quote from a MISO SME:

> We have invested heavily in getting the right people with the right backgrounds together. There has been resistance to this because we are bucking the norm in terms of assignment patterns and qualifications. We are ruffling a few feathers. When we write our requirements for positions, we are very specific about education, training, and experience. For planners, in general, we look for someone with an advanced degree in a particular discipline. For MISO, we look for someone with an advanced degree in the behavioral or social sciences. For operational, we look for someone with a statistics degree. Someone with a general political science degree is not going to work without the right experience. I'd like to have economists or [School of Advanced Military Studies] grads. We have two—but we want more—individuals from the war colleges.[106]

Placing the best and brightest in the assessment process would signal follow-through on what, until now, has been perceived as mere lip service by assessment practitioners. If assessment is important, it needs people who are intellectually curious. IO assessment requires critical thinking and an intellectual curiosity, individuals who know what data they need and who have the right tools or the right logic model or theories of change to improve planning. This is the only way to ensure that assessment does not fail and, by extension, that the mission does not fail.[107]

Improve Training for Assessment: Develop a Training Pipeline

By creating a military occupational specialty and formal course of instruction for operational assessment, Schroden believes, DoD could develop a proper training pipeline for developing personnel who could provide training to others in the conduct of operational assessment. Staff officers placed in assessment billets and individuals formally trained in operations research and systems analysis (ORSA) are commonly selected for these duties, with scant training on the specifics of assessment. As a result, weighted-average roll-ups of metrics and stoplight charts are considered benchmarks. "In the

---

[105] Author interview with Ronald Rice, May 9, 2013.

[106] Author interview on a not-for-attribution basis, July 30, 2013.

[107] Author interview with LTC Scott Nelson, October 10, 2013.

absence of sound doctrine and training, we have left practitioners either to flounder on their own or to steal flawed products from others, both of which are recipes for failure," says Schroden.[108]

Confusion between the terms *operational assessment* and *analysis* (or *operations research*) has led many to believe that ORSAs are trained to conduct this type of assessment, but that is not necessarily the case.

Yet by tasking ORSAs with operational assessment, "we are unconsciously sacrificing our capability to conduct operations *analysis* (i.e., to optimize our performance)," which leads Schroden to conclude that both a formal course of instruction for operational assessment *and* a dedicated military occupational specialty are requirements for success.[109]

The ORSA community may not be the best fit or have the full set of skills for complex environments. "We need to access a broader set of skills than ORSA, and lots of creative/flexible thinkers and analysts," Schroden concludes.[110] There is a dire need for critical and creative thinking in the area of assessment.[111] ORSAs were handed this responsibility by default, since there is no training or career path for assessment, and it might be time to rethink that assignment.[112]

For some, the rather obvious point here is that "you have to engage more social scientists."[113] When asked about the most critical piece to improving assessment, John-Paul Gravelines, a strategic communication assessment specialist in Afghanistan, replied, "Training is the one area I'll go back to as being critical."[114] Finally, an SME who asked not be named stated, "We are not funded, manned, trained, or equipped to do assessments, period."[115]

Train Staff on How to Interpret Polling Data

It is essential to train IO assessors in social science so that they can read and interpret polling data and understand the application and limitation of those data. One respondent reported confusion that so much money was spent on Gallup and other polling organizations, while analysts were not sufficiently trained to interpret or apply the results of the polls. Doing so would require an understanding of concepts like sampling error. In Nelson's view, these are the types of tools to invest in and to use as a

---

[108] Schroden, 2011, p. 95.

[109] Schroden, 2011, p. 97.

[110] Author interview with Jonathan Schroden, November 12, 2013.

[111] Military Operations Research Society, 2012.

[112] Military Operations Research Society, 2012.

[113] Author interview with Andrew Parker, April 26, 2013.

[114] Author interview with John-Paul Gravelines, June 13, 2013.

[115] Author interview on a not-for-attribution basis, July 30, 2013.

basis for staff training. Assessment personnel do not have to be experts or specialists, but they do need an understanding of how to apply social science principles to the problems they are trying to address.[116]

Advocate for Assessment to Provide an Impetus for Change

Identifying an advocate for operations assessment is one important step in breaking the failure cycle. The lack of advocacy for assessment within DoD is troubling for several reasons. For example,

- Without an advocate, there will be no impetus for change.
- Just tweaking current doctrine rather than completely reconceptualizing the design and implementation of operational assessments will lead to shortcomings in terms of comprehensiveness and effectiveness.
- The lack of a center of gravity or knowledge repository will leave a dearth of established experts in operations assessment.

Instead, according to Schroden, "we will continue to cannibalize other military occupational specialties, most notably the ORSA pool, to conduct assessments."[117]

Find the Right Balance in Assessment Cell Organization

Some organizational issues are unique to DoD, including assessment cell organization. Ideally, DoD would balance the need for a dedicated assessment cell with the need for assessments to be integrated into routine operations.[118] A way to address these needs is to connect the assessment process organizationally or embed it within the planning process for a larger campaign or in the planning cell of a particular task force or other unit.[119]

The USNORTHCOM effects assessment team established an effective assessment process via the command's Influence Assessment Capability. The primary IO task at USNORTHCOM was building partner capacity for IO, and that was the focus of the effects assessment team. While the team consisted of a director, a deputy, two branch chiefs, a research staff, and an assessment analysis staff, the real power of the organization was in the data developers. They had the required methodological skills and tools and were able to rely on participant observation to actually conduct needed empirical research. This allowed them to describe what they saw in SME exchanges and ensured that their descriptions were focused on the specific things the command

---

[116] Author interview with LTC Scott Nelson, October 10, 2013.

[117] Schroden, 2011.

[118] Military Operations Research Society, 2012.

[119] Military Operations Research Society, 2012.

wanted to measure (e.g., the dynamics of the group, rank structure), generating rich, valuable data.[120]

Assessment and Intelligence

Assessment requires data to populate measures—and intelligence is potentially a good data source. Assessors need to better leverage available intelligence, and the intelligence community needs to better support IIP and IIP assessment. Currently, there is too little interaction between operations and intelligence, and this has led to assessment in a vacuum. What is required is an honest broker between the operations and intelligence communities, especially in the area of predictive assessments.[121] The intelligence community needs to be trained in how to support IO and IO assessment. Because it has not been trained in how to support IO efforts, IO assessment teams keep getting asked to do their own intelligence preparation of the battlefield.[122]

Some organizations with no intelligence support are able to get by on their own. When an organization lacks intelligence support or does not have its own resources to collect and validate the information needed, defining the IE is an almost impossible task.[123] Some of our interviewees believed that it was probably easier to get J2 (the intelligence staff section) involved when focusing on an area or region on which J2 was already focused. (Even better, it is probably easier to get data from J2 if it is already collecting the data that are needed, rather than making a request that will necessitate new data collection.)[124]

A combatant command SME described experiences with assessments that focused on behavioral changes. Because these things are not, by and large, intelligence collection priorities, assessment personnel mostly did their own data collection. This dynamic echoes the one between IIP planning and execution and the assessment of those efforts. There are prospects for improvement on the intelligence side, and a move toward integrated data collection will bring a much-needed focus on nonkinetic operations, their role in larger campaigns, and their utility in furthering progress toward broader joint goals.

Another SME suggested that if one were able to recognize at the outset the need for J2 support, it could be included in IIP planning and assessment from the begin-

---

[120] Author interview with LTC Scott Nelson, October 10, 2013.

[121] Military Operations Research Society, 2012.

[122] Author interview with LTC Scott Nelson, October 10, 2013.

[123] Author interview with LTC Scott Nelson, October 10, 2013.

[124] Author interview on a not-for-attribution basis, December 5, 2012.

ning.[125] JP 5-0 also advocates this approach and includes "Commander's critical information requirements" as an output in the mission analysis stage in JOPP; thus, there is no doctrinal reason that these requirements could not include assessment-related data collection. Such support for IIP efforts and their assessment has the added benefit of helping intelligence staff understand these requirements and could "motivate a shift from intelligence traditions."[126] Of course, unit-level intelligence collection (and intelligence collection plans) also requires command-level support and prioritization.[127] Still, the intelligence community needs to do a better job of providing baseline intelligence.

There is a gap on the intelligence side in part because there is no support for IO in the Defense Intelligence Analysis Program (DIAP), which lists and coordinates the different intelligence requirements and responsibilities across the intelligence community. There is a need for dedicated intelligence specialist support to IO. One barrier is that some IRC operators are highly self-sufficient and may be reluctant to request the intelligence support they need, perhaps having learned from experience that they are unlikely to get it.[128] There are also tangible language barriers between the IRC operators, other operators, and the intelligence community. One SME, who chose not to be named, offered two suggestions for improving assessment integration with J2: (1) encourage J2 to change its priorities and (2) learn what data J2 staff are already collecting and backward plan so that the available data guide the assessment process.[129]

In the 1980s, the PSYOP community could make requests to the intelligence community and receive adequate feedback. But the intelligence tradition has gone back to prioritizing information necessary to support kinetic engagement. To get back to better support from intelligence, there will need to be a push from the senior levels, and it needs to be made a priority.[130] According to one SME, "Our ability to assess is limited by J2's ability to collect, but also by our understanding of the operational environment."[131]

Summary

When organizing for assessment, IIP should be broadly integrated into routine DoD processes, as well as within DoD assessment practices. There is a lack of shared or

---

[125] Author interview on a not-for-attribution basis, December 5, 2012.

[126] Author interview on a not-for-attribution basis, December 5, 2012.

[127] Author interview on a not-for-attribution basis, January 23, 2013.

[128] Author interview on a not-for-attribution basis, July 30, 2013.

[129] Author interview on a not-for-attribution basis, July 30, 2013.

[130] Author interview on a not-for-attribution basis, July 30, 2013.

[131] Author interview on a not-for-attribution basis, December 5, 2012.

complete understanding of IIP assessment among both DoD leadership and the intelligence community, so it is necessary to be much more explicit about processes and assumptions. In this way, IIP assessment stands in sharp contrast to the shortcuts and heuristics that characterize the assessment of kinetic activities.

It is also important to recognize different assessment roles: data collector, assessor, validator, integrator, and recommender. Some of these roles can be accomplished by the same individuals or at the same organizational levels, but the data collector and assessor roles should be separate from validator, integrator, and recommender roles.

Some best practices for DoD include ensuring that assessors are sufficiently independent and empowered to identify and address problems in execution or assumptions when evaluation reveals them, avoiding over-optimism through independent assessment or formal devil's advocacy, and an increased focus on collaboration, particularly among experts from different disciplines within DoD. The following points summarize some specific best practices related to resources, leadership, intelligence, and organizational culture:

- Assessment requires resources. Not all assessment needs to of the same quality and depth. (A general rule of thumb is that roughly 5 percent of a program's resources should be dedicated to assessment.)
- Assessment requires a strong commitment from leadership. Leaders who value assessment and make decisions supported by assessment output are typically more willing to allocate resources to assessment. Furthermore, leaders cannot be afraid of bad news; the only way to improve is to recognize what is not working and fix it.
- Assessment requires intelligence support. The intelligence community could assist IIP assessment efforts by sharing information it is already collecting, including cultural intelligence, network analyses, and cognitive states and behaviors of noncombatant populations. In addition, intelligence can be an excellent source of data to populate assessment frameworks. Since current intelligence structures may be unable or unwilling to meet some IIP assessment data requirements, other ways to collect needed data will need to be identified, planned, and resourced.
- Assessment requires an organizational culture in which it is prioritized. Organizations that do assessment well usually have cultures that value assessment. Changing organizational culture can be difficult, but we identified several starting points for such a shift:
  - leadership buy-in and leadership support
  - leading by example
  - a preference for management by evidence over management by intuition
  - distinguishing between assessment and auditing
  - using a spectacular failure as an example to show how assessment could have prevented it, identified it sooner, or even fixed it

- – disseminating assessment results back through all layers that contributed to encourage improvement
- – putting a *why* with the *what*—offering motivation and explanations to personnel, rather than just giving guidance and instruction
- – including assessment in plans and including planners in assessment design to ensure that assessment considered from the beginning improves buy-in.

Determining What's Worth Measuring: Objectives, Theories of Change, and Logic Models

Many of the principles identified in Chapter Three concern the importance of clear goals and objectives, the importance of clear logical connections between IIP activities and IIP objectives, and the importance of measuring these things. This chapter focuses on goals and objectives, the foundation for both operational and assessment success. The discussion highlights the properties that objectives should have and offers advice for setting (or refining) objectives so that they will have these desirable properties. The chapter then addresses the expression of a theory of change or logic of the effort that connects activities with the properly articulated objectives of the effort. Defining (or refining) objectives in an assessable way and articulating a theory of change are foundational for assessment success.

Setting Objectives

Setting objectives for an IIP effort or activity is a nontrivial matter. While it is easy to identify high-level goals that at least point in the right direction (e.g., win, stabilize the province, promote democracy), getting from ambiguous aspirations or end states to useful objectives is challenging. Yet, as we argued earlier in this report, clear objectives are necessary not only for the design and execution of effective IIP efforts but also for their assessment. This section describes some of the challenges and tensions inherent in setting IIP objectives and offers some advice regarding considering and setting objectives.

Characteristics of SMART or High-Quality Objectives

The received wisdom on assessment holds that objectives should be "SMART"—that is, specific, measurable, achievable, relevant, and time-bound.[1] Table 5.1 summarizes

---

[1] Author interview with Thomas Valente, June 18, 2013; Jessica M. Yeats and Walter L. Perry, "Review of the Regional Center Enterprise Measures of Effectiveness Plan," unpublished RAND research, 2011, p. 9; interview with Anthony Pratkanis, March 26, 2013.

Table 5.1
Characteristics of SMART Objectives

An Objective Is . . .	If...
Specific	It is well defined and unambiguous and describes exactly what is expected
Measurable	One can measure the degree to which the objective is being met
Achievable	It is realistic and attainable
Relevant	The achievement of the objective contributes to progress toward high-level strategic and policy goals
Time-bound	It has deadlines or is grounded within a deadline

SOURCE: Yeats and Perry, 2011, p. 9.

each of these criteria; each is then explored in greater detail, along with a selection of additional virtues to which objectives should aspire.

Specific

As noted in Chapter Three, specificity is essential; how can you talk about progress toward or accomplishment of a goal if you have not specified what the goal really is? This is particularly important for IIP efforts and their assessment because objectives in this area need to, according to one SME, "be very literal." It can be a source of difficulty when objectives are "abstract or wishy-washy."[2]

IIP objectives need to specify what behavior or behavior change is desired and from what audience or group.[3] Army FM 3-13, *Inform and Influence Activities*, presents a scheme for generating objective statements that, if followed, would certainly help a user meet the "specific" requirement. According to FM 3-13, an inform and influence objective statement should have four elements, each of which should be clearly articulated: the desired effect or outcome, the specific target, the desired target behavior, and the rationale for getting the target to perform that behavior (connecting the behavior to the outcome).[4] Figure 5.1 illustrates this construct.

It is important that objectives specify *what* is to be accomplished, not how it is to be accomplished. As noted in JP 5-0, "An objective does not infer ways and/or means—it is not written as a task."[5] Consider some of the objectives that correspond to the DoD IIP examples used in this report so far. The objective to promote voter turnout is fairly clear, but it could be more specific. The desired action is clear: Get the target audience to vote. The previous discussion made the purpose clear: Support

---

[2]   Author interview with Emmanuel de Dinechin, May 16, 2013.

[3]   Author interview with Anthony Pratkanis, March 26, 2013.

[4]   Headquarters, U.S. Department of the Army, 2013a, p. 7-2.

[5]   U.S. Joint Chiefs of Staff, 2011a.

Figure 5.1
Sample Inform and Influence Activities Objective Statement
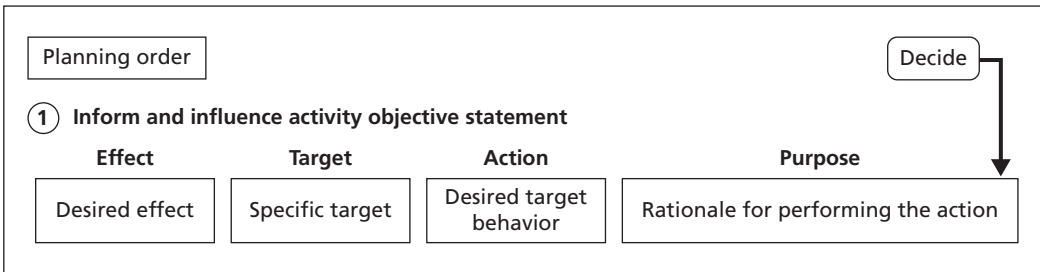
SOURCE: Headquarters, U.S. Department of the Army, 2013a, Figure 7-1.
RAND *RR809/1-5.1*

democratization and governance processes. What is not clearly specified is the target audience, which could be all eligible partner-nation citizens or perhaps one or more traditionally underrepresented groups. The extent of the desired effects could also be better specified: Among the target audiences, what is the desired level of increased voter turnout? Five percent? Ten percent? Specificity to that level forces more-careful planning and encourages proactive refinement if interim measures show that the effort has not made as much progress as desired.

Measurable

A measurable objective is one that can be observed, either directly or indirectly. High-quality objectives will allow observation of the degree to which the objective is being met (percentage of population adopting desired behavior or frequency with which targeted audience engages in desired behavior) rather than all or nothing (extremist rhetoric eliminated from radio broadcasts).

Some objectives, especially those that are not behavioral and cannot be directly observed, can still be meaningfully measured. Customer satisfaction is one example, as are various desired sentiments or attitudes. While perception of security cannot be directly observed, it can be self-reported in an interview, survey, or focus group, and it is likely to be highly correlated with proxy behaviors that *can* be directly observed. Pedestrian and vehicular traffic in an area, the number of people in the market on market day, and the percentage of school-age children who actually attend school are all observable and measurable things that could be proxy indicators for perceptions of security.

One way to move toward measurable objectives is to ask as part of the objective-setting process, "How will we know if we are meeting the objective?" If that question produces a clear idea about something to observe, or a clear indicator or measure to capture, then the objective is probably already measurable. If, on the other hand, that question prompts no clear answer, the objective should probably be refined.

Another way to test the measurability of an objective is to consider what it might mean to different people. Suppose we were to show ten data collectors videos of people engaged in various activities as a way to assess whether an influence objective had been achieved. If almost all (80–90 percent) agreed on what constituted accomplishment of the objective and what did not, then the objective is clearly and observably stated.[6]

Some objectives are just too complex or high level to be meaningfully observed directly, such as democratization or legitimacy. These are still worthwhile strategic goals, but they should be supported by measurable subordinate objectives (see the discussion of nested objectives in Box 3.1 in Chapter Three). Measure development is discussed in greater detail in Chapter Six.

Achievable

An objective must be something that one can reasonably expect to achieve. No IIP program is going to solve world hunger.[7] As the evaluation researchers Rossi, Lipsey, and Freeman note,

> Program advocates often proclaim grandiose goals (e.g., improve the quality of life for children), expect unrealistically large effects, or believe the program to have accomplishments that are clearly beyond its actual capabilities. Good evaluation questions deal with performance dimensions that are appropriate and realistic for the program.[8]

IO SMEs informed us that DoD IIP efforts are certainly not immune to this kind of objective inflation. Nor is public diplomacy. As the public diplomacy expert Phil Seib reminded us, "Success doesn't mean loving America." It is much more beneficial to set reasonable standards and benchmarks on objectives that are more realistic and useful.[9]

Achievable objectives are a balance between reasonable goals and reasonable expectations. Changing behaviors can require significant investments of time and resources, and it does not always work.[10] Those planning and executing IIP efforts must be patient and not expect to see immediate or extreme results. This is another area in which breaking objectives into smaller incremental chunks can be helpful, as the level of effort that turns out to be required to achieve the earliest and simplest of nested and progressive objectives can provide some indication of how difficult it will be to achieve subsequent objectives—if, in fact, the full scope of objectives is achievable in a reasonable time frame.

---

[6]   Author interview with Victoria Romero, June 24, 2013.

[7]   Author interview on a not-for-attribution basis, July 30, 2013.

[8]   Rossi, Lipsey, and Freeman, 2004, p. 71.

[9]   Author interview with Phil Seib, February 13, 2013.

[10] Author interview with Larry Bye, June 19, 2013.

Goals can be unachievable in two ways: The goal could be impractical or the timeline for achieving it could be impossible. Getting 100-percent voter turnout or reducing the incidence of violence in a troubled province to zero is just not possible. Increasing voter turnout from 50 to 60 percent or reducing violent incidents from 50 per month to fewer than 15 per month might be possible but could not be accomplished in a single week. The SMART characteristics are mutually reinforcing; if objectives are specific, it is much easier to ascertain whether or not they are achievable.

Relevant

Nesting objectives such that they are clearly connected also helps ensure that objectives are relevant to overall end states or campaign goals. If one is not careful, it is entirely possible to specify objectives that are observable and measurable but not actually connected to the mission or desired end state. Irrelevant (but achievable) objectives are harder to avoid if the implied or explicit theory of change or logic of the effort does not adequately connect intermediate or tactical objectives with campaign or long-term objectives. This is what happens in situations analogous to winning all the battles but losing the war. As JP 5-0 states, "An objective should link directly or indirectly to higher level objectives or to the end state."[11]

Irrelevant objectives are usually "missing a link" in their theory of change. A defense SME shared an anecdote about a "tip line to nowhere."[12] In the country of interest, an IIP effort sought to persuade local citizens to report suspicious activity to a tip line. IIP activities were conducted, and a line was established. A few months after the effort began, the tip line began receiving a significant number of calls, and the effort was considered successful. However, while the effort met the stated objective of changing local behavior to report suspicious activity to a tip line, it was not successful in any real sense. Why? Because the line was not "connected" to anything. That is, there was no procedure in place to validate the tips through other sources and then pass them to local authorities (or anyone else) to investigate or act on them. Tips were simply recorded in a logbook that then *just sat there*. The objective of collecting tips, was, by itself, not relevant to the campaign; only when and if collecting tips was connected to superordinate and longer-term objectives related to the reduction of criminal or insurgent behavior and the capture of perpetrators would it have become relevant.

Time-Bound

Finally, an objective should include a time horizon for its completion. Objectives that are not time-bound invite efforts in perpetuity that are making little or no real progress. Even if the desired end state is a generational change in international relationships, the intermediate objectives should have some kind of indicated time scope. Time boundaries need not be more precise than the science will allow, and they can be phrased as

---

[11] U.S. Joint Chiefs of Staff, 2011a.

[12] Author interview on a not-for-attribution basis, March 13, 2014.

opportunities to assess progress and revisit plans rather than times after which progress will be considered to be lagging. The timing of objectives can be tied to other natural temporal boundaries. How much progress on this chain of objectives do you think you will have made by the elections next year? How much progress on this objective will you make during your duty rotation? Timing should be specified, and so should the preliminaries of what should happen (be it taking a benchmark measure, some kind of scrutiny, revisiting the theory of change, launching the next phase of the effort, or considering canceling the activity) when a time boundary is reached.

Promoting voter turnout is an example of an IIP objective that is naturally time-bound (by the election date). Combatant command–sponsored SME exchanges (SMEEs) with partner nations are an example of one that is not. In a SMEE, U.S. and partner-nation military personnel meet and discuss their nation's security challenges and goals, finding common ground and learning from each other. The objectives are somewhat nebulous but may include building partner capacity through expertise, relationship and network building, laying a foundation for trust, and opening lines of communication for future engagement. What are the time bounds? How does one know when such efforts have succeeded and are no longer necessary? In some sense, because of the rotational nature of military service in almost all countries, there is a constantly renewed need for new generations of U.S. and partner-nation personnel to network and connect. That said, there is a time at which military-to-military relations should have matured to the point that a next step is possible (perhaps joint exercises, exchanges for professional military education, or another type of initiative). Where combatant command staffs have identified a target next step to build toward and at least a preliminary timeline for progress to that next step, SMEEs are more likely to remain relevant to broader strategic objectives.

Behavioral Versus Attitudinal Objectives

There is debate within the defense IIP community about whether objectives should be exclusively behavioral or whether attitudinal objectives are also permissible. The argument goes something like this: If influence is to contribute to military objectives, it will be because it gets people to do (or not do) certain things (engage in behaviors) that support broader military objectives. There is general agreement that changes in attitude might lead to the adoption of the desired behaviors; but, if you know what those desired behaviors are, you should specify them as part of the objective. For example, if the objective is reduced support for the insurgents, desired behavior changes might include decreased provision of havens to the insurgents, decreased provision of money or supplies to the insurgents, or decreased turnout at insurgent demonstrations or protests. While many of these behaviors might correlate with or even stem from attitudes that are less supportive of the insurgency, the objective is really about the behaviors, even if changing attitudes is part of the planned effort.

The crux of the debate, of course, concerns the extent to which attitudes lead to behaviors, with one view holding that attitudes are poor predictors of behavior and the opposed view holding that attitudes are good predictors of behavior.[13] Appendix D describes a number of social and behavioral science theories of influence, some of which include a role for attitudes and some of which do not.

To the extent that attitudes prove to be good predictors of behavior, the debate begins to lose meaning. However, where attitudes do *not* predict behavior well, the debate matters, and specifying behavioral objectives should be strongly preferred. Fortunately, articulating a clear theory of change/logic of the effort that connects planned activities with desired end states (as we advocate) allows the specification of both attitudinal and behavioral intermediate objectives and allows them to be tested as hypotheses in context as part of assessment. If a theory of change specifies a path promoting, first, attitudinal change, then behavioral change, and then achievement of the desired end state, the validity of this path can be tested.

The debate goes further. Depending on your view about the relationship between attitudes and behaviors, stopping at attitudinal objectives either (1) promotes sloppy thinking, because they likely stop short of being SMART because they fail to fully connect to the desired end state, or (2) is good because they are flexible, allowing the people whose attitudes have been changed to choose the specific behaviors through which they will express these changed attitudes and beliefs, encouraging the behaviors you want and possibly other unconsidered but beneficial behaviors as well.

While we do not resolve this debate here, if the ultimate goal or end state requires that something demonstrable has changed (be it an adversary's capitulation, the election of a government friendly to the United States, or something else), it is probably best to specify the behaviors that will lead to those end states rather than stopping at attitudes favorable to those end states. And if (as we advocate) planners have specified a string of nested and progressive intermediate objectives, there is no harm (and there may be a benefit) in having these nested objectives include a mix of attitudinal and behavioral elements. Again, behavioral objectives are strongly preferred over attitudinal objectives. Attitudinal changes may be included as subordinate or supporting objectives and as part of a longer chain of logic, but ultimate objectives should be some kind of consequential behavioral change.

Intermediate Versus Long-Term Objectives

Related to the time-bound aspect of SMART objectives is the potential tension between intermediate and long-term objectives. Many IIP end states are long-term and do not

---

[13] For the first view, see, for example, Andrew Mackay, Steve Tatham, and Lee Rowland, "The Effectiveness of US Military Information Operations in Afghanistan 2001–2010: Why RAND Missed the Point," *IO Sphere*, December 3, 2012. On the latter view, see, for example, Arturo Muñoz, "Response to 'Why RAND Missed the Point,'" *IO Sphere*, January 15, 2013.

lend themselves to intermediate measures of progress.[14] However, some easily achiev-able intermediate objectives can end up being accomplished but not actually contrib-ute to any higher-level objectives or end states. Earlier in this chapter (in the section "Relevant" in our discussion of SMART objectives), we gave the example of the tip line to nowhere, in which an IIP effort to promote use of a tip line was successful in collecting tips, but the tips were never shared with the authorities who could act on them. Since tips were not passed on or actioned in any way, they made no contribution to the broader objectives of reducing criminal or insurgent behavior and capturing the perpetrators.

The solution, of course, is to have both intermediate and long-term objectives. Specify the long-term objective as precisely as possible and keep it available as a con-stant reference. Then, identify the incremental steps that you believe will lead you to that end state: "Define what conditions will change at each phase and how to detect the new behavior or function."[15] These intermediate objectives provide actionable and assessable objectives in the short- and medium-terms. Further, beliefs about the steps necessary to reach a desired end state can be tested as hypotheses. Does the second intermediate objective actually lead to the third intermediate objective? If not, revise it (sooner rather than later) so that a solid logical connection can still be made between intermediate objectives and the ultimate long-term objective.

For example, the ultimate objective for the tip line could have been to take action against insurgents based on synthesis of citizen tips and corroborating intelligence, with a secondary objective to increase citizen participation in legitimate government processes, such as the reporting of criminal or insurgent behavior. Intermediate objec-tives, then, would include not only establishing and advertising the tip line but also transmitting tips received to relevant parties (such as law enforcement), the timely vali-dation of tip intelligence, and timely action based on the tips.

How IIP Objectives Differ from Kinetic Objectives

As noted in Chapter Two, there is considerable shared understanding about how kinetic military efforts function (mostly based on a combination of physics and experience), but this is not the case for IIP efforts. Because of this, there are numerous shortcuts, heuristics, and correct shared assumptions in the planning and assessment processes for kinetic efforts that are not available for IIP planning.

This difference extends to objectives, too. The same shared understanding and known valid assumptions allow shortcuts in specifying kinetic objectives. While they still should be SMART, different elements of SMART are often assumed (often cor-rectly) for kinetic objectives. Take, for example, the tactical objective "destroy that bridge." This objective is not all that specific: It does not indicate a percentage of

---

[14] Author interview on a not-for-attribution basis, August 1, 2013.

[15] The Initiatives Group, 2013, p. 21.

destruction or a specific number of feet of bridge surface that must be rendered unusable, but it is specific enough for military purposes; there is a shared understanding about what *destroyed* means, and that is sufficient. That same shared understanding makes the objective measurable, and such measurements will be performed through traditional battle damage assessment processes, so further clarification is not necessary. The objective is certainly achievable; there is no bridge in the world that could not be reduced through fires. The challenge is to match the correct number of sorties, strikes, or shells to the task, and there is (again) a traditional way to make the necessary calculations and allocate sufficient resources. This objective is not guaranteed to be relevant, however. The objective as stated does not indicate how the destruction of the bridge serves the overall campaign plan, but if the superordinate commander's intent is known, the connection will be obvious. If the connection is *not* obvious, the addition of a simple clause to the objective—"in order to . . ."—will wholly satisfy the requirement for relevance. Finally, as stated, the objective is not time-bound. Again, that will be either elaborated or assumed. The implied time bound will often be "within the next air tasking order cycle," or it may be implicit in the shared understanding that connects the bridge's destruction to the commander's intent. For example, if the purpose is to deny an enemy the ability to use the bridge to supply its forces or bring in reinforcements, then the time bound is to destroy the bridge before the enemy uses it for these purposes. Once again, even if the assumptions underlying the objectives are not perfectly clear, planners can still rely on a preexisting understanding of the overall process to make the necessary specification, or they know to ask for clarification.

These shared assumptions and understandings are just not there when it comes to IIP efforts. For example, if, instead of "destroy that bridge," what if the objective were "get that formation of enemy troops to surrender"? In this IIP objective, at least there is a shared understanding of what the end state would be. There is a clear behavioral goal, with enemy soldiers abandoning their vehicles, weapons, and positions and moving toward U.S. forces in a nonthreatening manner with hands raised, possibly hoisting a white flag. But that is where shared understanding ends and more SMART clarity is required. How many enemy forces? Within what time frame? All at once or a few at a time? Is the objective actually to take them all prisoner or for some of them to just desert, abandoning weapons and uniforms and returning home? Shared assumptions continue to be lacking even after the objective is clarified and discussion begins to address *how* to persuade the enemy formation to surrender. It is easy to imagine a number of different approaches: Bomb them until they surrender; cut off their lines of communication, retreat, resupply, and wait for them to surrender; use loudspeakers to demand their surrender (in their native tongue), and provide them procedural instructions; drop leaflets demanding their surrender, pointing out that the leaflets could just as easily have been bombs; or explore a range of other options drawn from history or from plausible notions. What is not easy is identifying which (or which combination) of these options is most likely to produce the desired result on the desired timeline.

Unlike destroying a bridge, military officers lack shared understanding about how effective each of these approaches is likely to be and how long it is likely to take. Different officers may prefer different combinations of approaches, but there is little in the way of evidence (or training or experience shared with other officers) to justify their views. So, IIP objectives must be both SMART and much more explicit than objectives for kinetic campaigns. Theories of change/logics of efforts for IIP efforts cannot be assumed in the way that they can for kinetic efforts, and therefore they must be made explicit as well.

IO and kinetic efforts (specifically, fires) differ in other ways as well. The *Marine Corps Operating Concept for Information Operations* identifies four such differences:

> One: information must compete for the attention of the intended target while fires have no such requirement. Two: although the target of fires may have few choices about the effects to which it is subjected, the target of an information operation can choose what signals to heed or ignore through the application of social and cultural filters. Three: although the effects of fires remain limited to targets within the designed radius of the ordnance, information effects can propagate well beyond the intended target and perhaps pick up strength, change, and create unintended consequences. Four: although the physical effects of fires are self-demonstrating, information must be interpreted by their target, which does so according to its own frame of reference. Hence what is ultimately received may not be intended by the sender, and what is received by one target may be different than the one received by another. Additionally effects within the IE, especially within the cognitive dimension, are often difficult to measure and assess.[16]

The point of the observation that informing, influencing, and persuading differs from kinetic military action is not to plead for exceptionalism or to argue that the former is harder than the latter. Rather, the intent is to point out that they are not equally ingrained; IIP efforts do not benefit from the same intuition and assumptions that facilitate the planning and assessment of kinetic efforts, and this necessitates greater levels of explicit detail in IIP planning and assessment. Stating the theory of change (and being prepared to modify it based on contact with the context) is critical in IIP planning and assessment in a way that it is not for exclusively kinetic operations. The planning process for IIP operations and kinetic operations is fundamentally the same, but IIP planning requires that more (assumptions, theory of change/logic of the effort, details of the objective) be made explicit. These explicit details should be generated during mission analysis and COA development (steps 2 and 3) in JOPP, and they are an essential product of operational design for IIP efforts. COAs might even include competing theories of change for how to achieve an IIP objective (or an objective supported with IIP elements) for exploration during COA analysis and war-gaming.

---

[16] U.S. Marine Corps, 2013, p. 6.

Efforts to inform, influence, and persuade differ from kinetic efforts in many important ways. Because military planners can more perfectly intuit the relationships between actions and outcomes in the kinetic realm, shortcuts preserve meaning and are effective. However, because the social and psychological processes required of influence efforts are not part of standard military intuition, it is important that connections (and assumptions) be spelled out.

How to Identify Objectives

Much of the discussion so far has focused on the characteristics of well-formed IIP objectives. Often, just identifying the desired characteristics will push a planner toward better-specified objectives. However, it is sometimes the case that the overall goal is clear but how to describe the objectives effectively is not. In our research, we encountered a number of suggestions regarding processes for identifying and refining objectives.

One piece of advice was to work with stakeholders to better refine goals and objectives. The evaluation researcher Stewart Donaldson suggests that stakeholders, activity planners, and evaluators work together to develop a common understanding of what the objectives are and what the theory of the program is.[17] Another evaluation researcher, Eric Biersmith, recommends building a logic model that includes objectives (outcomes) and involving stakeholders in that process; this approach, he argues, helps all stakeholders reach a shared vision of the effort.[18] This admonition to engage with stakeholders when defining and refining objectives is certainly applicable in the defense context. JP 5-0 suggests that "frequent interaction among senior leaders, combatant commanders (CCDRs), and subordinate joint force commanders (JFCs) promotes early understanding of, and agreement on, strategic and military end states, objectives, planning assumptions, risks, and other key factors."[19]

Part of this engagement with stakeholders can involve asking strategic questions. If initial guidance from higher levels is not sufficiently specific, return with clarifying questions: Who? What? How much? By when?[20] Even absent broad stakeholder engagement, these are good questions. If objectives are insufficiently articulated in guidance from the higher level, those at the planning and execution level can try to refine objectives until they are SMART. These refined objectives can then be pushed back up to the higher level for approval. This form of "leading up" can be highly effective. If the right refinements have been made at the lower level, then the higher level

---

[17] Stewart I. Donaldson, *Program Theory–Driven Evaluation: Strategies and Applications*, New York: Lawrence Erlbaum Associates, 2007, p. 10.

[18] Eric Biersmith, "Logic Model as a Tool to Evaluate Prevention," paper presented at Evaluation 2013, the annual conference of the American Evaluation Association, Washington, D.C., October 14–19, 2013.

[19] U.S. Joint Chiefs of Staff, 2011a, p. x.

[20] Ketchum Global Research and Analytics, undated, p. 6.

will undoubtedly approve; if, on the other hand, assumptions made at the lower level do not match the unexpressed intent at the higher level, comments and guidance that come down from the higher level with the rejection of the proposed revisions should help move things in the right direction. The third chapter of JP 5-0, "Operational Art and Operational Design," urges commanders to collaborate with their higher head-quarters to resolve differences in interpretation regarding objectives and achieve clar-ity. This should be done as part of the "understand the strategic direction" element of operational design, and it should take place in JOPP during the planning initiation or mission analysis step (or perhaps between them).

Time permitting, objectives can also be a subject for formative research. Think-ing about goals and objectives as research questions for evaluation can help improve strategy articulation.[21] Especially in areas where validated theories of change are lack-ing, formative research can help lead to SMART objectives by determining what can actually be observed, what kinds of changes are realistic to expect, and how long they are likely to take. Formative research can help improve both assessments and IIP efforts themselves, and this topic is discussed in greater detail in Chapter Eight.

The scholar Ralph Keeney acknowledges that identifying objectives can be tricky but suggests that objectives can eventually be identified if you start by creating a list of values: What should the effort accomplish and what should it *not* accomplish. He asserts, "Once values are presented in the form of a list, it is not difficult to systemati-cally convert them into objectives."[22] He suggests several different rhetorical devices for listing values, which can then be used to move on toward objectives. Table 5.2. lists these devices.

The measurement specialist Douglas Hubbard recommends a process he refers to as "clarification chains" to help take clear but imprecise goals and make them more specific and measurable.[23] He describes this process as a thought experiment, a decom-position, identifying something we think of as intangible in an effort to discover what is tangible (and, most important, measurable):

> How could we care about things like "quality," "risk," "security," or "public image," if these things were totally undetectable, in any way, directly or indirectly? If we have reason to care about some unknown quantity, it is because we think it cor-responds to desirable or undesirable results in some way.[24]

---

[21] Author interview with Craig Hayden, June 21, 2013.

[22] Ralph L. Keeney, "Developing Objectives and Attributes," in Ward Edwards, Ralph F. Miles, Jr., and Detlof von Winterfeldt, eds., *Advances in Decision Analysis: From Foundations to Applications*, Cambridge, UK: Cam-bridge University Press, 2007, p. 110.

[23] Hubbard, 2010, p. 27.

[24] Hubbard, 2010, p. 27.

Table 5.2
Devices to Help Articulate Values

Device	Questions
Wish list	What do you want? What do you value? What would be ideal?
Alternatives	What is a perfect alternative, terrible alternative, a reasonable alternative, the status quo? What is good or bad about each?
Consequences	What has occurred that was good or bad? What might occur that you care about?
Goal and constraints	What are your aspirations to meet the stated goals and constraints? What limitations do these place on you?
Different perspectives	What would your competitor or constituency or other stakeholders be concerned about? At some time in the future, what would concern you?
Strategic values	What are your ultimate values that may be represented in a mission statement, a vision statement, or a strategic plan? What are your values that are absolutely fundamental?
Generic values	What values do you have for your customers, you employees, your shareholders, yourself? What environmental, social, economic, or health and safety values are important?
Why do you care?	For each stated value, ask why it is important. For each response, ask why it is important.
What do you mean?	For each stated value, specify its meaning more precisely. For broad values, identify major component parts.

SOURCE: Keeney, 2007, p. 110, Table 7.3. Used with permission.

In the DoD context, this process might be aimed at one of the more nebulous higher-level objectives mentioned previously, such as democratization or stability. What stability means and what is necessary to achieve stability will differ in different contexts.[25] A clarification chain for developing SMART objectives for a stability operation might start by discussing what is contributing to instability in that context (perhaps armed gangs, lack of employment opportunities, weak infrastructure, ethnic tensions, or a lack of community leadership), and then trying to identify which of those things are connected to each other, which are mutually reinforcing, and which might disappear on their own if others are removed. Such a process could have multiple benefits: not only the specification of more-tangible (and otherwise SMART) objectives but also the beginnings of a theory of change for how these things connect and what needs to be done about them. Such an exercise would be a reasonable part of mission analysis in JOPP, and it would be important to understanding both the strategic direction and the operational environment in operational design.

---

[25] Jan Osburg, Christopher Paul, Lisa Saum-Manning, Dan Madden, and Leslie Adrienne Payne, *Assessing Locally Focused Stability Operations*, Santa Monica, Calif.: RAND Corporation, RR-387-A, 2014.

Setting Target Thresholds: How Much Is Enough?

A combination of the specific, achievable, and time-bound aspects of SMART informs the step of setting target thresholds for objectives. How much is enough? What proportion of a target audience needs to adopt a desired behavior for the effort to be considered a success? What level of progress do you need to make toward an intermediate objective before you launch activities that aim to build on that progress and before you move the effort toward accomplishing a later subordinate objective? At what threshold have your efforts accomplished all they can toward this objective, indicating that it is time to transition to different efforts and objectives or to take the program elsewhere?

First and foremost, being specific about such targets is good practice, and this approach was advocated broadly by the SMEs with whom we spoke and in the literature we reviewed.[26] In the words of the brand strategist Olivier Blanchard, "The specificity of targets drives accomplishment. The more specific, the more likely the desired outcome will be reached. The less specific the goal, the less likely it will be met. Always set targets."[27]

Once again, your desired end state and ultimate goal should help drive thresholds. In an election, 51 percent voting for your preferred candidate is an unambiguous success.[28] However, for an effort promoting voter turnout, what amount of improvement is desired? Almost no IIP effort should expect 100-percent change or accomplishment, whatever the objective is. Even where an objective is relative, seeking an increase or decrease in a behavior (such as "decrease insider attacks in province X"), it should be accompanied by a target threshold—expressed either in percentage terms or in absolute terms.

Sometimes, the overall end state is not sufficiently specific to identify thresholds for IIP efforts. When this is the case, asking the questions that can get to those thresholds can substantially improve both the focus of IIP efforts and their assessment. Consider this example from the pages of *IO Sphere*:

> The commander stated one of his objectives was to "remove noncombatants from the town." Designing an MOE to meet that objective would require a PSYOP officer to clearly understand what the commander meant by "remove" and "noncombatant." He could gain that information from the commander's written intent and desired end states, or he could ask the [commander] for specific parameters. How many—quantity—will have to leave to meet the commander's intent: 100% of all persons not carrying weapons, 80% of women, children, and men over age 60, etc.? How far from Fallujah—distance—should they go to be considered

[26] See, for example, Donna M. Mertens and Amy T. Wilson, *Program Evaluation Theory and Practice: A Comprehensive Guide*, New York: Guilford Press, 2012.

[27] Blanchard, 2011, p. 18.

[28] Author interview with Mark Helmke, May 6, 2013.

"removed"? How long should they stay away—persistence? The answers to these questions establish the standards of judgment; they make assessing PSYOP results easier because they can be defined, their attributes analyzed, and their parameters/bounds determined.[29]

What exactly a target threshold will be should be driven, in part, by the problem at hand and, in part, by what your theory of change leads you to believe is possible, as well as the time span within which you believe it to be possible. Comparison with past performance or with other, similar efforts can be useful in setting thresholds.[30] What have similar efforts been able to accomplish in other contexts? For example, if a violence-reduction program in Haiti reduced violent incidents by 15 percent over six months, that might be a useful starting point as an objective for a similar effort somewhere else (adjusting for the initial level of violence, the relative scale of the effort, and any other input derived from the theory of change).

Another way to think about the target threshold is in a decisionmaking context. Remember that assessment should support decisionmaking. How much of something do you need to see in order to reach a decision point, or for you feel compelled to choose a different course of action?[31]

If an effort does not achieve its specified threshold of an objective within the targeted time, there is an opportunity for scrutiny. Why did it fall short? It may be that the initial expectation was a little unrealistic and that things appear to be on a trajectory to meet the target but are happening just a bit late. Or it may be that there were performance problems with some elements of the effort, and the shortfalls are directly related. Or it may be that some of the assumptions in the theory of change did not hold, or relationships were not as strong as assumed, and the theory of change needs to be updated. Clear target thresholds can help mitigate against open-ended commitments (where improvement continues to be sought long after enough of whatever was improving has been gained), and they can help turn "good enough" into "better" the next time by identifying weaknesses in theory or practice.

An effort should have termination criteria—clear guidelines for what constitutes sufficient accomplishment to move on to the next stage of the effort or to consider the effort complete.[32] Termination criteria should be developed as part of operational design, according to JP 5-0. Programs have a life cycle, and it should be viewed positively when a program accomplishes its objective and can be allowed to end.[33] Related

---

[29] Robert L. Perry, "A Multi-Dimensional Model for PSYOP Measures of Effectiveness," *IO Sphere*, Spring 2008, p. 9.

[30] Mertens and Wilson, 2012.

[31] Hubbard, 2010.

[32] U.S. Joint Chiefs of Staff, 2011c.

[33] Author interview on a not-for-attribution basis, July 30, 2013.

to termination criteria, ideally, an objective should include indicators of failure, too.[34] Good objectives need to at least imply what failure would look like. How will you know if you have not succeeded? Complete failure is often easy to recognize: A tip line receives no calls (or generates no actionable tips), an election-participation campaign yields no increase in voter turnout, security improves but perception of security does not, no enemy soldiers respond to calls to surrender. Distinguishing partial success from partial failure can be particularly difficult, but is not just an "is the glass half full, or is it half empty?" dilemma. Some results are equivalent to zero and should thus be considered failures. For example, if the margin for error for a 2-percent increase in voter turnout is plus or minus 2 percent or more, that is equivalent to failure. If an objective includes a target threshold, it could include multiple target thresholds— perhaps a minimum success threshold and a desired target threshold—based on what accomplishing the objective is supposed to contribute to the larger campaign. If a commander wants an enemy formation to surrender in order to minimize loss of life to both sides, conserve blue force resources, and minimize property damage, that will influence the target threshold. Of course the commander would like 100 percent of the enemy soldiers to surrender, but that may be unlikely. Perhaps mission analysis or COA analysis reveals that if 70 percent surrender, the remainder can be scattered or captured without resorting to significant indirect fires and incurring heavy friendly casualties. Perhaps the same analysis reveals that if 30 percent surrender, that would still significantly weaken the enemy's fighting strength and would sufficiently reduce blue force casualties to be worthwhile. In this case, anything less than 30 percent, though still some kind of accomplishment, would not justify the time spent conducting the effort. While the minimum success threshold for a similar effort in a different time and place may be different, for this effort, anything below 30 percent would be considered a failure.

Logic Model Basics

One of the recurring themes of this report is the importance of (and the benefits from) specifying a theory of change/logic of the effort for an IIP effort. A logic model is one way to collect and express the elements of a theory of change: "The logic model is supposed to make the program's theory of change explicit. A theory of change describes how the activities, resources, and contextual factors work together to achieve the intended outcome."[35] We explore theories of change and their use in greater detail in Chapter Five. Also see the section "Effective Assessment Requires a Theory of Change

---

[34] Author interview with Victoria Romero, June 24, 2013.

[35] Mertens and Wilson, 2012, p. 244.

or Logic of the Effort Connecting Activities to Objectives" in Chapter Three for an introduction to theories of change and their benefits.

Logic models traditionally include program or effort inputs, outputs, and outcomes. Some styles of logic model development also report activities and impacts. Figure 5.2 presents these elements in sequence.

Inputs, Activities, Outputs, Outcomes, and Impacts

The *inputs* to a program or effort are the resources required to conduct the program. These will of course include personnel and funding, but are usually more specific than this, perhaps indicating specific expertise required or the number of personnel (or person-hours of effort) available. An effort's *activities* are the verbs associated with the use of the resources, and are the undertakings of the program; these might include the various planning, design, and dissemination activities associated with messages or products, and could also include any of the actions necessary to transform the inputs into outputs. In fact, some logic model templates omit activities, as activities just connect inputs to outputs and can often be inferred by imagining what has to be done with the inputs to generate the outputs. We include activities here because of the focus on informing, influencing, and persuading, and the fact that assumptions are not always shared, and there is certainly no harm in being explicit about what activities will transform the inputs into outputs.

The *outputs* are produced by conducting the activities with the inputs. Outputs include traditional MOPs and indicators that the activities have been executed as planned. These might include execution and dissemination indicators, measures of reach, measures of receipt/reception, and indicators of participation. *Outcomes* (or effects) are "the state of the target population . . . that a program is expected to have changed."[36] This is the result of the process: The inputs resource the activities, and

Figure 5.2
Logic Model Template
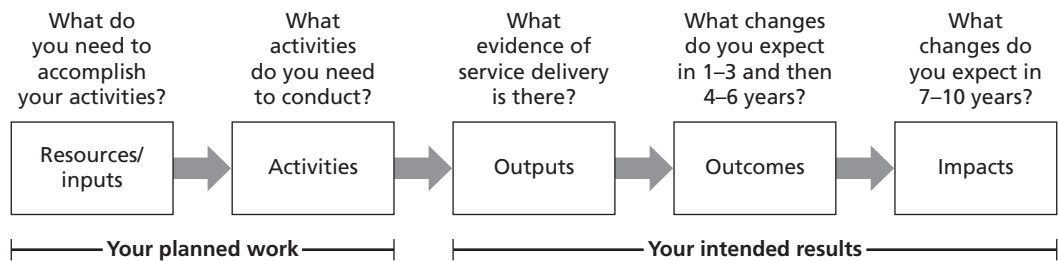
SOURCE: Mertens and Wilson, 2012, p. 245, Figure 7.1. Used with permission.
RAND *RR809/1-5.2*

---

[36] Rossi, Lipsey, and Freeman, 2004, p. 204.

the activities produce the outputs. The outputs lead to the outcomes. This is a critical juncture from a theory of change perspective, as the mechanism by which the outputs (messages disseminated, messages received) connect to the outcomes (behaviors changed) is critical and is a potentially vulnerable assumption in influence and persuasion. Outcomes are characteristics or behaviors of the audience or population, not of the program or effort. The outputs are related to the program or effort, and describe the products, services, or messages provided by the program. *Outcomes* refers to the results (or lack of results) of the outputs produced, not just their delivery or receipt.[37]

The *impact* of a program or effort is the expected cumulative, long-term, or enduring contribution, likely to a larger campaign or superordinate goal. There is no clear dividing line between immediate and short-term outcomes, medium-term outcomes, and long-term impacts. In fact, there is not an agreed-upon difference between outcome and impact. To some, it means the difference between an individual change and a system change;[38] to others, it means a difference in design in that outcomes are not proven to be causally linked to the activities and outputs, but impacts are those outcomes that can be attributed to the intervention due to evidence from (typically) experimental studies.[39] To others, the difference is just a time horizon or level of analysis, with impacts being long-term and expanded outcomes.[40] Under this scheme, if the outcome is the changing of a specific set of behaviors or attitudes, the impact is the durability of that change and the broader consequences of that change. For example, if the outcome of a defense IIP effort is increased participation in an election in a partner nation, the hoped-for impact might be a combination of increased participation in future elections, and increased support for democracy and democratic values.

JP 5-0 both explicitly and implicitly follows logic models. For each of the elements of operational design and each of the JOPP steps, JP 5-0 explicitly lists the inputs so that element or step and the expected outputs. In both processes, many of the outputs of earlier steps or elements are then inputs to later steps. The overall presentation supports a logic model framework. For example, the emphasis in operational art on ends, ways, and means corresponds with logic model language: The ends are the outputs and outcomes, the ways are the activities, and the means are the inputs.

---

[37] Rossi, Lipsey, and Freeman, 2004.

[38] Amelia Arsenault, Sheldon Himelfarb, and Susan Abbott, *Evaluating Media Interventions in Conflict Countries*, Washington, D.C.: United States Institute of Peace, 2011, p. 16.

[39] Author interview with Julia Coffman, May 7, 2013.

[40] Author interview with Maureen Taylor, April 4, 2013.

Logic Models Provide a Framework for Selecting and Prioritizing Measures

A logic model encapsulates a theory of change or the logic of the effort and, done well, suggests things to measure.[41] Each layer in the logic model suggests clear measures. One might ask:

- Were all of the resources needed for the effort available? (inputs)
- Were all activities conducted as planned? On schedule? (activities)
- Did the activities produce what was intended? Did those products reach the desired audience? What proportion of that audience? (outputs)
- What proportion of the target audience engaged in the desired behavior? With what frequency? (outcomes)
- How much did the effort contribute to the overall campaign? (impacts)

These questions point directly to possible measures, and also help to prioritize. Not everything needs to be measured in great detail or particularly emphasized in data collection.[42] For example, the level of assessment data collection for inputs may be quick, simple, and holistic: "Were all the resources needed for this effort available?" "Yes." (Were the answer "no," some relatively simple follow-up questions about which resources were lacking would come next, but the exact degree of deficiency would still not be all that relevant.) Some activities may be similarly simple (activities regarding printing, or securing broadcast time, for example), while others may require more-precise measurement. Outputs and outcomes deserve the greatest measurement attention.

Consider the example theory of change offered in Chapter Three that connects training and arming local security guards and promoting awareness of security and participation in local government (outputs) to improvements in security, improvements in perception of security, improvements in governance (outcomes), and, ultimately, stability (longer-term outcome). Measures would follow the key nodes, and should include measure(s) of the number of local security guards armed and trained; indicator(s) that they are (or are not) present and patrolling, and perhaps how many are doing so; proxy measure(s) of security over time to show whether or not tangible security has improved; measures relating to the delivery and receipt of materials prompting awareness of improved security; measure(s) of perception of security over time (perhaps through a recurring small survey, or perhaps through observations of tangible markers of perception of security, as discussed earlier); measures of the delivery and receipt of materials encouraging participation in local government; some kind of measure(s) of participation in local government (perhaps attendance at the equivalent of town council meetings or the number of unfilled billets in local government); measure(s) of local

---

[41] Author interview with Christopher Nelson, February 18, 2013.

[42] Author interview with Ronald Rice, May 9, 2013.

governance over time (this would be very context specific but perhaps something to do with contracts let, or disputes resolved, or frequency of meetings); and proxies for stability measured over time.

The benefit to measuring aspects of all of the different layers in the logic model is at its greatest when an effort is not working, or is not working as well as imagined. When the program does not produce all the expected outcomes and one wants to determine why, a logic model (or another articulation of a theory of change) really shines.

Program Failure Versus Theory Failure

A program or effort does not produce the desired results (outcomes) for one of two fundamental reasons: either *program failure*, in which some aspect of the effort failed to produce the needed outputs, or *theory failure*, where the indicated outputs were produced but did not lead to the intended outcomes. Figure 5.3 illustrates the logic of theory failure versus program failure.

Logic model–based assessment can help identify which is the case, and help initiate steps to improve the situation. If program failure is occurring, scrutiny of resources and activities can lead to process improvement and getting outputs on track. If the theory is flawed, it can be diagnosed, tweaked on the fly and experimented with, or replaced with an alternative theory (and supporting inputs, activities, and outputs).

Consider the security, governance, and stability theory of change or logic model and associated measures described in the previous section. There are a number of ways that chain could be broken through theory failure or program failure, but here is an example that makes the distinction clear: If the local security forces never received the

Figure 5.3
Program Failure Versus Theory Failure
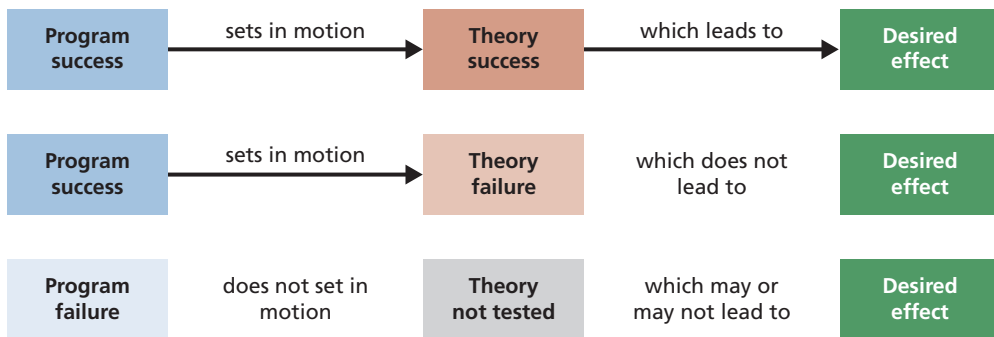
intended arms and training, then that would be program failure. If the local security forces received arms and training but just went home and never patrolled or positively contributed to the security situation, that would be theory failure.

Both program failure and theory failure have the potential to be fixed. It is usually easier to identify how to fix program failure, even if it may be hard to actually generate the inputs required: In this case, see that the arms and training are delivered. Theory failure can be more challenging, as it requires amending the theory of change—and probably the program, too. In this instance, if the armed would-be guards are not engaging in security activities, a number of possible amendments and workarounds suggest themselves: Screen possible participants for those already inclined to patrol, increase pay or other incentives to encourage the trained forces to actually do what they have been prepared to do, or add an IIP component to the training, seeking to increase the likelihood that trainees will engage in security-enhancing behaviors after training is complete.

Constraints, Barriers, Disruptors, and Unintended Consequences

In addition to specifying inputs, activities, outputs, outcomes, and impacts, logic modeling (or other forms of articulating a theory of change/logic of the effort) provides an opportunity to think about things that might go wrong. Which assumptions are the most vulnerable? Which of the inputs are most likely to be late? Which of the activities might the adversary disrupt, or which activities are contingent on the weather? These things can be listed as part of the logic model, and placed next to (or between) the nodes they might disrupt. For example, if local contractors might abscond with funds allocated for printing, or if the contractors are vulnerable to long power outages that can stop their presses, then these things could be noted between the relevant input and activity. If friendly force–caused collateral damage can prevent the translation of a short-term outcome into a long-term impact, it could be noted between outcomes and impacts.

Note that these disruptors can be anything outside the direct control of the program or effort.[43] For IIP efforts, this could include contextual factors (language, culture, history), exogenous shocks (natural disasters, economic crises, significant political action), actions by adversaries, actions by third parties in the information environment, and kinetic actions by friendly forces. The kinetic actions of a force send messages with far greater power than spoken or written messages.[44] If a picture is worth 1,000 words, then a JDAM (joint direct attack munitions) is worth 10,000.[45]

If these potential disruptors can be conceived of as part of the logic modeling process, then, as needed, they can also be included in the measurement and data col-

---

[43] Author interview with Christopher Nelson, February 18, 2013.

[44] Author interview with Steve Booth-Butterfield, January 7, 2013.

[45] Paul, 2011.

lection plan. The collection of such information can further facilitate the adjustment of situations involving apparent program or theory failure, or awareness that failure has come from an unanticipated and external source, and that neither the theory nor the program has actually failed—they have just been temporarily derailed by outside circumstances.

Barriers or disruptors do not necessarily completely disrupt processes (though some do), but all will at least slow down or diminish the rate of success. Perhaps they are best conceived like the "coefficient of friction" in physics. If desired levels of results (be they outputs or outcomes) are not being produced and an identified disruptor is measured as being present, adjustments can be made. These adjustments might simply be to put more of an input or activity in place (realizing that a certain amount is being lost to "friction"), or to identify some kind of workaround to minimize or remove the impact of the disruptor.

Returning to the security, governance, stability logic model, the example program failure (failure to deliver arms and training) and the example theory failure (trained and armed forces not patrolling or otherwise contributing to security) occurred for some reason. If the reason can be identified, it can be added to the logic model as a disruptor and then worked around, both in the current iteration of the program and in future iterations. For example, training and arms might not have been delivered because of a failure to get entry visas in a timely fashion for the civilian contractors scheduled to provide the training. A possible workaround is simple: Get the visas and then execute the training; in the future, start the application process sooner. If visas are being delayed indefinitely, alternative workarounds might be engagement at the political level or the use of personnel already in country to deliver the training. Trained and armed forces not patrolling might be due to a number of possible disruptors: insufficient pay or fear of being overmatched by foes, for example. Or the disruptor might be a hybrid of multiple disruptors, none of which is a showstopper by itself but instead is a source of friction, but together they stop the process. Perhaps half the trainees feel that they are insufficiently paid and will not patrol. The other half would patrol, but because half their squadmates are absent, they fear overmatch and so will not patrol on their own. Possible workarounds could include raising pay, which would (in this narrative) get close to 100 percent of the force in the field, or training and arming additional forces, so that those who feel sufficiently paid to patrol also feel that they have a sufficient number of comrades to patrol with.

Building a Logic Model, Theory of Change, or Logic of an Effort

A theory of change/logic of an effort helps ensure that there are clear logical connections specified (either as assumptions or hypotheses, or a combination of both) between the activities of a program or effort and the objectives. Especially in the cognitive and

behavioral realm, where shared understanding of such connections is lacking, explicitly specifying the theory of change can be critical to both execution and assessment. A logic model, as described above, is one way to articulate a theory of change. This section offers some concrete advice for the building or development of a program theory of change.

Various Frameworks, Templates, Techniques, and Tricks for Building Logic Models

Building a logic model is fundamentally about articulating the underlying logic of the program or effort.[46] To a certain degree, the framework of inputs to activities to outputs to outcomes to impacts is sufficient to begin to develop a logical model. Begin at the right, with SMART objectives, and work backward to the left.[47] What has to happen in order for those objectives to be met? What do you need to do to make those things happen? What resources do you need to do those things? A graphical depiction of this process of working backward appears in Figure 5.4.

Find and Fill Gaps in the Logic Model

Sometimes working backward from SMART objectives will result in more and more uncertainty at the levels of activities and inputs. In some situations (especially IIP situations), it is unclear what activities are most likely to produce the outputs needed to reach desired outcomes. When this occurs, additional information is needed.

Figure 5.4
Working Backward to Articulate a Theory of Change
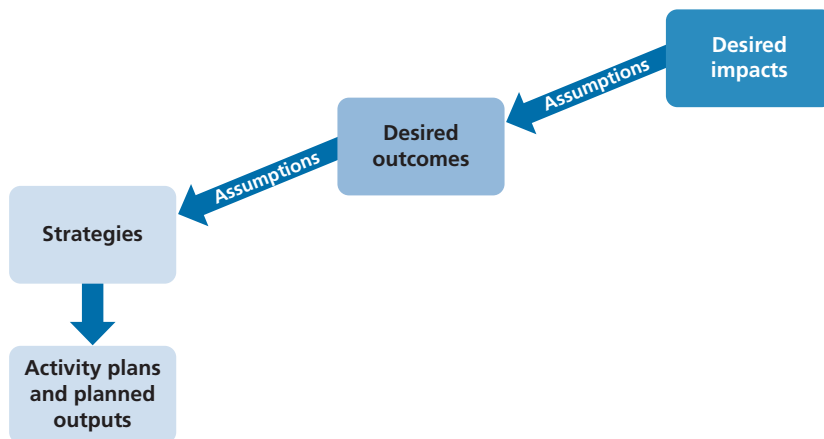
SOURCE: NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 8, Figure 2.
RAND *RR809/1-5.4*

---

[46] UK Ministry of Defence, 2012.

[47] North Atlantic Treaty Organization, Joint Analysis and Lessons Learned Centre, *A Framework for the Strategic Planning and Evaluation of Public Diplomacy*, Lisbon, Portugal, June 2013.

One approach to resolving uncertainty about the best activities to achieve desired outcomes is formative research. Thomas Valente noted that, in his experience, it is sometimes difficult to directly influence a desired behavioral outcome, but that it is often possible to influence mediating factors that can then lead to the desired behavioral change.[48] Formative research can help identify the mediating factors and test which kinds of messages or activities have the most influence on those factors; after such formative research, one is left not only with a thoughtfully articulated logic model but also with one that is at least partially validated. Formative research for this purpose might involve quick field experiments, pilot tests of draft products, or other ways to test activities in a limited way. Alternatively, formative research could involve consultations, workshops, or focus groups with SMEs (either influence SMEs or contextual SMEs, or a combination of both) to get their views on the best ways to effect desired changes.[49] Methods and approaches to formative research are discussed further in Chapter Eight.

Another approach to decreasing uncertainty about which activities will lead to desired outputs and outcomes is a literature review. Look at the existing psychological and behavioral science literature on behavior change, especially programs that have sought similar outcomes.[50] A literature review is a quick and relatively inexpensive way to learn from the experiences (both successes and failures) of others. The social psychologist and influence expert Anthony Pratkanis recommended the review of some of the memoirs of successful influence practitioners of the World War II and Vietnam eras as particularly useful for contemporary defense IIP.[51] One caveat to existing theories: No single social or behavioral science theory explains everything, and different theories will be appropriate to different populations.[52] Finding the right theory for a given objective in a given context may involve synthesizing (and testing) a new theory from various existing theories.

As part of this project, we reviewed a collection of major theories of influence, both from the existing social and behavioral science literature and from the theories implied in existing practice. These are reported in detail in Appendix D. A quick review would, perhaps, allow a practitioner to recognize his or her implicit theory of change or

---

[48]   Author interview with Thomas Valente, June 18, 2013.

[49]   Author interview with Christopher Nelson, February 18, 2013.

[50]   Author interview with Joie Acosta, March 20, 2013.

[51]   Author interview with Anthony Pratkanis, March 26, 2013. Pratkanis recommended the following publications: Martin F. Herz, "Some Psychological Lessons from Leaflet Propaganda in World War II," *Public Opinion Quarterly*, Vol. 13, No. 3, Fall 1949; William E. Daugherty and Morris Janowitz, *A Psychological Warfare Casebook*, Baltimore, Md.: Johns Hopkins University Press, 1958; Ronald De McLaurin, Carl F. Rosenthal, and Sarah A. Skillings, eds., *The Art and Science of Psychological Operations: Case Studies of Military Application*, Vols. 1 and 2, Washington, D.C.: American Institutes for Research, April 1976; and Wallace Carroll, *Persuade or Perish*, New York: Houghton Mifflin, 1948.

[52]   Author interview with Thomas Valente, June 18, 2013.

to find one, or more, that is sufficiently compelling to incorporate into a preliminary logic model for a new program or effort.

Another way to find and fill gaps in a logic model is based on operational experiences. The after-action review process is dedicated specifically to learning from both success and failure. As much as the tradition of the after-action review warrants praise for its ability to extract lessons learned from successful and unsuccessful campaigns, the approach has a major shortcoming that makes it an imperfect analogy for the assessment process: It is retrospective and timed in a way that makes it difficult for campaigns that are going to fail to do so quickly. On the other hand, JP 5-0 describes operational design as an iterative process, not just during initial planning but also during operations as assumptions and plans are forced to change. Operational design also advocates continuous learning and adaptation, and well-structured assessment can support that. As we advocate in Chapter One, fail fast! If a logic model contains uncertain assumptions, plan not only to carefully measure things associated with those assumptions but also to measure them early and often. If faulty assumptions are exposed quickly, this information can feed back into a new iteration of operational design, producing a revised logic model and operational approach.

Start Big and Prune, or Start Small and Grow

There is at least as much art as science to achieving the right level of detail in a logic model or theory of change. For example, a theory of change might begin as something quite simple: *Training and arming local security guards will lead to increased stability*. While this gets at the kernel of the idea, it is not particularly complete as a logic model. It specifies an outcome (increased stability) and some outputs (trained local security guards and armed local security guards), and further implies inputs and activities (the items needed to train and arm guards), but it does not make a clear logical connection between the outputs and the outcome. Stopping with that minimal logic model could lead to assessments that would only measure the activity and the outcome. However, such assessments would leave a huge assumptive gap. If training and arming go well but stability does not increase, assessors will have no idea why. To begin to expand on a simple theory of change, ask the questions, "Why? How might A lead to B?" (In this case, how do you think training and arming will lead to stability?) A thoughtful answer to this question usually leads one to add another node to the theory of change, or an additional specification to the logic model. If needed, the question can be asked again relative to this new node until the theory of change is sufficiently articulated.

How do you know when the theory of change is sufficiently articulated? There is no hard-and-fast rule. Too many nodes, too much detail, and you end up with something like the infamous spaghetti diagram of Afghan stability and counterinsurgency dynamics.[53] Add too few nodes and you end up with something too simple that leaves

---

[53] In 2009, GEN Stanley McChrystal, then commander of U.S. and NATO forces in Afghanistan, received a PowerPoint slide meant to convey the complexity of the coalition military strategy for counterinsurgency and

too many assumptive gaps. If an added node invokes thoughts such as, "Well, that's pretty obvious," perhaps it is overly detailed.

Elicit an Implicit Theory of Change

As noted, one challenge that can come up in logic modeling is when the inputs, activities, outputs, and outcomes are all clear, but it is not clear how the outputs are supposed to lead to the desired outcomes. This is a situation with an implicit logic of the effort, and the goal then becomes making it explicit. Faced with this situation, assessors can start by asking why and how questions (as suggested in the previous section), but it is possible that they will not be able to come up with satisfactory answers. This is particularly likely to be the case if the planner or assessor building the logic model is not expert in the area of activity, or is not intimately familiar with the specific program or activity. One way to resolve this is to engage stakeholders in the logic modeling process, or otherwise trying to elicit the implicit theory of change.[54] Presumably, those engaged in the planning and execution of a program or activity have some idea why they do the things they do. Engaging stakeholders may quickly reveal missing connections in a theory of change. However, it is also possible that while stakeholders intuit how their actions connect to desired outcomes, they have a hard time articulating it. In such a case, the theory of change remains implicit, but working with stakeholders can still bring it to light. Ask stakeholders the same kind of questions for refining logic models noted above. Begin with some specific program element and ask, "Why are you doing that?"[55] Break it down, walk through activities, and try to expose the internal logic of the effort or its shared understandings.

Specific Frameworks

There are a number of specific frameworks, worksheets, and guidebooks that can help with articulating a logic model or theory of change. We found two to be particularly relevant: The NATO Joint Analysis and Lessons Learned Centre's (JALLC's) *A Framework for the Strategic Planning and Evaluation of Public Diplomacy* and USAID's LogFrame template. We discuss each in turn.

The NATO JALLC framework provides several useful worksheets and templates that can help IIP program or effort planners and clarify goals and objectives, specify the theory of change, and devise an activity plan. The worksheets can be found in Appendix A of that document and can be downloaded from the JALLC website. Step-

---

stability operations in that country. The slide prompted two strains of commentary: one declaring that the Afghanistan strategy had gotten out of hand and another declaring that the military's use of PowerPoint had gotten out of hand. We revisit both these points in Chapter Eleven, on the presentation and uses of assessment.

[54] Donaldson, 2007, pp. 32–39.

[55] Rossi, Lipsey, and Freeman, 2004, p. 148.

by-step planning instructions from the second chapter of that document are summarized here:[56]

- In worksheet 1, the public diplomacy goals are described, along with an indicator of success for each goal.
- In worksheet 2, each of the target audiences is described and mapped to applicable goals.
- Worksheet 3A is the strategic matrix of desired impacts, which connects goals from worksheet 1 to audiences from worksheet 2.
- Worksheet 3B constructs the theory of change by mapping each desired impact to desired outcomes via the planning assumptions (what needs to happen to achieve the impact) and maps the desired outcomes to their requisite conditions.
- Worksheet 3C develops SMART key performance indicators (KPIs) for each impact, outcome, and condition articulated on Worksheet 3B, along with measures, data sources, baseline measurements, targets, and methods for assessing each KPI. It also specifies the research questions that will be used during the evaluation to investigate qualitative aspects of the desired impacts that cannot be assessed by the KPI.
- Worksheet 4 is the Public Diplomacy Activity Plan and presents the mix of communication, outreach, and engagement activities that are planned to accomplish each desired impact. For each activity, the user develops one or more KPIs, targets, objectives, and monitoring methods.
- Worksheets 5A, 5B, and 5C are the Evaluation Data Collection Plan, the Monitoring Plan Data Collection Matrix, and the Monitoring and Evaluation Data Collection Summary, respectively, which combine the complete list of data collection requirements for each of the desired impacts, providing a convenient way to plan coordinated research.

For media interventions in the international development context, many donors and sponsors require that the intervention and associated evaluation plan be placed in a logical framework matrix, frequently referred to as a "LogFrame." The LogFrame provides a structured way to specify the theory of change that links inputs (resources and activities), outputs, objectives (or purposes), and goals and to map that logic model to an evaluation plan consisting of indicators or measures and data sources. The evaluation design, methods, measures (output, outcome, and impact), and processes are reported in the LogFrame, which is agreed upon prior to the initiation of the project.[57] Figure 5.5 displays the LogFrame matrix template used by USAID.

---

56 NATO Joint Analysis and Lessons Learned Centre, 2013, pp. 19–20, A1–A5.

57 Arsenault, Himelfarb, and Abbott, 2011, pp. 17–19.

Figure 5.5
USAID's LogFrame Template

Narrative Summary	Indicators	Data Sources	Assumptions
Project goal:			
Project purpose:			Affecting the purpose-to-goal link:
Outputs:			Affecting the output-to-purpose link:
Inputs:			Affecting the input-to-output link:

SOURCE: U.S. Agency for International Development, "Logical Framework Template: Basic," web page, undated.

RAND *RR809/1-5.5*

Updating the Theory of Change

Fortunately, if an initial theory of change or stated logic of the effort is not sufficiently detailed in the right places or does not fit well in a specific operating context, iterative assessments will point toward places where additional detail is required. As assessment proceeds, whenever a measurement is positive on one side of a node but negative on the other and you cannot tell why, either a mistaken assumption has been made or an additional node is required. Following the example discussion of a logic model for increasing stability as outlined above, imagine a situation in which measures show real increases in security (reduced significant activities [SIGACTs], reduced total number of attacks/incursions, reduced casualties/cost per attack, all seasonally adjusted), but measures of perception of security (from surveys and focus groups, as well as observed market or street presence) do not correspond. If planners are not willing to give up on the assumption that improvements in security lead to improvements in perception of security, they need to look for another node. They can speculate and add another node, or they can do some quick data collection, getting a hypothesis from personnel operating in the area or from a special focus group in the locale. Perhaps the missing node is awareness of the changing security situation. If preliminary information confirms this as a plausible gap, then this not only suggests an additional node in the theory of change and an additional factor to measure but also indicates the need for a new activity: some kind of effort to increase awareness of changes in the security situation.

This example is more than just hypothetical. One of our interview respondents shared a personal operational experience with counterinsurgency in the Pacific.[58] His unit managed to make significant improvements in the security situation in his area of responsibility; personnel were operating at night and had driven the adversary from the area. However, local perceptions of the security situation were unchanged, and the population remained quite fearful. Locals did not see security forces operating, since they were doing so at night, and knew nothing of the disposition of the adversary, so they perceived the context to still be highly insecure. Finally recognizing the problem, the respondent's forces began to share information about their successful nighttime exploits and increased their presence during daylight hours. Their active efforts to promote changes in perception to match the tangible security improvements they had achieved were ultimately successful.

Improvements to the theory of change improve assessments, but they can *also* improve operations. Further, articulating a theory of change during planning allows activities to begin with some questionable assumptions in place—and with the confidence that they will be either validated by assessment or revised. Theory of change–based assessment supports learning and adapting in operations. (Again, as we advocate in Chapter One, fail fast.) This approach can also help tailor generic operations and assessments to specific contexts. By treating a set of generic starting assumptions as just that, a place to start, and testing those assumptions as hypotheses, a theory of change (and the operations and assessments it supports) can evolve over time to accommodate contextually specific factors, whether such factors are cultural, the result of individual personalities, or just the complex interplay of different distinct elements of a given environment or locale.

A preliminary theory of change might evolve not only because of the inclusion of new connective nodes but also by asking after missing disruptive nodes (disruptors). Again, by articulating the possible disruptor as part of the theory of change, it can then be added to the list of things to attempt to measure. For example, when connecting training and arming local guards to improved willingness and capability to resist insurgents, we might red-team disruptive factors such as "trained and armed locals defect to the insurgency" or "local guards sell weapons instead of keeping them." We can then add these disruptors as an alternative path on our theory of change and attempt to measure the possible presence of these disruptors. In the same way that no plan long survives contact with the enemy, logic models often require revision when exposed to reality. Iteration and evolution are important to (and expected of) theories of change.

---

[58] Author interview on a not-for-attribution basis, February 13, 2013.

Validating Logic Models

Logic models should be validated. Sometimes IIP programs or efforts are predicated on incorrect assumptions. Sometimes, IIP efforts are based on a thoughtful foundation derived from existing psychological research, but it is not applicable in the given cultural context. For example, much psychological research is based on findings from experimental results with American college students and may not be generalizable to other cultures.[59] As noted in the previous section, one way to validate a logic model is to execute based on it, revise it through trial and error, and declare it valid when it finally works. The summative evaluation for a successful effort or program validates the program's logic model.[60]

Logic models can also be validated in other ways. One such approach is similar to the formative research recommended above for building a logical model: some sort of SME engagement. If a preliminary logic model survives scrutiny by a panel of both influence and contextual experts, then it is likely to last longer and with fewer subsequent changes than a logic model not validated in this way. In JOPP, this could be part of COA analysis and war-gaming, though may require input from SMEs outside the standard staff.

Another way to validate logic models is with significant dedicated formative research—foundational research on influence or on influence in certain cultures and contexts.[61] This could be in laboratory experiments,[62] or field experiments conducted in contexts of interest.[63]

Summary

This chapter focused on two topics: how to establish and specify high-quality objectives for an IIP program or effort and how to establish and articulate a theory of change/logic of the effort for such an effort. This discussion has left several key takeaways, which we have organized into two categories: setting objectives and articulating theories of change or expressing them as logic models.

First, regarding setting objectives:

- The quality of an effort's goals directly relates to the quality of its associated assessment measures. Clearly articulated and specific goals are much easier to connect to clear and useful measures.

---

[59] Author interview with Victoria Romero, June 24, 2013.

[60] Author interview with Christopher Nelson, February 18, 2013.

[61] Author interview with Victoria Romero, June 24, 2013.

[62] Author interview with Devra Moehler, May 31, 2013.

[63] Author interview with Matthew Warshaw, February 25, 2013.

- Good inform, influence, and persuade objectives should specify the observable behaviors sought, and whom (the target audience) they are sought from.
- While there is some debate, behavioral objectives are strongly preferred over attitudinal objectives. Attitudinal changes may be included as subordinate or supporting objectives and as part of a longer chain of logic, but ultimate objectives should be some kind of consequential behavioral change.
- A program's theory of change contains assumptions about how the world works and what kinds of activities will lead to desired goals and why. Assessment can help distinguish between theory failure (one or more of the assumptions is wrong) and program failure (the program is not being executed properly); assessment can also help identify ways to correct either of these failings.
- Good objectives are SMART: specific, measurable, achievable, relevant, and time-bound.
- Good objectives need to at least imply what failure would look like. How will you know if you have not succeeded?
- Breaking objectives into smaller "bite-sized" incremental subordinate objectives can make it easier to articulate a logic model or theory of change and make it possible to demonstrate incremental progress.

Second, regarding articulating theories of change or expressing them as logic models:

- Efforts to inform, influence, and persuade differ from kinetic efforts in many important ways. Because military planners more perfectly intuit the relationships between actions and outcomes in the kinetic realm, shortcuts preserve meaning and are effective. However, because the social and psychological processes required of influence efforts are not part of standard military intuition, it is important that connections (and assumptions) be explicitly spelled out.
- Specifying a theory of change involves identifying overall objectives—and the inputs, outputs, and processes necessary to achieve those objectives—and describing the logic that underpins it all (an explanation of how the proposed actions will lead to the desired outcomes). A logic model is one structure for presenting a theory of change.
- In addition to describing the logical connections between activities and objectives, a good theory of change should include possible barriers, disruptors, threats, or alternative assumptions. If things that might divert progress along the logical path and prevent objectives from being achieved are identified at the outset, then their possible presence and impact can be included in the assessment process and needed adjustments can ensue.
- Formative research can help with planning and can also support development of a theory of change.

- In the same way that no plan long survives contact with the enemy, logic models often require revision when exposed to reality. Iteration and evolution are important to (and expected of) theories of change.
- There are a number of different frameworks and templates in use in different industries that can support logic modeling. There are also a number of techniques or tricks that help when developing a logic model. For example, begin with a literature review; synthesize multiple existing theories; start with a long logic model, then prune; start with a short logic model, then elaborate; work backward; and involve stakeholders and users in logic modeling.
- Logic models should be validated. This can be accomplished through SME engagement, through other research efforts, or through trial and error as part of assessment within a program of activities.
- When the program does not produce all the expected outcomes and one wants to determine why, a logic model (or another articulation of a theory of change) really shines.

From Logic Models to Measures: Developing Measures for IIP Efforts

This chapter addresses the processes and principles that govern the development of valid, reliable, feasible, and useful measures that can be used to assess the effectiveness of IIP activities and campaigns. The development of measures is decomposed into two broad processes: first, deciding which constructs are essential to measure, and second, operationally defining the measures. The chapter begins by defining the hierarchy of terms associated with measure development and identifying the types of measures that IIP assessment stakeholders are likely to encounter or employ. It then discusses techniques for identifying the constructs worth measuring, including the role of the logic model and underlying theories of change. The balance of the chapter addresses the desired attributes of measures and best practices for constructing valid and feasible measures to capture the constructs identified as worthy of measurement.

Hierarchy of Terms and Concepts: From Constructs to Measures to Data

In evaluation research, *data* are generated to *measure* the *variables* that represent the *constructs* we are interested in studying. The construct is the abstract idea or concept we want to measure, such as health, sentiment, economic well-being, religiosity, satisfaction, or violence. In program evaluation, constructs include the program outputs and outcomes and other mediating factors that ought to be measured to capture program effects. Constructs are not in themselves directly observable but can be represented or operationally defined by *variables*. A variable is a characteristic or event associated with the construct that varies across different individuals or times when measured. The measure, or the process by which the variable is measured, is the operational definition of the variable. Variables are measured by one or more *data items*. Table 6.1 illustrates the hierarchy of measurement terms using three examples with survey-based data-generating processes.

Producing usable data from measures involves several components with varying complexity, depending on the data-generating process (e.g., surveys, direct observations, tests). The construct or phenomena you seek to observe must first be captured

Table 6.1
From Constructs to Measures to Data: Three Survey-Based Examples

Construct	Variable	Measure (Operational Definition)	Data Item Example
Unit cohesion	Solidarity within military units, directing toward common goals	Platoon Cohesion Index (20 items)	How important is each of the following to first-term soldiers in your platoon?
Health	Functioning, including six domains of health	Short-form health survey (36 items)	Would you say your health is excellent, very good, good, fair, or poor?
Income	Total household income	National opinion research center/general social survey	What was your total household income, before taxes, from all sources in 2012?

through the use of assessment tasks, including written tasks (e.g., filling out a survey, taking a test) or operational tasks (e.g., behavioral observation, exercises, drills, games). The assessment task may then need to be *scored* by an external rater and subsequently *adjusted* to account for variations in the task. This process produces a *metric*, the yardsticks of the measure, such as the *number* or *rate* of incidents, the *degree* or prevalence of a belief, or the *time* required to complete a process. Metrics are aggregated to produce data.[1] These components are illustrated in Figure 6.1.

Figure 6.1
Measure Components
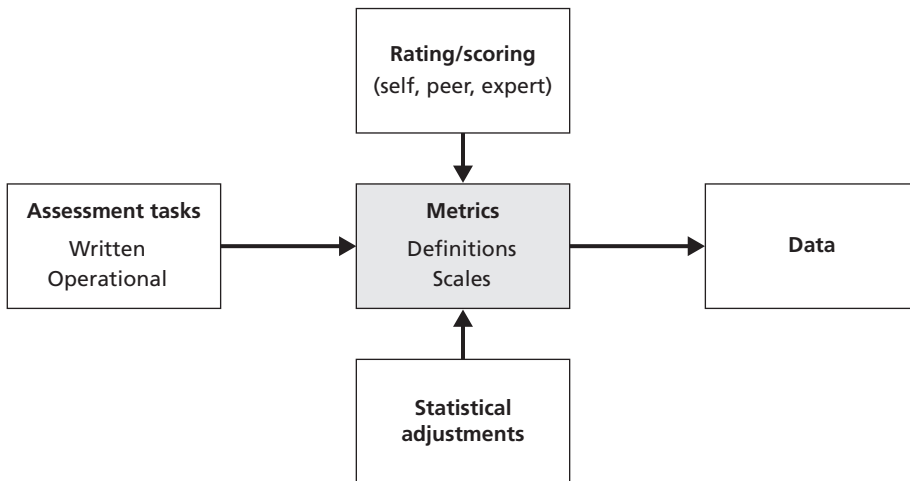
SOURCE: Derived from author interview with Christopher Nelson, February 18, 2013.
RAND *RR809/1-6.1*

---

[1]   Author interview with Christopher Nelson, February 18, 2013.

Types of Measures

Good evaluations capture a spectrum of measures along the logical sequence from inputs to outputs to behavioral outcomes and system-level change. Ideally, the assessment should include a measure to gauge every cause-and-effect relationship specified in the program logic model. If you do not have a measure at each step, "you'll never be able to walk the dog back and make a causal connection."[2]

DoD assessment doctrine emphasizes the distinction between measures of performance (MOPs) and measures of effectiveness (MOEs). MOPs include input, process, and output measures. DoD defines an MOP in the information environment as a "criterion used to assess friendly actions that is tied to measuring task accomplishment" and that "describes what and how . . . forces need to communicate to achieve the desired effect."[3] Input measures capture the extent to which the necessary resources are in place to implement a project (e.g., units and associated personnel, air time, pamphlets). Process measures capture whether a campaign or activity is progressing on time and as planned. Output measures capture the immediate or direct products of a particular activity (e.g., commercials aired, pamphlets distributed).[4]

MOEs, by contrast, are concerned with program outcomes and impacts. According to DoD guidance for inform and influence activities, an MOE is a "criterion used to assess changes in system behavior, capability, or operational environment that is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect" and that "describes what the specific target (audience) needs to do to demonstrate accomplishment of a desired effect."[5] MOEs should relate to the effect, not the tasks used to create the effect.[6]

Some organizations refer to these measures as *key performance indicators* (KPIs). JALLC's *Framework for the Strategic Planning and Evaluation of Public Diplomacy* defines a KPI as a measure of achievement against a planned objective that is "directly linked to a desired impact or desired outcome and is generally represented by a numeric value."[7]

In IIP evaluation, MOEs or KPIs are typically associated with attitudinal and behavioral changes at the individual and group levels. Whether attitudinal change constitutes an effect is controversial (see the discussion in Chapter Five under the heading "Behavioral Versus Attitudinal Objectives"), which demonstrates a limitation to the MOP-versus-MOE construct. Christopher Rate and Dennis Murphy argue that

---

[2] Author interview on a not-for-attribution basis, February 20, 2013.

[3] U.S. Joint Chiefs of Staff, 2011b; Headquarters, U.S. Department of the Army, 2013a, p. 7-4.

[4] Haims et al., 2011.

[5] Headquarters, U.S. Department of the Army, 2013a, p. 7-3.

[6] The Initiatives Group, 2013, p. 18.

[7] NATO, Joint Analysis and Lessons Learned Centre, 2013, pp. 6, 18.

excluding attitudinal change from the scope of MOE is myopic and leads evaluators to ignore short-, intermediate-, and long-term effects "that may suggest that the influence campaign is impacting the target audience towards a desired condition." They point out that most influencers affect behavioral change indirectly through "processes in the cognitive domain of the information environment."[8]

Borrowing from the education evaluation literature, *value-added* measure may be a more useful term for capturing the contribution of an IIP activity to an end state. Whereas the MOE concept does not distinguish between changes in desired outcomes that are due to the program and changes that may be due to something else, value-added measures capture only those changes that the intervention is responsible for.[9] Moreover, this framing more appropriately reflects the causal relationship between IIP activities and end states. IIP activities contribute (i.e., add value) to outcomes but are unlikely to wholly cause them.

While appreciating the conceptual differences between measure types can be valuable, assessment reports should avoid being overly concerned with the difference between MOPs and MOEs, because this focus is overly narrow and potentially distracting. In reality, there is a spectrum of measure "types," and the MOE-MOP dichotomy can mislead evaluators into thinking that there are only two relevant measures. In the discussion of MOPs and MOEs in joint doctrine (JP 5-0), MOPs are characterized as helping to answer the question, "Are we doing things right?" While MOEs help answer the question, "Are we doing the right things?"[10] This is an inappropriately simplistic view, however. While connecting MOPs to adequate execution is fine, connecting outcome measurement (MOEs) exclusively to "doing the right things" ignores a host of other possible factors, disruptors, and contextual conditions. While MOPs and MOEs conceived in this way can help discriminate between program failure and theory failure (see the discussion in Chapter Five), if theory failure is indicated (MOPs are strong, and MOEs are weak), it is important to remember that there is more to the theory (and the theory of change, ideally) than just the activities to be performed. Adversary action might be interfering, or there may be some other slight theoretical disconnect that could be easily remedied. A premature jump to concluding that the wrong things are being done on the basis of a single MOE might lead to the termination of an otherwise promising effort.

Moreover, a preoccupation with the assessment lexicon can distract from central issues and challenges regarding the "how" of assessment. Jonathan Schroden argues that this focus contributes to doctrinal deficiencies for assessment, a primary reason behind the failure of operations assessment. He observes that JPs 3-0 and 5-0 "mainly

---

8   Christopher R. Rate and Dennis M. Murphy, *Can't Count It, Can't Change It: Assessing Influence Operations Effectiveness*, Carlisle Barracks, Pa.: U.S. Army War College, March 14, 2011, p. 9.

9   Author interview with Christopher Nelson, February 18, 2013.

10   U.S. Joint Chiefs of Staff, 2011a.

focus on making clear the distinctions between [MoEs] and [MoPs]. Nowhere do they discuss in detail *how* to do operations assessment. Thus, to a practitioner they provide little more than a beginner's lesson in vocabulary."[11]

Identifying the Constructs Worth Measuring: The Relationship Between the Logic Model and Measure Selection

As highlighted in Chapter Three, the quality of the measures will depend to a great degree on the quality of the influence objectives and associated intermediate objectives articulated during the planning phase. In this sense, measures and objectives are two sides of the same coin, and the principles guiding the development of objectives and logic models can be recast as guiding the development of the measurement system. Chapter Five discussed the attributes of objectives, and the logic models for achieving those objectives, that facilitate effective measurement and assessment. This section calls on many of those concepts, bridging the connection between the logic model specification for planning purposes and the measure selection process for assessment purposes. This and subsequent sections seek to answer the following question: "Of all of the cause-and-effect relationships that were specified in the planning phase, which are essential to measure?"

Separating what is important to measure from what is less important "is what measure development is all about."[12] The program logic model provides the framework for selecting the constructs that are worth measuring, but evaluators should not assume that all important measures will simply fall into their laps in the course of planning. As Christopher Nelson pointed out, goals and objectives can be unclear or unmeasurable, and program managers often disagree on the ultimate goal that a program is designed to serve.[13] Moreover, it is too costly to measure every cause-and-effect relationship and mediating variable within the system that ties program inputs to outputs to outcomes.

The importance of measuring something, or the information value of a measure, is a function of uncertainty about its value and the costs of being wrong. When identifying constructs worth measuring, assessors should therefore give priority to load-bearing and vulnerable cause-and-effect relationships within the logic model. These can be identified by drawing on IIP theories, empirical research, expert elicitation, and rigorous evaluations of similar programs implemented in the past.[14] Moreover, the information value of a measure takes precedence over its validity and reliability. Even the most valid and reliable measurement instruments cannot improve the value of the

---

[11] Schroden, 2011, p. 92.

[12] Author interview with Christopher Nelson, February 18, 2013.

[13] Author interview with Christopher Nelson, February 18, 2013.

[14] Author interview with Christopher Nelson, February 18, 2013.

measure if it is measuring a construct that is irrelevant to assessment stakeholders and the decision they need to make. Assessors should therefore try to measure every truly important variable even if the measurement instrument has weak validity. Douglas Hubbard emphasizes this point in *How to Measure Anything*: "If you are betting a lot of money on the outcome of a variable that has a lot of uncertainty, then even a marginal reduction in your uncertainty has a computable monetary value."[15] For example, if the planned operation is a demonstration of force (such as bombers flying near an adversary's airspace or sailing a carrier battle group near an adversary's territorial waters) to dissuade the adversary from a certain COA, there are some things that are very important, but also very difficult, to measure. The MOPs are easy, and almost irrelevant; it is trivial to know whether the sorties or the fleet movements have been executed as planned. It should be easy to observe the core outcome, whether or not the adversary has chosen the dispreferred COA.

The steps that are hard to measure (but any sort of measure might help follow the process and decrease uncertainty) pertain to the logical nodes connecting the demonstration of force with the adversary's decisional result. First, did the adversary notice the demonstration? Were blue forces surveilled with radar? Second, if observed, did the adversary take note of the demonstration? Does intelligence report increased traffic between adversary higher headquarters and the forces that might have monitored the demonstration? Did adversary forces approach or challenge (visually or verbally) blue forces? Do adversary representatives publicly decry the provocation? If so, they got the message! Third, how was the demonstration perceived by adversary decisionmakers? This is particularly difficult to measure, as it requires observation of a relatively small number of hard-to-access individuals, but any kind of information that could help reduce uncertainty could be of value. Perhaps intercepted communications could provide hints, as might public statements from adversary representatives. Finally, how did the demonstration affect the decisional process? Again, this is very hard to measure, barring serendipitous intercepts of decisional communications, or a highly placed human intelligence (HUMINT) source, but potentially incredibly illuminating. Hints of information about the decisional process might help planners decide whether a further demonstration would help or hurt the effort, or whether some additional and different IIP efforts might further contribute, and how.

Capturing the Sequence of Effects, from Campaign Exposure to Behavioral Change

IIP summative evaluations should include a measure of exposure to the campaign and several measures that capture the internal processes by which exposure influences behavioral change. As introduced in Chapter Five, the three major internal processes that many IIP evaluations measure are changes in knowledge, attitudes, and practices or behavior (commonly referred to as the KAP construct). However, the exact pro-

---

[15] Hubbard, 2010, p. 36.

cesses that need to be measured will be determined by the hierarchy or sequence of effects, from exposure to behavioral change, articulated or implied by the theories of change motivating the program logic model. For international development programs, for example, the specific measures depend on the Logical Framework, or Log-Frame, governing the program (see the discussion in Chapter Five).[16]

Knowledge, attitudes, and behaviors are themselves composed of intermediate processes, so it is useful to decompose the KAP construct into a hierarchy or sequence of several discrete and measurable effects along the path from exposure to sustained behavioral change. The more processes that are measured, and the more measures employed to gauge each process, the more confidence evaluators can have in the estimated effects of the IIP activity and the better the researchers will be able to understand how to improve the efficacy of the intervention in future iterations.[17] Consider the example from the preceding section concerning using a demonstration of force as a deterrent to certain adversary behaviors. Because of the host of possible interpretations of, or responses to, such a demonstration, it is key to spell out the intended sequential changes in knowledge, attitudes, and behaviors and check them against what actually occurs. Measures like the ones suggested above may clearly indicate that adversary decisionmakers are aware of the demonstration (a knowledge element), and that they rightly perceived it as a threat and are concerned (attitude), but they may not know what to do or not do in response (failure to reach desired behavior because of an omitted knowledge element). If that is the case, the desired results might be realized with a small addition to the effort—a public or private statement of demands, or requests, of the adversary, so that decisionmakers know what is wanted from them.

There are many other relevant outcomes for IIP campaign evaluation beyond knowledge, attitudes, and behaviors. Mediators and processes of behavioral change that should be measured, if possible, include knowledge/awareness, salience, attitude, norms, self-efficacy, behavioral intentions, behavior, behavioral integration, skills, environmental constraints, media change, and policy change.[18] Because individual behaviors can rarely be directly observed when targeting a large group, Martin Fishbein and colleagues contend that the most-important outcome measures for communication campaigns for behavioral change are attitudes toward the behavior, norms about the behavior, and behavioral intention.[19]

---

[16] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[17] Author interview with Ronald Rice, May 9, 2013.

[18] Charles K. Atkin and Ronald E. Rice, "Advances in Public Communication Campaigns," in Erica Scharrer, ed., *The International Encyclopedia of Media Studies*, Vol. 5, London: Wiley-Blackwell, 2013; interview with Thomas Valente, June 18, 2013.

[19] Martin Fishbein, Harry C. Triandis, Frederick H. Kanfer, Marshall Becker, Susan E. Middlestadt, and Anita Eichler, "Factors Influencing Behavior and Behavior Change," in Andrew Baum, Tracey A. Revenson, and Jerome E. Singer, eds., *Handbook of Health Psychology*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2001.

The hierarchy of effects in William McGuire's input-output communication matrix (see Appendix D) is a useful, illustrative model for identifying the constructs worth measuring (though it is certainly not the only one). Table 6.2 shows 12 such constructs derived from McGuire's hierarchy. As shown, program designers and evaluators can obtain estimates of expected effects by specifying the logical sequence, expected exposure levels (50 percent in this example), and expected rate at which the campaign moves individuals between steps (75 percent in this example; in reality, this is highly variable and changes with each step).

Upstream and Downstream Measures

When choosing which of a host of nodes in a theory of change or connections in a logic model to measure, priority should be given to load-bearing elements—elements that are central inputs, outputs, or outcomes—or those that are vulnerable, either because they contain or relate to untested assumptions or are vulnerable to outside forces (either contextual factors or adversary action). Also, priority should be given to measures that help reduce uncertainty in decisionmaking.

Table 6.2
12 Constructs to Measure from McGuire's Hierarchy of Effects Model

Category	Construct to Measure	Notional Cumulative Success Rate (75%)
Exposure	1. Exposed to message	50%
	2. Recalls message	38%
Knowledge	3. Comprehends message	28%
	4. Knows how to change behavior	21%
Attitudes and behavioral intention	5. Likes message	16%
	6. Considers the message important (saliency)	12%
	7. Recognizes positive impact of behavior	9%
	8. Believes they can change behavior (self-efficacy)	7%
	9. Intends to practice behavior	5%
Behavior	10. Begins to practice behavior	4%
	11. Experiences benefits of behavioral change	3%
	12. Sustains behavior/proselytizes to others	2%

SOURCE: Adapted from Valente, 2002, p. 41, and William J. McGuire, "Theoretical Foundations of Campaigns," in Ronald E. Rice and Charles K. Atkin, eds., *Public Communication Campaigns*, 2nd ed., Newbury Park, Calif.: Sage Publications, 1989.

NOTE: Notional success rates assume a 75-percent effectiveness rate, defined as the rate that the campaign moves individuals between steps.

Another consideration in prioritizing measures is choosing the appropriate mix of *downstream* and *upstream* measures or variables. *Downstream* measures capture the elements of the logic model that are those closer to the desired outcomes, or the end of the logical sequence of effects. Downstream measures are helpful because they place more attention on what stakeholders ultimately care about, and they allow reasonable variation in how outcomes are attained. *Upstream* measures capture the processes that are closer to program inputs or outputs and further from desired outcomes. Upstream measures are valuable because they avoid the issue of "coproduction" of outcomes by focusing only on what the program can control, they are often more feasible and cost-effective, and they test assumptions about the carrying capacity of the program environment.[20] While it is optimal to measure each step along the sequence of effects, doing so is not always feasible or cost-effective. If everything in an effort or program goes well, then downstream measures are more attractive; if, however, something is wrong and the effort is not performing as expected, upstream measures are more likely to support the identification and remediation of the problem.

Recall the example from Chapter Two of an effort to recruit partner-nation police. The most obvious downstream measure is the outcome, the actual number of new recruits applying to join the police. If applications are up to the target threshold, then additional measurement is not useful. However, if there has been no increase in applications, or a modest increase that does not meet the minimum success criteria, then upstream measures (measures related to the distribution and reach of recruitment materials, or attitudinal data from surveys or focus groups about how potential recruits viewed these materials, for example) might help determine why. This is why it is important to collect upstream measures.

Sometimes, downstream outcomes cannot be readily observed or take years to become evident. In this case, evaluators can use logic models to identify the key upstream or *intermediate* variables that correlate with the likelihood of achieving the unobservable outcome. Measuring upstream variables (e.g., exposure) can be sufficient to measuring outcomes if the causal linkages specified in the logic model between upstream factors and the outcome of interest have been validated by empirical research.[21] This might be the case, for example, if a campaign promoting respect for human rights has been successful (and successfully monitored) in three regions of a country, and is being applied in a fourth region. The theory of change/logic of the effort has been validated in that context, and upstream success (successful implementation and execution) can be safely assumed to be leading to downstream success.

Discussions concerning the appropriateness of using upstream versus downstream measures can be seen in debates surrounding the importance of measuring reach. Some argue that there is an overemphasis in the literature from marketing and public diplo-

---

[20] Author interview with Christopher Nelson, February 18, 2013.

[21] Author interview with Christopher Nelson, February 18, 2013.

macy on measuring reach. Katherine Brown, a former public affairs officer at the U.S. embassy in Kabul, and Phil Seib, the director of the Center for Public Diplomacy at the University of Southern California, both criticized program managers for mistaking reach for impact when selling their programs.[22] Amelia Arsenault claimed that this is due to government organizations that depend on "perfunctory" indicators such as the number of people reached rather than true measures of effect.[23]

Others argue that measuring audience can be sufficient to measuring outcomes if the logic model is validated. Kim Andrew Elliot, an audience research analyst with the U.S. International Broadcasting Bureau, explained that measuring *audience* can be sufficient for measuring *outcomes* if content is validated through formative research methods, such as product testing and focus groups.[24] Mark Helmke, a former professional staff member on the Senate Foreign Relations Committee with responsibility for overseeing strategic communication programs, echoed this sentiment in the defense context: "If your messages are validated as having the right effects with strategic audiences, all that matters is getting it out."[25]

Behavioral outcomes can also be estimated by measuring validated *mediators* of behavior. Thomas Valente illustrates this point with the example of a youth tobacco-prevention health communication campaign that aims to rebrand cigarette smoking as "uncool." If formative research demonstrates that the primary reason kids start smoking is that they perceive cigarettes to be cool, then the evaluation just needs to measure the extent to which the intervention is changing perceptions of smoking to approximate the effect of the campaign or test whether the intervention is changing perceptions of smoking to approximate the effect of the campaign on the incidence of tobacco use among youths.[26]

This discussion reinforces the critical importance of using formative and empirical research to validate the program logic model. The "best way," according to Valente, to identify predictive upstream variables or other mediators of behavioral change is to conduct good formative research.[27] These causal linkages can be validated to various degrees by many methods, including controlled laboratory experiments, empirical analyses of past interventions, behavioral observation, and qualitative methods like focus groups and in-depth interviews. For more on logic model development and validation, see Chapter Five, on logic model development, and Chapter Eight, on formative research methods.

---

[22] Author interview with Phil Seib, February 13, 2013; interview with Katherine Brown, March 4, 2013.

[23] Author interview with Amelia Arsenault, February 14, 2013.

[24] Author interview with Kim Andrew Elliot, February 25, 2013.

[25] Author interview with Mark Helmke, May 6, 2013.

[26] Author interview with Thomas Valente, June 18, 2013.

[27] Author interview with Thomas Valente, June 18, 2013.

Attributes of Good Measures: Validity, Reliability, Feasibility, and Utility

The previous section addressed the selection of constructs or phenomena worth measuring. Once the evaluators have identified and prioritized the outcomes, outputs, and other constructs worth measuring, they must be operationalized with measures that capture variability associated with the construct. This section discusses the desired attributes of measures that should guide the measure-development process.

The quality of a measure is typically evaluated on the basis of its *validity*, *reliability*, *feasibility*, and *utility*. Validity and reliability represent the two types of measurement errors. Validity is the correspondence between the measure and the construct, or freedom from systemic error (bias). Reliability is the degree of consistency in measurement, or freedom from random error (e.g., signal to noise). Feasibility is the extent to which data can actually be generated to populate the measure with a reasonable level of effort. Utility is the usefulness of the measure to assessment end users and stakeholders.[28]

Assessing Validity: Are You Measuring What You Intend to Measure?

Measurement or *instrument validity* is the degree to which a variable represents the concept it is intended to measure.[29] Validity can be assessed on several dimensions. *Face validity* asks whether the measure subjectively measures what it purports to measure, or whether an untrained observer would perceive it as obviously capturing the construct. *Discriminant validity* asks whether the measure can discriminate between constructs. If a measure could be seen as capturing several unrelated constructs, the measure has low discriminant validity. Conversely, *convergent validity* asks whether the measure overlaps with other measures that capture the same construct. If a measure is trending synchronously with another validated measure of the same construct, it has high convergent validity.

Convergent validity is particularly important for assessing the quality of measures used in IIP evaluations. Because there are significant limitations to the quality of and quantity of data associated with any particular MOE in the information environment, the most valid measures of success are those that converge across multiple quantitative and qualitative data items. This highlights the need to triangulate multiple methods and measures. It is easy to identify weaknesses with any single measure, but when a collection of measures suggests the same general trend, it is easier to have confidence in the conclusions.[30]

---

[28] Author interview with Christopher Nelson, February 18, 2013.

[29] Valente, 2002, pp. 89–90.

[30] Author interview with Steve Booth-Butterfield, January 7, 2013.

Consider an example in which the construct of interest is the level of security within an area. One possible measure would be the number of attacks or incidents reported within the area. This measure has *face validity*, as it makes sense to connect security with incidents. A skeptic, however, might ask about the mechanism of reporting or observing incidents, and might propose an alternative explanation: Perhaps the number of incidents remains the same, but reporting is diminishing because of intimidation by threat elements and lack of confidence in authorities. Number of attacks or incidents reported lacks *discriminant validity*, as it does not discriminate between the construct "security increased" and the construct "security remained the same, intimidation increased." However, this concern could be ameliorated if incidents reported were considered in concert with other measures that have convergent validity. Perhaps reduced reports of attacks is joined by reduced volume of threat-communication traffic in the region and by survey data indicating that the population feels more secure and is more likely to report incidents to authorities.

Some SMEs argued that it is important to think not just about the constituents of technical validity but also about the *political validity* of a measure, the credibility associated with the measurement instrument to audiences and stakeholders. Political validity is often discussed in the context of high-stakes education evaluation. Tarek Azzam, an assistant professor at Claremont Graduate University who focuses on the real-world application of evaluation efforts, identified several factors that contribute to political validity, including the stakes surrounding the measure and ingrained preferences for the measure, which relate to how the measure has been used historically.[31] For example, congressional representatives are well aware of the strengths and weaknesses of opinion polls (from their experiences during election campaigns) and may be particularly skeptical of IIP assessment reporting if too much weight is put on survey research, or if a survey has not been conducted with sufficient rigor.

Assessing Reliability: If You Measure It Again, Will the Value Change?

A measure is reliable if you would get the same result if you measured the same subject or phenomenon over and over again in the same way. Measures can be unreliable if the meaning of the measure behaves differently for different groups (common in IIP), if it is difficult for the respondent to choose the right response, or if the data collection and managing process is unreliable (e.g., lack of standardization between interviewers or coders). Concerns over measurement reliability highlight the advantages of using multi-item measures or scales. Scales have higher reliability than single-item measures because multiple items can correct for poor reliability in one or more of the items.

Test and retest reliability is measured by administering the same instrument to the same group of individuals twice at different points in time and correlating the scores. It is useful for measuring the reliability of measures populated by data from survey

---

[31] Author interview with Tarek Azzam, July 16, 2013.

instruments. *Interrater reliability* assesses the degree to which different observers agree in their ratings of the same phenomenon when using the measure or instrument. It is useful when the measurement procedure depends on human observation.

Because assessing the achievement of most IIP outcomes depends on subjective judgment, interrater is an important and useful gauge of the specificity of influence objectives and their associated measures. Steve Booth-Butterfield illustrated how DoD may use interrater reliability to gauge the specificity and measurability of an influence objective. The influence objective (and associated measure) should be defined well enough that if you gave the definition to ten different observers, and showed them video tapes of people engaging in ten different behaviors, eight to nine of the observers would agree on which video showed an action that represented the achievement of the objective.[32]

Assessing Feasibility: Can Data Be Collected for the Measure with a Reasonable Level of Effort?

A measure is feasible if data of sufficient quality can be collected for the measure with a reasonable level of effort. Feasibility can be assessed by fully mapping out the process by which data will be generated, as discussed earlier and illustrated in Figure 5.2 in Chapter Five. Feasibility will depend on accessibility, the amount of technical assistance that will be required, and the degree to which data collection is aligned with existing measurement practices or systems.[33] An associated implication is that the measure or indicator must be defined clearly enough such that it implies the type of data needed to be collected for evaluation.

The feasibility of generating data is an underappreciated criterion for measure development. As Nelson noted, "Selecting and developing theoretical measures is relatively easy; finding data to populate the measures is much harder." Developing feasible and sustainable data-generating processes is expensive and involves coproduction by numerous semiautonomous actors. Measuring upstream factors requires the willingness to make internal processes widely visible.[34]

Assessing Utility: What Is the Information Value of the Measure?

The utility of the measure gauges the information value that the measure provides to end users and stakeholders. Utility is assessed by asking whether the data or results are actionable—for example, whether there are clear linkages from the information provided by the results of the measure to decisionmaking levers, and whether the information is perceived as useful to individuals with the "wills, skills, bills," and opportunities

---

[32] Author interview with Steve Booth-Butterfield, January 7, 2013.

[33] Author interview with Christopher Nelson, February 18, 2013

[34] Author interview with Christopher Nelson, February 18, 2013.

needed to act.[35] In *Data-Driven Marketing*, Mark Jeffery suggests using the "80/20" rule to select measures with the highest utility: Determine "the 20 percent of data that will give 80 percent of the value," and focus on generating those data first.[36]

Experts from the marketing and social marketing sectors often lament the poor information quality coming from social media. "If the golden rule of business measurement is 'measure what matters,'" says Olivier Blanchard, "the golden rule of social media measurement is 'just because you can measure it doesn't mean that it matters.'"[37] However, if the cost of collecting the data is low, it may be worth collecting data on a broad swath of measures. Hubbard points out that, while most of the variables have an "information value" near zero, "usually at least some variables have an information value that is so high that some deliverable measurement effort is easily justified."[38]

Feasibility Versus Utility: Are You Measuring What Is Easy to Observe or Measuring What Matters?

There is often tension between the feasibility of a measure and its utility. Often, what is important or useful to measure cannot be easily observed or cannot be observed in the near term. The danger, according to Nicholas Cull, historian and director of the master's program in public diplomacy at the University of Southern California, is that because it cannot be evaluated easily, it will not be done, so program designers will just go after the low-hanging, easy-to-measure but less important outputs.[39] Simon Haselock, founding director of Albany Associates and former NATO spokesman in Sarajevo, has observed this dynamic playing out in the field. For example, in response to questions about how much progress is being made, program managers may give answers like, "We trained 50 journalists" because "numbers are easier than communicating the complexity of the situation."[40] Stephen Downes-Martin describes this problem as "blinkered metrics collection." Assessment cells will often identify up front which metrics are hard to collect and then set them aside without regard for their importance.[41] Recall the example of the demonstration of force earlier in this chapter. There were several constructs that were easy to measure: the inputs (the MOP about the conduct of the demonstration) and the outcome (did the adversary follow the averred COA). The most important constructs, however, are much harder to collect,

---

[35] Author interview with Christopher Nelson, February 18, 2013.

[36] Mark Jeffery, *Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know*, Hoboken, N.J.: John Wiley and Sons, 2010, p. 23.

[37] Blanchard, 2011, p. 32.

[38] Hubbard, 2010, p. 36.

[39] Author interview with Nicholas Cull, February 19, 2013.

[40] Author interview with Simon Haselock, June 2013.

[41] Downes-Martin, 2011, p. 108.

including downstream measures about adversary decisionmakers' perception of the demonstration and their decisional deliberations. The easy constructs should certainly be measured, but some effort should also be made to capture information about those constructs that are more informative and difficult to measure.

To improve the information value of measures, evaluators should treat feasibility as a necessary but not sufficient condition and should avoid the temptation to only measure what is easy to observe. Constructs are worth measuring if knowing their value reduces uncertainty about the effects of the program.[42] Assessors should *first* identify the constructs worth measuring and *subsequently* determine what *can* be measured given resource and environmental constraints.

Desired Measure Attributes from Defense Doctrine

DoD and NATO doctrine emphasize the practical attributes of measures, such as feasibility and utility, over the technical or academic qualities—i.e., validity and reliability. The Joint Chiefs of Staff *Commander's Handbook for Assessment Planning and Execution* states that measures should be *relevant*, *measurable*, *responsive*, and *resourced*. *Relevant* measures are those that can inform decisions associated with the operation. *Measurable* measures have qualitative or quantitative standards or yardsticks that they can be measured against and should have a baseline collected prior to execution. *Responsive* measures detect situational changes quickly enough to enable timely responses by decisionmakers. *Resourced* measures are those for which resources are planned and available for data collection and analysis.[43]

The *NATO Operations Assessment Handbook* identifies separate attributes for MOPs and MOEs. MOPs help managers determine whether actions are being executed as planned and therefore must be directly tied to a specific action rather than other elements of the plan. MOEs help managers determine whether the program is on track to achieve the desired end states and therefore must be repeatedly measured across time to determine changes in system states.[44] The handbook distinguishes between attributes that a measure *must have* and those that it *should have*. These necessary and desired attributes are identified in Table 6.3.

Constructing the Measures: Techniques and Best Practices for Operationally Defining the Constructs Worth Measuring

Once the evaluators have identified and prioritized the outcomes, outputs, and other constructs worth measuring, they must be operationalized with measures that capture

---

[42] Hubbard, 2010, p. 36.

[43] U.S. Joint Chiefs of Staff, 2011c, p. III-6.

[44] NATO, 2011, p. 3-3.

Table 6.3
Necessary and Desired Attributes of MOPs and MOEs from the NATO Assessment Handbook

MOPs Must . . .
☑ MOPs only: Align to one or more own-force actions
☑ Describe the system element or relationship of interest that must be observed
☑ Be observable in a manner that produces consistent data over time
☑ Describe as specifically as possible how the action is to be executed (MOP) or how the element is expected to change (MOE)
☑ Be sensitive to change in a period of time meaningful to the operation
☑ Have an associated acceptable condition
☑ MOPs only: Have a known deterministic relationship to the action
☑ MOEs only: Be culturally and locally relevant

Measures Should . . .
☑ Be reducible to a quantity (as a number, percentage, etc.)
☑ Be objective
☑ Be defined in sufficient detail that assessments are produced consistently over time
☑ Be cost-effective and not burdensome to the data collector
☑ Have an associated rate of change
☑ MOEs only: Have appropriate threshold(s) of success or failure

SOURCE: Adapted from NATO, 2011, pp. 3-5–3-6.

the variability associated with the construct. The last section presented the criteria that define good measures and that should guide the measure-development process. This section explores the techniques for applying those criteria to define and field valid, reliable, feasible, and useful measure that capture key constructs for evaluating IIP campaigns. This section presents some best practices to implement, key points to consider, and general recommendations.

Best practices:

- Assemble real or virtual panels or semistructured workshops with subject-matter and evaluation experts.
- Hold simulations, exercises, premortems, and clarification workshops with program managers and stakeholders.
- Review rigorous evaluations of similar campaigns implemented in the past.
- Review historical texts or memoirs to find creative proxies for influence used in the field.
- Avoid measures that can be easily manipulated by the program being assessed.
- Include flexible measures that can capture unintended consequences.

- Consider the incentive or accountability system to avoid developing measures that create perverse incentives, such as "teaching to the test."
- Use several outcome measures per influence objective without overdoing it.
- Where appropriate, use numeric proxies for influence, such as Klout Scores.
- If the measure is numeric, express it in terms of a ratio to more easily interpret change from the baseline.

Key points to consider:

- Define the measure such that it captures failure as well as success.
- For exposure measures, the measure denominator should be the target or strategic audience, rather than the population at large.
- More is not always better and measures such as "number of engagements" may stress the carrying capacity of the program or the recipient audience.
- The affiliated data collection management plan, including sampling rates, should be implied or specified along with the measure definition.
- DoD IIP MOEs should specify the desired direction of change, the target audience, what DoD is trying to influence the target audience to do, and the numerical percentage threshold of effectiveness.
- DoD IIP MOPs should specify the message quantity (e.g., number of broadcasts or deliveries), medium, delivery, and target audience.

General recommendations:

- Develop a repository or clearinghouse of validated IIP measures.
- Maintain a "wiki" casebook of IO campaigns that were known successes or failures.
- Select measures that can produce usable data and specify or imply the data-generating process in the measure definition.

Operationally defining key outcomes and outputs requires a precise understanding of what program managers and stakeholders mean by their stated objectives. Vague objectives are commonplace in DoD and complicate efforts to operationally define them with valid measures. Hubbard encourages the use of "clarification workshops." These workshops begin with stakeholders stating their initial, ambiguous objective. The evaluators then follow up by asking, "What do you mean?" and "Why do you care?" This dialogue is repeated until the objective is sufficiently measurable. Often, once stakeholders come to agreement about what they actually mean, the issue starts to appear much more measurable.[45]

---

[45] Hubbard, 2010.

Likewise, Nelson encourages evaluators to elicit expertise from SMEs and practitioners to identify the load-bearing nodes of the logic model and associated processes for measuring them. This can be facilitated through the use of panels, semistructured workshops, simulations, and exercises or "premortems," in which participants assume that the program has failed and work backward with SMEs, program managers, and stakeholders to identify the sources of failure.[46]

Evaluators should conduct an extensive literature review of past evaluations, case studies, and even memoirs to identify measures used explicitly and implicitly to evaluate the effectiveness of similar interventions implemented in the past. Some campaigns are widely perceived to have succeeded or failed. What informed those judgments, and can those rationales be operationalized as measures for future campaigns? To facilitate this process, DoD should consider maintaining a wiki casebook of what has worked and what has not in past IO campaigns.[47] Behavioral change theory should also be reviewed to identify the measures of influence used in academic settings that can be applied to an operating environment.

Matthew Warshaw suggested that the issue of measure development may have less to do with developing new measures than leveraging those that have already been used or written about. In his view, "lots of great work has been done," but no one has the time or resources to dedicate to a thorough literature review, because "once the money is in the door, they're already past due for the first deliverable."[48]

In reviewing the literature, Anthony Pratkanis urged DoD to look beyond the usual suspects. He suggested reviewing historical texts and memoirs from Vietnam or World War II to identify the type of indicators that soldiers looked for and trusted as measures of "winning over villages." These texts are "full of gems" that can be applied or adapted to today's influence environment. For example, during World War II, the assistant chief of the Allied Expeditionary Force Supreme Headquarters' Psychological Warfare Division, R. H. S. Crossman, would measure the influence of the information pamphlets he was dropping by visiting the office that was keeping track of all of the "rumors" to see if his own were making the list.[49] The following are among the classics that he recommends:

- Martin F. Herz, "Some Psychological Lessons from Leaflet Propaganda in World War II," *Public Opinion Quarterly*, Vol. 13, No. 3, Fall 1949, pp. 471–486.
- William E. Daugherty and Morris Janowitz, *A Psychological Warfare Casebook*, Baltimore, Md.: Johns Hopkins University Press, 1958.

---

[46] Author interview with Christopher Nelson, February 18, 2013.

[47] Author interview with Nicholas Cull, February 19, 2013.

[48] Author interview with Matthew Warshaw, February 25, 2013.

[49] Author interview with Anthony Pratkanis, March 26, 2013.

- Ronald De McLaurin, Carl F. Rosenthal, and Sarah A. Skillings, eds., *The Art and Science of Psychological Operations: Case Studies of Military Application*, Vols. 1 and 2, Washington, D.C.: American Institutes for Research, April 1976.
- Wallace Carroll, *Persuade or Perish*, New York: Houghton Mifflin, 1948.

We conclude our discussion of measure construction with a few recommendations, best practices, and pitfalls to avoid.

Planners should consider developing an indicator clearinghouse of validated and potential IIP measures and indicators. This repository could show where the measures have been used before, how well they worked, and the extent to which they have been validated by social science methodologies. Invalidated measures could be kept in the repository but visually crossed out to discourage evaluators from repeatedly using invalid measures.

Each influence objective should be tied to several specific measures, because some measures will have insufficient or unreliable data. For example, if your goal is to reduce the influence of a particular mullah, your measures could assess (1) the population's self-reported impressions of him, (2) his attendance at his mosque, and (3) how often he is mentioned in communications from various organizations or the press.[50] As addressed in the discussion of convergent validity, the most valid measures of success are those that converge across multiple quantitative and qualitative data items.[51]

On the other hand, evaluators and planners should be careful to avoid "metric bloat" or "promiscuous" measure collection. Having too many measures per objective can complicate analysis and the interpretation of results.[52] If planners find that the number of measures is becoming unmanageable, they should discard the lower-performing ones, as determined by the attributes identified in the preceding section. It is also worth noting that measuring the same outcome twice does not satisfy two layers of the assessment scheme. In one SME's experience, planners often adhere to a circular logic where they want an observed effect to cause the same effect.[53] In this situation, they might call something that is fundamentally the same by two different names. For example, "reductions in the number of attacks and incidents will lead to increased security" almost sounds sensible but is less so if rephrased as "increases in security will lead to increased security."

While many outcomes and constructs are difficult to quantify, planners and assessors should make an effort to express IIP measures in quantifiable terms. The

---

[50] Author interview with Anthony Pratkanis, March 26, 2013.

[51] Author interview with Steve Booth-Butterfield, January 7, 2013.

[52] William P. Upshur, Jonathan W. Roginski, and David J. Kilcullen, "Recognizing Systems in Afghanistan: Lessons Learned and New Approaches to Operational Assessments," *Prism*, Vol. 3, No. 3, 2012, p. 91; Downes-Martin, 2011, p. 108.

[53] Author interview on a not-for-attribution basis, February 20, 2013.

NATO Framework for Public Diplomacy evaluation suggests that KPIs for influence be represented by a numeric value such as the Klout Score, an indicator of social media influence with widely accepted validity that takes into account Facebook, Twitter, LinkedIn, YouTube, and other platforms.[54] *Where the measure is numeric, assessors should express the measure in terms of a ratio so that progress from the baseline to future states can be easily determined.* In this formulation, the baseline value is the denominator, and changes due to the IIP activity are reflected in the numerator.[55]

Planners should avoid the temptation to only collect data on indicators of success. Measures or indicators should be defined or scaled such that they capture failure or regression as well as success.[56] The measurement system should also be flexible enough to capture unintended consequences.[57]

The incentive and accountability system tied to the measures should be carefully considered. Measures that create perverse incentives, such as "teaching to the test" or "buying likes," should be avoided. Metrics can take on a "life of their own."[58] If indicators are not defined carefully enough, it may be possible for an activity to satisfy the indicator without affecting the construct that stakeholders are interested in measuring, invalidating the measure, and distorting program activities.[59] Measures of exposure are particularly susceptible to perverse incentives.[60] A recent State Department Inspector General's report accused the Bureau of International Information Programs of "buying likes" on Facebook as a way to improve the perceived reach of the program.[61]

Dennis Affholter cites an example of a state using the number of new foster homes licensed as an indicator of successful placement of children. The social services system responded by aggressively recruiting and licensing new homes without building the capabilities of foster parents to work with the children. As a result, the indicator moved upward, but the placement of children in appropriate homes did not improve.[62]

Through a principal-agent analysis, Leo Blanken and Jason Lepore modeled the impact of measurement on military organizations to show that the undervaluing of the

[54] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 18.

[55] The Initiatives Group, 2013.

[56] Author interview with Steve Booth-Butterfield, January 7, 2013.

[57] Author interview with James Pamment, May 24, 2013.

[58] Military Operations Research Society, 2012.

[59] UK Ministry of Defence, 2012, p. 364.

[60] Author interview with Craig Hayden, June 21, 2013.

[61] Office of the Inspector General, U.S. Department of State, *Inspection of the Bureau of International Information Programs*, May 2013; Craig Hayden, "Another Perspective on IIP Social Media Strategy," *Intermap*, July 23, 2013.

[62] Dennis Affholter, "Outcome Monitoring," in Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds., *Handbook of Practical Program Evaluation*, San Francisco: Jossey-Bass, 1994, quoted in Rossi, Lipsey, and Freeman, 2004, p. 227.

incentive properties of measures creates systemic positive bias of information and deleterious incentive structures for agents within the organization. They show that effective measurement is possible only if the incentives of the agents (implementers) align with the goals of the principals (stakeholders), the principal knows the agent's motivations, and the agent understands that his or her actions contribute to the metric. The authors use the example of Vietnam to demonstrate the consequences of misaligned incentives. The "body count" metric led to an overproduction of violence that worked against the principal's goal of a stable South Vietnam.[63]

Measures that are obviously used to promote programs or redistribute resources are at a high risk of being manipulated by the program being assessed. Past examples of manipulated or "captured" metrics in counterinsurgency environments have included exaggerated reports of operational readiness of host-nation forces or of enemy casualties and reduced reporting of civilian casualties.[64]

Excursion: Measuring Things That Seem Hard to Measure

It is difficult to measure the impact of a program when the outcomes are long-term and there are many intervening variables that might explain observed outcomes. But these challenges, as Professor Craig Hayden notes, should "not serve as a cover for not doing the measurement that needs to be done. We just need better measures."[65] In *How to Measure Anything*, Hubbard provides several novel suggestions for measuring things that seems "impossible" to measure:

- If what you are trying to observe hasn't left a trail, add a "tracer" so that it starts to leave a trail. For example, an activity aimed at increasing recruitment could encourage individuals to sign up through a unique, traceable channel linked to the IIP activity. Alternatively, units returning from patrol can be systematically debriefed with a checklist to assess their observation of atmospheric indicators.
- If you can't follow a trail at all, conduct an experiment to create the conditions to observe it.[66]
- Work through the hypothetical consequences of success and failure. If the activity worked, what should you expect to see? What about if it failed? Think of extreme cases, and work backward to more reasonable ones.[67]

---

[63] Leo J. Blanken and Jason J. Lepore, *Performance Measurement in Military Operations: Information Versus Incentives*, Monterey and San Luis Obispo, Calif.: Naval Postgraduate School and California Polytechnic State University, November 12, 2012, p. 1.

[64] Dave LaRivee, *Best Practices Guide for Conducting Assessments in Counterinsurgencies*, Washington, D.C.: U.S. Air Force Academy, December 2011, p. 18.

[65] Author interview with Craig Hayden, June 21, 2013.

[66] Hubbard, 2010, p. 130.

[67] Hubbard, 2010, p. 130.

- Decompose the construct you are trying to measure so that it can be estimated from other measurements. Place each element of the decomposition into one or more methods of observation: trails left behind, direct observation, tracking with "tags" or tracers, or experiments.[68]
- If it is important, it can be defined in terms of the *cost* of being wrong and the *chance* of being wrong. And if the outcome is possible, it can be observed.
- Try to compute the value of observing something by identifying the threshold where it begins to reduce your uncertainty about outcomes.
- Review the social science methods. Even a basic knowledge of random sampling, experimental design, or expert elicitation techniques can significantly reduce uncertainty.[69]
- Just do it. Start collecting observations, before the measurement system is fully validated. You may be surprised by what you find with the first few observations.[70]

MOE and MOP Elements in Defense Doctrine

While more IIP assessment doctrine is needed, DoD IIP assessors can nonetheless draw from existing DoD and intelligence community guidance to inform MOP and MOE development. Two useful sources are the Army's field manual for inform and influence activities (FM 3-13) and the *Information Environment Assessment Handbook* issued by the Office of the Under Secretary of Defense for Intelligence. This section reviews some of the guidelines for MOE and MOP development from those documents.

FM 3-13 provides guidance for the construction of MOEs and MOPs in the information environment. According to the manual, IIP MOEs should specify the *activity* (desired direction of change), *descriptor* (target audience), *subject* (what the campaign is trying to influence the target to do), and the *metric* (numerical percentage threshold of effectiveness derived from higher-level guidance). Implicit in the specification of the metric is the *baseline*, or the historical measure from which the threshold level of effectiveness is derived. IIP MOP components identified in the manual include *quantity* (number of broadcasts or deliveries), *medium* (product format used to disseminate message), *delivery* (how and where U.S. forces delivered the medium), and *target* (selected audience).[71]

Figure 6.2 shows how MOPs feed into MOEs in support of the objective to increase voter turnout in an upcoming election, which would show support for a democratically elected government. In this example, the MOE is to increase (activity) votes (subject) among registered voters at the polls (descriptor/target audience) by 33 percent (metric) compared with the UN-monitored election two years ago (baseline). The

---

[68] Hubbard, 2010, p. 130.

[69] Hubbard, 2010, p. 287.

[70] Hubbard, 2010, p. 137.

[71] Headquarters, U.S. Department of the Army, 2013a, p. 7-3.

MOP is to distribute 275 (quantity) handbills (medium) door-to-door (delivery) to the registered voters (target audience).[72] The figure also shows the planning order, which ensures that the process is tailored to the effort's objectives.

The *Information Environment Assessment Handbook* issued by the Office of the Under Secretary of Defense for Intelligence identifies the four elements to a repeatable, measurable, and operationally relevant MOE: ratio, IE condition, measurement characteristics, and IE object. The *ratio* identifies a baseline and represents progression toward the objective. The *IE condition* is the specific condition within the IO that the

Figure 6.2
Illustrating the Assessment Methodology for Army Inform and Influence Activities
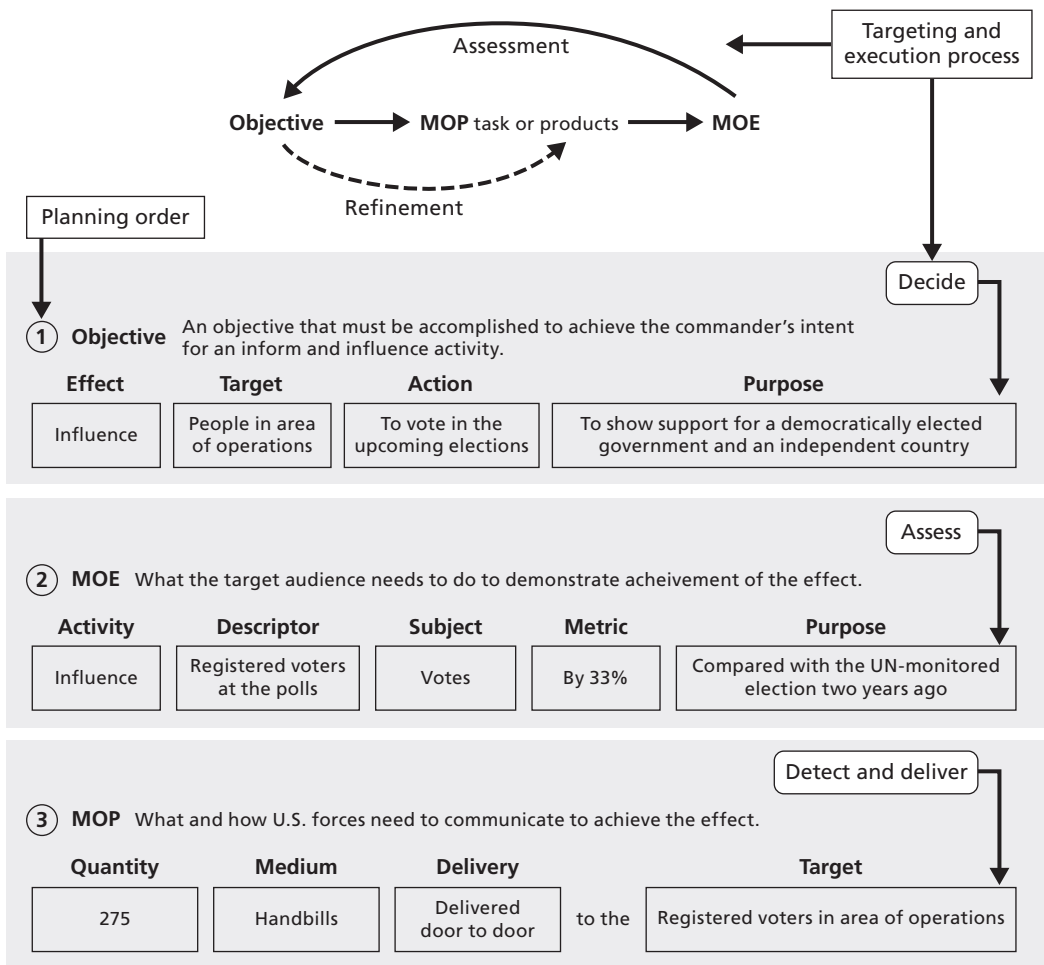
SOURCE: Adapted from Headquarters, U.S. Department of the Army, 2013a, p. 7-4.
RAND *RR809/1-6.2*

---

[72] Headquarters, U.S. Department of the Army, 2013a, p. 7-3.

MOE assesses change in and which can change through one of three measurement characteristics: *amount*, *accessibility*, or *functionality*. The *IO object* is the entity within an IE system whose condition is changing (e.g., facility, individual, government).[73]

The handbook goes on to describe the five steps for developing MOEs with those four elements: effect decomposition, characteristic identification, baseline ratio definition, determining data collection requirements, and aggregation and assessment. Assessors begin by decomposing a well-defined effect to identify the conditions that would need to change to realize the effect and the IE object associated with the effect whose condition is changing (e.g., a facility, an individual, an organization, a government). Next, the characteristics of the desired change are identified along the three measurement dimensions (amount, accessibility, and functionality). The assessors should then make a "concerted effort" to express the MOE in quantifiable terms such that a ratio can be constructed to easily compare the baseline (denominator) with a future condition (numerator).[74]

Once the ratio is defined, the assessor identifies the observable elements of the condition of change that need to be monitored and specifies a data collection plan with sampling rates (e.g., weekly, monthly). Multiple variations of collections are specified so that the evaluation does not solely rely on a limited collection method. The collection plan and sampling rates should clarify the operational context, the type of indicators or observables (cultural, situational, technical, functional, and/or biometric), possible or expected outcomes, and standards for sampling. Finally, ratios across all MOEs are weighted and aggregated.[75]

Summary

This chapter addressed key concepts and best practices for developing the measures that can and should be used to evaluate the performance and effectiveness of IIP campaigns. It reviewed two general processes: deciding which constructs are essential to measure and operationally defining the measures. Key takeaways include the following:

- Good measures are valid, reliable, feasible, and useful.
- The quality of measures depends on the quality of the influence objectives and associated intermediate objectives enumerated within the program logic model.
- The importance of measuring something, or the information value of a measure, is determined by the amount of uncertainty about its value and the costs of being wrong. Assessors should therefore give priority to load-bearing or vulnerable pro-

---

[73] The Initiatives Group, 2013, p. 16.

[74] The Initiatives Group, 2013, p. 16.

[75] The Initiatives Group, 2013, p. 18.

cesses. These elements can be identified by drawing up IIP theories, empirical research, expert elicitation, and evaluations of similar campaigns implemented in the past.[76]

- There is tension between the feasibility of a measure and its utility. Often, what is important or useful to measure cannot be easily observed. Assessors should first identify the measures with the highest information value and subsequently determine what is feasible among those worth measuring.

- IIP evaluations should include a measure of exposure and several measures that capture the internal processes by which exposure to the campaign produces behavioral change. These processes will depend on the theory of change and can include changes in knowledge, attitudes, self-efficacy, norms, issue saliency, and behavioral intention.

- To aid in measure selection and operational definition, assessors should consider assembling real or virtual panels, semistructured workshops, simulations or exercises, premortems and clarification workshops with program managers, evaluation experts, cultural anthropologists, trusted local sources, and other stakeholders.

- DoD IIP evaluators should develop a repository or clearinghouse of validated IIP measures.

---

[76] Author interview with Christopher Nelson, February 18, 2013.

Assessment Design and Stages of Evaluation

The *design* of an assessment or evaluation is the plan that describes the research activities that will answer the questions motivating the evaluation. *Design determines the way in which the evaluation can (or cannot) make causal inference regarding the outputs, outcomes, or impacts of the intervention*. Design-related decisions govern the *structure* of data collection—that is, the number, timing, and type of data measurements—rather than the *methods* by which data are collected (the topic of the next two chapters). Broadly, evaluation designs can be classified as experimental (control with random selection), quasi-experimental (control without random selection), and nonexperimental or observational studies (no control). Evaluators should be familiar with a range of potential evaluation designs and their strengths and weaknesses so that they can design the best and most appropriate evaluation given stakeholder needs, populations affected, and available resources.[1] Deciding whether or not to pursue an assessment design capable of attributing causation is the first and likely largest assessment design decision. In the DoD planning context, that assessment design decision should follow from operational design undertaken during mission analysis (step 2 in JOPP). The iterative process during operational design that defines the problem, develops the solution, and allows for a clear statement of the commander's intent should also contribute to clarity about whether it is important to show the extent to which specific efforts *caused* the solution to be realized, or whether it is sufficient just to show the problem solved. Specific assessment design alternatives capable of meeting the assessment requirement should be prepared as part of COA development and evaluated during COA analysis, war-gaming, and comparison.

This chapter presents key concepts in IIP evaluation design. It begins with a discussion of the criteria and associated determinants of the quality of assessment design, including feasibility, usability, and various types of validity. The discussion of usability is accompanied by the introduction of the "uses and users" construct for matching the evaluation approach to the end users and stakeholders. The chapter then presents the three phases or types of evaluation: formative, process, and summative. Three sections

---

[1]   Valente, 2002, pp. 87–88.

describe design options associated with these three evaluation phases. The section on summative evaluation designs is the most extensive and addresses considerations in the use of experimental, quasi-experimental, and nonexperimental designs for evaluating the effectiveness of IIP campaigns.

Criteria for High-Quality Evaluation Design: Feasibility, Validity, and Utility

How should evaluators choose between possible evaluation designs? This section proposes that the best designs are valid, feasible, and useful (as characterized in the text to follow). However, there are tensions and trade-offs inherent in pursuing each of those objectives. Thomas Valente summarizes this dynamic by characterizing the "two conflicting forces in design" as (1) sufficient rigor and specificity to make firm conclusions and (2) limitations of money, time, cooperation, and protection of human subjects.[2] Evaluators should therefore select the strongest evaluation design, in terms of internal and external validity, among those designs that are useful and feasible with allocated resources.[3]

Moreover, the importance of selecting the most rigorous design varies with the importance and intended use of the results. Resources should therefore be allocated according to the importance of potential outcomes. Presuming that *resources* and *rigor* are closely correlated, Peter Rossi and colleagues advocate the "good enough" rule in evaluation designs: "The evaluator should choose the strongest possible design from a methodological standing after having taken into account the potential importance of the results, the practicality and feasibility of each design, and the probability that the design chosen will produce useful and credible results."[4] In a budget-constrained environment, evaluations are *simultaneously more important and harder to afford*. To allow room for more assessments within a constrained budget, there needs to be a mechanism for quick, cheap, and good-enough assessments (see the discussion of resource prioritization in Chapter Three). Remember that assessment ultimately supports decisionmaking. What level of methodological rigor is sufficient to support the decisions that need to be made?

The following sections address feasibility, validity, and utility as attributes of evaluation design. A complementary discussion of evaluation criteria can be found in the section on meta-evaluation in Chapter Eleven.

---

[2] Valente, 2002.

[3] Valente, 2002, pp. 89–90.

[4] Rossi, Lipsey, and Freeman, 2004, p. 238.

Designing Feasible Assessments

Acknowledging the importance of constructing the best and most valid evaluation possible given the available resources, Thomas Valente states that the first requirement of evaluation design "is that it be practical, which often prevents the use of the best design that might be theoretically possible."[5] Time, resources, and ethical or practical concerns with carrying out randomized experiments all constrain feasibility. The *Commander's Handbook for Assessment Planning and Execution* notes that in "fast-paced offensive or defensive operations or in an austere theater of operations a formal assessment may prove impractical."[6]

For example, planners may need to know whether a particular influence product is resonating or backfiring to inform a near-term decision about whether and how to scale distribution of that product or future products in that series. Given resource and time constraints, they are unable to conduct a pre-post or exposed-versus-unexposed quasi-experimental design using data from representative surveys (these designs are described in more detail in the section on summative evaluation designs). In this situation, the best approach may be to conduct an informal focus group, or in-depth interviews with a convenience sample consisting of trusted local sources and experts, preferably including a mix of subjects with various levels of exposure to the campaign or message, or to rely on information from intelligence sources (signals intelligence, HUMINT, or other relevant take).

More generally, commanders often need post facto assessments when baseline data were not collected (because someone forgot to plan to meet that need during initial planning), which necessarily limits the rigor of the assessment. In these cases, evaluators should consider post-only quasi-experimental designs that compare outcomes between those who were exposed and those who were not exposed, perhaps within propensity matched groups, as described in following sections. Importantly, challenges inherent in difficult situations should not serve as a cover for failing to consider the available options and selecting the best feasible design from a methodological perspective.

To gauge the feasibility of a new resource-intensive evaluation design, IIP evaluators should consider the use of pilot evaluations. Pilot evaluations test the evaluation design on a much smaller scale than ultimately envisioned by either studying the effectiveness of a small effort or by focusing on a subset of the target audience. BBC Media Action sponsored a pilot field experiment conducted by the Annenberg School for Communication at the University of Pennsylvania that looked at the effects of partisan radio programming on public transportation passengers in Ghana. The pilot gave the research team at BBC Media Action considerable insight into the utility and feasibil-

---

[5] Valente, 2002, p. 88.

[6] U.S. Joint Chiefs of Staff, 2011c.

ity of adding field experiments to its research portfolio.[7] Ronald Rice, the Arthur N. Rupe Chair in Social Effects of Mass Communication at the University of California, Santa Barbara, and a leading expert in public communication evaluation, echoed this suggestion, noting that pilot studies using cultural anthropologists help make sense of the social context and the conditions under which the evaluation can be conducted effectively.[8]

Time permitting, DoD IIP efforts should include both pilot tests of the effort's activities and pilot tests of the evaluation design. Such limited-scope formative efforts can ensure that money spent later on the full-scale efforts is well spent.

Designing Valid Assessments: The Challenge of Causal Inference in IIP Evaluations

Designing feasible evaluations is in tension with designing valid ones. Validity represents the extent to which a design or a measure is accurate or free from systemic bias. *Internal validity* is the extent to which the design supports the kinds of causal inferences or causal conclusions that need to be made within the evaluation. *External validity* (also known as *generalizability* or *ecological validity*) is the extent to which design is able to support inference (e.g., generalization) about the larger population of interest. Components and trade-offs associated with both forms of validity are discussed in the following sections. Of note, this section is concerned with study validity, the degree to which the evaluation design accurately measures program impact. Measurement validity, the degree to which a variable represents the concept it purports to measure, is addressed in Chapter Six.

Internal Validity

It is commonplace to assert the difficulty or impossibility of determining causality when it comes to isolating the contribution of an IIP intervention. Adding to the challenge of reliably ascertaining outcome measures is "the question of whether observed changes in attitudes and behavior can be directly attributed to any specific influence activity."[9]

Several SMEs commented on the scale of this challenge. Victoria Romero hadn't "seen anyone deal with [the challenge of inferring causality] well."[10] In the DoD context, the contribution of the IIP effort often cannot be separated from "background

---

[7]   Author interview with James Deane, May 15, 2013. The results of the Ghana experiment have been documented in a working draft: Jeffrey Conroy-Krutz and Devra Coren Moehler, "Moderation from Bias: A Field Experiment on Partisan Media in a New Democracy," draft manuscript, May 20, 2014.

[8]   Author interview with Ronald Rice, May 9, 2013.

[9]   Rate and Murphy, 2011, p. 10.

[10]   Author interview with Victoria Romero, June 24, 2013.

noise" and the myriad factors operating at operational, tactical, and strategic levels.[11] In public relations, causal inference is considered a "$64,000 question."[12] Charlotte Cole, the senior vice president of global education who oversees research on the effects of international programming at Sesame Workshop, stressed that the most important takeaway of all of this is that measuring the impact of media interventions is "enormously complicated—much more than people think."[13]

Adding to the complexity of identifying the contribution of U.S. actions versus environmental factors is the challenge associated with isolating the contribution of influence tactics within the broader context of a military campaign. This issue is of primary importance due to the role of assessments in driving resource allocation and choosing between competing capabilities. In Mark Helmke's view, we cannot "evaluate communication strategy in a vacuum, because it is one weapon in a broader strategy that cannot be separated from noncommunicative aspects."[14] Nicholas Cull illustrated this challenge with the example of identifying the contribution of the communication campaign to the military intervention in Haiti: "What are you going to do? Run a controlled experiment where you invade a Caribbean island without explaining it to everyone?"[15]

Threats to internal validity are the factors that limit the ability to draw causal inference. The most-valid evaluations are those that included the most-effective controls against those factors. The threats to internal validity that are most relevant to IIP evaluation research include confounding variables, selection, maturation, history, instrumentation, attrition, and regression toward the mean. These are explained in Table 7.1.

Threats to internal validity are controlled by design choices. The higher the internal validity, the more controlled, complex, and therefore (typically) resource intensive the design will be. Table 7.2 lists six study designs identified by Valente and Patchareeya Kwan in increasing order of control against threats to internal validity. In the pre- and postprogram (4 and 5) and the Solomon four-group (6) designs, both the treatment and control groups are tested before and after the intervention. This controls for history, maturation, and testing threats to internal validity, but it does not control for sensitization, in which the premeasure may sensitize the subject. The Solomon four-

---

[11] David C. Becker and Robert Grossman-Vermaas, "Metrics for the Haiti Stabilization Initiative," *Prism*, Vol. 2, No. 2, March 2011.

[12] Author interview with David Michaelson, April 1, 2013.

[13] Author interview with Charlotte Cole, May 29, 2013.

[14] Author interview with Mark Helmke, May 6, 2013.

[15] Author interview with Nicholas Cull, February 19, 2013.

Table 7.1
Threats to Internal Validity and Challenges to Establishing Causality in IIP Evaluation

Confounding variables: environmental	An extraneous variable or factor influences the outcome of interest (dependent variable). This is the primary threat to internal validity associated with post-only or quasi-experimental research designs in media effects research. Other competing factors in the environment can influence the same behaviors you are seeking to change.[a] Even if they are identified, it is not possible to control for every confounder in media evaluation. For drunk driving awareness campaigns, Tony Foleno noted that it is simply impossible to "control for every enforcement effort."[b] In the international security context, observed outcomes in a particular country or region "cannot be separated from background noise" and everything else "happening on a national and international level."[c] It is "very difficult to know the effect of your message when there are competing messages in the same marketplace."[d]
Confounding variables: other U.S. actions	In addition to environmental factors, other coalition and U.S. government kinetic and nonkinetic activities confound the outcome of interest. Because communication is one weapon in a broader strategy, it is difficult to isolate the effects from the noncommunicative aspects of the campaign.[e] For example, if you conduct an IIP intervention and two weeks later the United States gives $1 billion to a country, "people might have a favorable opinion of the U.S. but it's probably not due to the campaign."[f]
Unobserved confounding variables	The influences bearing on behavioral changes are not always observable, because people cannot assess what changed their behavior. In these cases, "pinpointing the actual cause of a behavior change is next to impossible."[g]
Election bias	There are systematic differences between the subjects in the treatment group and those in the control group. This is a large issue in exposed versus unexposed quasi-experimental designs measuring media impact: The individuals who voluntarily exposed themselves to the product (the treatment group) may be predisposed to the message.
Maturation	Maturation is the naturally occurring process of change that affects both the control group and the treatment group and that interacts with the intervention. This is particularly relevant to IIP interventions because the outcomes of interest are typically only observed over the long term.[h]
History	Uncontrollable events coincide with the treatment and have an effect on outcomes that cannot be distinguished from the intervention.
Contamination or diffusion of the treatment	The control group may be contaminated by individuals in the treatment group (those exposed to the intervention) sharing or talking about the media with members of the control group. According to Amelia Arsenal, in IIP evaluations it is important to check for spillover effects, such as whether those exposed to the intervention contaminated the control group (e.g., by discussing or sharing the media) or whether the comparison group was otherwise unintentionally exposed to the intervention.[i]
Testing	Taking a pretest may increase planners' knowledge of the subjects.
Sensitization	The pretest sensitizes the subjects to the topic of the intervention.
Instrumentation	Changes in measurement tools or procedures may result in differences between the pretest and posttest.
Hawthorne effect	Subjects may react positively to being part of the treatment group; this is also called the observer effect.

[a] Interview with Doug Yeung, March 14, 2013. [b] Interview with Tony Foleno, March 1, 2013. [c] Becker and Grossman-Vermaas, 2011. [d] Interview with Marc Patry, June 6, 2013. [e] Interview with Mark Helmke, May 6, 2013. [f] Interview with Julianne Paunescu, June 20, 2013. [g] Interview with Marc Patry, June 6, 2013. [h] Interview with Craig Hayden, June 21, 2013. [i] Interview with Amelia Arsenault, February 14, 2013.

group design controls for the sensitization effect by adding an un-pretested control group and an un-pretested treatment group.[16]

Table 7.2
Study Designs and Internal Validity

Design	Baseline Observation	Intervention	Follow-Up Observation	Controls for
1A. Postprogram only	—	X	X	None
1B. Postprogram only, with				
Exposed (treatment)	—	X	X	History, some confounds
Unexposed (control)	—	—	X	
1C. Postprogram only within propensity matched groups, with				
Exposed (treatment)	—	X	X	Selection, history, some confounds
Unexposed (control)	—	—	X	
2. Pre- and postprogram	X	X	X	Selection
3. Pre- and postprogram with				
Treatment group	X	X	X	Testing
Post-only treatment group	—	—	X	
4. Pre- and postprogram with				
Treatment group	X	X	X	History and maturation
Control group	X	—	X	
5. Pre- and postprogram with				
Treatment group	X	X	X	Sensitization
Control group	X	—	X	
Post-only treatment group	—	X	X	
6. Solomon four group with				
Treatment group	X	X	X	All of the above
Control group	X	—	X	
Post-only treatment group	—	X	X	
Post-only control group	—	—	X	

SOURCE: Adapted from Valente and Kwan, 2012, pp. 89–90.

---

[16] Thomas W. Valente and Patchareeya P. Kwan, "Evaluating Communication Campaigns," in Ronald E. Rice and Charles K. Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2012.

External Validity

External validity, or the extent to which evaluation results can be generalized beyond the sample, is limited by geography, time, target-audience consideration, and experimental or laboratory conditions. There is often a trade-off between external and internal validity. Designs with the highest internal validity often have weak ecological validity because the laboratory-like conditions required to control for the threats to internal validity do not appropriately reflect conditions in which the focal audience would interact with the program "in the wild" or under generalizable circumstances.[17] Likewise, field experiments taking place in the wild have the highest ecological validity but are the hardest to control for threats to internal validity.

Craig Hayden, an assistant professor in the International Communication Program at American University, noted that this is particularly true with international strategic communication activities. Studies that are capable of demonstrating causality in a rigorous way "have more-narrow parameters that do not correspond to the messy boundaries" of international strategic communication campaigns.[18] Cole argues that it is a mistake to prioritize randomized controlled trials at the expense of studies that observe how the audience engages the media naturalistically: "If you don't know how people are actually using the medium naturalistically, you haven't shown anything about your impact."[19]

Designing Useful Assessments and Determining the "Uses and Users" Context

As emphasized in the introductory chapters of this report (specifically, Chapters One through Three), assessment is a decision-support tool. Evaluations must therefore be designed so that end users are able to inform decisionmaking with the results, and the nature of the assessments has significant implications for design. For example, if end users need to know whether a specific activity is influencing a particular target audience, the design should assign priority to those conditions that allow valid causal inference regarding the extent to which that population was affected by the activity. If the design will be used to answer broader questions regarding how a program shapes the views of a larger audience, it may place less emphasis on experimental controls and more on the generalizability of the population under study. If it will be primarily used to inform midcourse process improvements, observational and nonexperimental approaches may suffice. CDR (ret.) Steve Tatham, the UK's longest continuously serving information activities officer, argued that one of the key lessons from the past three decades of defense evaluation practice is that unrealistic or poorly managed stakeholder expectations about the nature, benefits, and risks of evaluation lead to unde-

---

[17] Author interview with Marie-Louise Mares, May 17, 2013.

[18] Author interview with Craig Hayden, June 21, 2013.

[19] Author interview with Charlotte Cole, May 29, 2013.

Box 7.1
The Challenge of Determining Causality in IIP Evaluation

The preceding sections presented several daunting challenges to establishing causality in IIP evaluations. But despite these difficulties, it is not impossible to obtain reasonable estimates of causal effects. Craig Hayden expressed concern that the preoccupation with the challenge of causality has become "cover for not doing the measurement that needs to be done." In his view, these challenges are just reasons why evaluators need more-thoughtful designs and better, more-creative measures that capture long-term effects (see the section "Techniques and Tips for Measuring Effects That Are Long-Term or Inherently Difficult to Observe" in Chapter Nine).[a] Moreover, it is better to frame IIP interventions as *contributors to* rather than *causes of* change, because programs are "long-term and there are many intervening variables that might provide an explanation for an outcome."[b]

A DoD MISO practitioner commented that much of the concern over causality is driven by a lack of awareness of alternatives to true experimental design.[c] In *Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know*, Mark Jeffery responds to the objection that there are too many factors to isolate cause and effect: "The idea is conceptually simple: conduct a small experiment, isolating as many variables as possible, to see what works and what does not."[d]

Ultimately, there are a number of designs described in this chapter that can lead to assessments of DoD IIP activities with high internal validity and allow strong causal claims. These designs tend to be more resource intensive and require an unambiguous commitment to some kind of experimental or quasi-experimental structure in program delivery and assessment. This, then, turns back to the matter of feasibility. If you want to be able to make causal claims, are you willing to put forward the time and effort necessary to make that possible?

While experimental or quasi-experimental designs are often comparatively resource intensive, many quasi-experimental designs are more feasible in the defense context than many planners might think. A functional quasi-experimental design may simply require a delay in delivery of all or part of a program's materials and outcome measurement at a few additional times. Quasi-experiments are not as rigorous as randomized controlled experiments, but they still provide strong grounds from which to assert causation—sufficient for many assessment processes. See the section "Summative Evaluation Design," later in this chapter, for more discussion of quasi-experimental designs in IIP evaluation.

[a] Author interview with Craig Hayden, June 21, 2013.

[b] Author interview with Craig Hayden, June 21, 2013.

[c] Author interview on a not-for-attribution basis, July 30, 2013.

[d] Jeffery, 2010.

---

sirable conflicts and disputes, lack of evaluation utilization, and dissatisfaction with evaluation teams and the evaluations they produce.[20]

Assessment design, processes, and degree of academic rigor and formality should be tailored to the assessment end users and stakeholders. Monroe Price, director of the Center for Global Communication Studies at the University of Pennsylvania's Annenberg School for Communication, noted that the core questions governing evaluation design are *whom* and *what decisions* the evaluation is informing. Field commanders will have a different set of questions from congressional leaders.[21] Gerry Power, former chief operating officer of InterMedia and former director of research at the BBC World Service Trust (now BBC Media Action), echoed many of the other SMEs we interviewed

---

[20] Author interview with UK Royal Navy CDR (ret.) Steve Tatham, March 29, 2013.

[21] Author interview with Monroe Price, July 19, 2013.

in arguing that if assessment is to inform strategic and tactical decisions, academic rigor must be balanced with stakeholder needs, appetite for research, and cost considerations. InterMedia, for example, tends to use more-sophisticated analytic techniques, such as structural equation modeling, when it is working with a client that can appreciate, understand, and interrogate InterMedia on those techniques.[22] Tarek Azzam at Claremont Graduate University noted that prioritizing validity and rigor over stakeholder needs sacrifices the likelihood that the assessment will be used down the road and the certainty that users are making decisions on good data.[23] Part of successful assessment design is balancing stakeholder needs with feasibility and rigor.

To design useful evaluations, evaluators must first understand the assessment audience (users and stakeholders) and the decisions that evaluations will inform (assessment uses). Christopher Nelson, a senior researcher at the RAND Corporation, calls this the "uses and users" context. He encourages evaluators to identify and characterize the key *users* (end users and stakeholders) and *uses* of the assessment (that is, what decisions it will inform) prior to designing the evaluation and measurement system.[24] *End users* are those users with formal or institutional responsibility and authority over the program and who have an active interest in the evaluation. In the IO context, program managers, military leadership, and Congress represent potential end users, depending on the level of evaluation. *Stakeholders* include a broader set of "right to know" audiences that has a more passive interest in the evaluation. Stakeholders could include the target audience, media, and internal program management and staff.[25]

As noted in Chapter Two, the three motives for assessment (improve planning, improve effectiveness and efficiency, and enforce accountability) can be categorized even more narrowly by noting that assessments are primarily either up- and out-focused (accountability to an external stakeholder) or down- and in-focused (supporting planning or improvement internally). This categorization focuses on the users of the assessments. The characteristics of both categories are described in Table 7.3.

A matrix that maps each assessment user to an assessment use and, where appropriate, identifies when and for how long (continuous versus a single point in time) the assessment results will be needed can help planners identify the uses-and-users context.[26] Table 7.4 provides a basic template for a use-user matrix. Table 7.5 provides an example of a use-user matrix for an evaluation of a hypothetical IIP program.

---

[22] Author interview with Gerry Power, April 10, 2013.

[23] Author interview with Tarek Azzam, July 16, 2013.

[24] Author interview with Christopher Nelson, February 18, 2013.

[25] Author interview with Christopher Nelson, February 18, 2013.

[26] Author interview with Christopher Nelson, February 18, 2013.

Table 7.3
Accountability- Versus Improvement-Oriented Evaluations

Characteristic	Accountability-Oriented Measures	Improvement-Oriented Measures
Audience	External (e.g., funders, elected officials)	Internal (e.g., program managers, staff)
Decisions supported	Judging merit and worth (e.g., reauthorization, termination)	Design, identification of gaps, corrective action plans
Data requirements	Comparable over time and across programs	Targeted to program-specific (or campaign-specific) concerns
Evaluation type or stage	Summative evaluation	Formative evaluation or process evaluation

SOURCE: Adapted from interview with Christopher Nelson, February 18, 2013.

Table 7.4
Uses and Users Matrix Template

		Likely Uses		
		Accountability	Improvement	Combined/Other
Likely Users	End users			
	Stakeholders			
	Others			

Table 7.5
Uses-Users Matrix Example for Evaluating a Notional DoD IIP Program

		Likely Uses		
		Accountability	Process Improvement	Develop Evidence Base
Likely Users	Congressional and DoD resource allocators	Should this program be funded? Is influence a priority funding area in this region?	Are the right campaign objectives in place?	Are we defeating the insurgency?
	Commander/ program director	Is the program staff implementing the program as directed?	What processes are underperforming? Is the logic model properly specified?	What influence efforts work across time?
	Public, researchers	Are elected officials making wise investments?		Academic IIP research

■ Immediate need
■ Medium-term need
■ Long-term need

SOURCE: Interview with Christopher Nelson, February 18, 2013.

A Note on Academic Evaluation Studies Versus Practitioner-Oriented Evaluations and Assessments

Academic evaluation research typically involves time- and resource-intensive experimental designs intended to advance theory within the field of study. Evaluation and assessments in the field, commonly referred to as *monitoring and evaluation* (M&E), are typically less rigorous from a methodological perspective but have faster turnaround, because they are intended to shape short-term decisions on the ground. As Valente noted, M&E practitioners "don't have time to pontificate," because they need to feed results back into the program quickly.[27] Amelia Arsenault at Georgia State University explained that the difference in rigor arises from publication standards: Academics want their results to be published in peer-reviewed journals, so they need maximally reliable measurements and must couch their research in a literature review and a theoretical framework. M&E practitioners want to use reliable measurements but, because they do not need to be published, will often cut corners when doing so accrues significant practical benefits (e.g., constructing a stratified sample from convenience or snowball sampling instead of a real random sample).[28]

Academic and field evaluations are complementary, and the two groups, academics and practitioners, have things to learn from each other. Academics advance the science of evaluation while practitioners put it to use. Valente makes the case for handing academics the more intensive theoretical work that advances the state of evaluation sciences (e.g., developing and refining the scales, doing the factor analysis), which can then be by practitioners.[29] Several SMEs embraced the view that more needs to be done to encourage collaboration between academics and practitioners. Valente suggested holding a conference to bring the two communities together. Maureen Taylor, chair of strategic communication at the University of Oklahoma and author of several studies related to evaluating media interventions in conflict environments, noted that one challenge to cross-collaboration between academics and practitioners is that the large organizations contracting M&E do not want the researchers to publish because they fear public critique of the work.[30]

Types or Stages of Evaluation Elaborated: Formative, Process, and Summative Evaluation Designs

Chapter Two introduced the formative-process-summative construct for distinguishing between the three stages of evaluation. This section elaborates on this distinction

---

[27] Author interview with Thomas Valente, June 18, 2013.

[28] Author interview with Maureen Taylor, April 4, 2013.

[29] Author interview with Thomas Valente, June 18, 2013.

[30] Author interview with Maureen Taylor, April 2013.

by describing the design-related characteristics of each evaluation type. Understanding the type of evaluation that is needed is a prerequisite for designing that evaluation. Making use of the wrong evaluation framework (e.g., jumping into effectiveness issues when formative research is more appropriate) can create misleading evaluations and result in a mismatch of research resources and aims.[31]

Broadly, there are three stages in evaluation: *formative evaluation*, *process evaluation*, and *summative evaluation*. All three are applicable in the IIP context. *Formative evaluation* consists of the preintervention planning stage designed to develop and test messages, determine baseline values, analyze audience and network characteristics, and specify the logic model and characteristics of the communication system that the intervention is designed to influence, including barriers to behavioral change. *Process evaluation* determines whether the program has been or is being implemented as designed, assesses output measures such as reach and exposure, and provides feedback to program implementers to inform course adjustments. *Summative evaluation*, including outcome and impact evaluation, is the postintervention analysis to determine whether the program achieved its desired outcomes or impacts.[32] Design considerations, being tied intimately to issues of causal inference, are most relevant to the summative phase.

Julia Coffman distinguishes among four types of evaluation in two broad categories: front-end preintervention evaluations (formative) and back-end postintervention evaluations (process, outcome, and impact evaluations). *Outcome evaluation* assesses outcomes in the target population or communities that come about as a result of the IIP strategies and activities, whereas *impact evaluation* measures community-level change or longer-term results that are achieved as a result of the campaign's aggregate effects on individuals' behavior.[33] The three stages of evaluations are inherently linked, and they should be at least conceptually integrated, connecting and nesting with each other. Valente observed that there are synergies between the phases: "If the formative and process evaluations are good enough, the summative evaluation takes care of itself."[34] Coffman suggested timing the data collection so that one phase is continually informing the other.[35]

Figure 7.1 maps the three evaluation phases to a notional sequence of activities in an IIP campaign process and the seven-stage PSYOP (now MISO) process.[36] Each box

---

[31] William D. Crano, "Theory-Driven Evaluation and Construct Validity," in Stewart I. Donaldson and Michael Scriven, eds., *Evaluating Social Programs and Problems: Visions for the New Millennium*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2003, p. 146.

[32] Author interview with Ronald Rice, May 9, 2013; interview with Thomas Valente, June 18, 2013.

[33] Julia Coffman, *Public Communication Campaign Evaluation*, Washington, D.C.: Communications Consortium Media Center, May 2002.

[34] Author interview with Thomas Valente, June 18, 2013.

[35] Author interview with Julia Coffman, May 7, 2013.

[36] Headquarters, U.S. Department of the Army, *Psychological Operations Leaders Planning Guide*, Graphic Training Aid 33-01-001, Washington, D.C., November 2005.

in the figure presents the generic IIP campaign element, with the corresponding stage of evaluation and the corresponding PSYOP process stage. Figure 7.2 presents characteristics and research activities associated with each of the three evaluation phases.

In defense doctrine, process evaluation is associated with measures of performance, which Christopher Rate and Dennis Murphy define in the IIP context as "cri-

Figure 7.1
The IIP Campaign Execution and Evaluation Process

Need identification and assessment Formative evaluation: problem definition and objective setting	**Designing and testing the message** Formative evaluation: focus groups, laboratory testing, etc.	**Produce the message** Process evaluation	**Collect baseline data** Basis for summative evaluation	**Disseminate the message/ implement the program** Process evaluation
PSYOP process steps: Phase I: planning Phase II: target audience analysis (TAA)	**PSYOP process steps:** Phase III: series development Phase IV: product development and design Phase VII: evaluation (send prototypes for pretest)		**PSYOP process steps (either or both):** Phase II: TAA Phase VII: evaluation (send prototypes for pretest)	**PSYOP process step:** Phase VI: production, distribution, and dissemination

RAND RR809/1-7.1

Figure 7.2
Characteristics of the Three Phases of IIP Evaluation

Formative evaluation	**Process evaluation**	**Summative evaluation**
Activities Focus groups In-depth interviews Secondary analysis Participant observation	Activities Implementation monitoring (e.g., viewer logs, broadcast schedule) Effects monitoring (e.g., sales data, visitation data, interviews)	Activities Analyze survey data Key informant interviews
Objectives, understand: Barriers to action Appropriate language Constellation of factors	Objectives, understand: Frequency of broadcasts Potential audience reach Preliminary data on effects	Objectives, understand: Level of effect Degree of efficiency
Design program	Launch program	Program ends

SOURCE: Based on a handout provided during author interview with Thomas Valente, June 18, 2013.
RAND RR809/1-7.2

teria used to assess friendly actions that are tied to measuring task accomplishment . . . and how well the influence activities involved are working (e.g., distribution of materials, campaign reach, how many people reached, etc.)." Summative evaluation is likewise associated with measures of effectiveness, which DoD guidance defines as criteria "used to assess changes in system behavior, capability, or operational environment that [are] tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect."[37]

Formative Evaluation Design

Formative evaluation is the preintervention research that helps to shape the campaign logic model and execution. Formative evaluation can define the scope of the problem, identify possible campaign strategies, provide information about the target audience, determine what messages work best and how they should be framed, determine the most-credible messengers, and identify the factors that can help or hinder the campaigns.[38]

Formative evaluation design can range from observational studies using focus groups, interviews, atmospherics, or baseline surveys to laboratory experiments for testing the efficacy of messages and media. Design considerations for formative research were not commonly discussed in our interviews, because the bulk of formative research consists of observational studies that do not seek to determine causality. However, as we address in Chapter Eight (which describes formative methods in detail), many SMEs indicated that more laboratory experiments should be conducted in the formative phase.

To inform decisionmaking, formative research must be turned around quickly. But Cole pointed out that this can be counterintuitive to academic researchers who want to conduct intricately designed and rigorous research.[39] In Coffman's experience, formative evaluations are often not done quickly enough to inform the subsequent campaign. In her view, doing formative evaluations well often has more to do with doing them quickly than the methods or approach employed. She suggested that DoD IIP evaluators look into "rapid response" or "real-time evaluation" methods commonly used to aid disaster relief and humanitarian assistance efforts.[40]

---

[37] U.S. Joint Chiefs of Staff, 2011b, p. GL-15.

[38] Coffman, 2002.

[39] Author interview with Charlotte Cole, May 29, 2013.

[40] Author interview with Julia Coffman, May 7, 2013. For more on real-time evaluation, see John Cosgrave, Ben Ramalingam, and Tony Beck, *Real-Time Evaluations of Humanitarian Action: An ALNAP Guide*, pilot version, London: Active Learning Network for Accountability and Performance in Humanitarian Action, 2009.

Formative evaluation should feed back into the logic model development and refinement process, including the specification of components and characteristics of the system within which the intervention is situated, allowing researchers to translate the theoretical model into a messaging campaign. For example, often you cannot directly influence the desired behavioral outcome, but you can influence a mediator who acts on the influences shaping behavior. Valente explained that formative research is what enables researchers and program designers to identify those mediators and to map the system logic model, including all of the different factors in the system that influence the outcome of interests. In his view, the "hardest part of all of this work is translating the theoretical factors into messages that people like and will respond to, which is why we do formative research—to translate the theory into a message."[41]

Process Evaluation Design

As shown in Figure 7.1, process or implementation evaluation can be conducted at several points in the campaign process, depending on the program logic model. *Message production evaluation* documents how the message was created. *Message dissemination evaluation* consists of measuring the *volume*, *channel*, and *schedule* (time and duration) for program dissemination.[42] While some researchers include measuring audience comprehension or exposure as process evaluation, this report addresses exposure measures separately.

Process evaluation serves several purposes and is underutilized. Process research can document implementation; guide program adjustments midimplementation; identify whether the necessary conditions for impact took place; identify the causes of failure (see the discussion of program versus theory failure in Chapter Five); identify threats to internal validity, such as contamination or interference from other campaigns; and generate information necessary for replicating and improving the program or campaign. Valente stressed that process evaluation is particularly important when programs fail but is frequently overlooked because researchers often assume that message implementation and reception are uniform.[43]

The *Information Environment Assessment Handbook* characterizes process assessment as any assessment function designed to improve the health or efficiency of an organization's internal system. For example, the Defense Readiness Reporting System (DRRS) focuses on a military organization's ability to carry out its critical missions, and, in the private sector, the Lean Six Sigma training program focuses on remov-

---

[41] Author interview with Thomas Valente, June 18, 2013.

[42] Valente, 2002, pp. 75–77.

[43] Author interview with Thomas Valente, June 18, 2013.

Box 7.2
The Importance of Tracking Interventions over Time

Collecting data for process evaluation serves the additional purpose of providing data on the key explanatory variable for summative evaluations (e.g., the timing and extent of the IIP intervention). This is an area in which DoD data collection efforts must improve. Process and summative evaluations in support of counterinsurgency operations in Iraq and Afghanistan have been complicated by a lack of data on U.S. efforts and activities. Jonathan Schroden voiced this concern: "We do a good job of cataloging what the insurgents do, but we do an abysmal job of cataloging what our own forces have done. This needs to be addressed systematically: It's impossible to know what's working if we don't know what we're doing."[a] Moreover, a vast amount of information is lost on U.S. activities during rotations.

Mark Helmke observed this problem across other U.S. IIP domains. To conduct trend analysis over time, "the U.S. needs to keep better records of its own engagements: What did we do? When? Whom did we network with?"[b] Collect and keep data not only on the details of the IIP efforts but also on other friendly force activities that can impact the relevant part of the information environment (remembering that any capability becomes an IRC when it affects the IE).

[a] Author interview with Jonathan Schroden, November 12, 2013.
[b] Author interview with Mark Helmke, May 6, 2013.

ing variation in an organization's critical processes.[44] In summary, process evaluations should be incorporated into the assessment cycle as complements to summative evaluations. Process evaluations are typically less involved from a design perspective but provide a valuable means to test hypotheses about why the program failed or fell short of theoretical optimal outcomes.

Summative Evaluation Design

Summative evaluations consist of postintervention research designed to determine the outcomes that can be attributed or tied to the IIP intervention or campaign. Determining causality—or the extent to which one or more influence activities contributed to or was responsible for a change in knowledge, attitudes, or behaviors—is a chief goal of summative IIP evaluation.[45] Summative evaluation designs can be classified as *experimental*, *quasi-experimental*, or *nonexperimental*. This section discusses the application of these designs to IIP evaluation, including strengths and weaknesses, variations, and notable examples.

In experimental designs, subjects are randomly assigned to treatment and control conditions and are observed, at minimum, after treatment. Experimental designs have the highest internal validity and therefore the strongest basis for causal inference. Quasi-experimental designs, or "natural experiments," such as longitudinal or cross-sectional exposed-versus-unexposed studies, are similar to experimental designs except

---

[44] The Initiatives Group, 2013, p. 3.

[45] Valente, 2002, p. 89; interview with Kavita Abraham Dowsing, May 23, 2013.

that the researchers cannot randomly assign subjects to treatment or control groups. Quasi-experimental evaluation designs can be mixed method, incorporating qualitative components. Quasi-experimental designs have lower internal validity than experimental designs but are often much more practical and cost-effective. Nonexperimental studies do not have a control and therefore have limited to no ability to make causal claims regarding the contribution of the program to outcomes but can nonetheless be useful for gathering information on perceptions of the campaign.

Within those broad categories there are many design variations, some of which are described below. Organizations with effective research groups often use several designs. The Sesame Workshop, renowned for its strong research culture and effective programming, uses four designs: experimental designs with random assignment, pre- and posttesting without a control (quasi-experimental longitudinal design), exposed-versus-unexposed post-only testing with a comparison (quasi-experimental cross-sectional design), and commissioned general market studies on reach and perceptions of the show (nonexperimental).[46]

Experimental Designs in IIP Evaluation

Experimental designs are characterized by random assignment to treatment and control conditions. The treatment group is given the intervention, while the control group receives no intervention or an innocuous one, and outcomes are observed for both groups. Often called the "gold standard" for assessing causal effects, randomized experiments are the most valid way to establish the effects of an intervention.[47] In a *field experiment*, researchers examine the effects of an intervention in its natural setting. In a *randomized controlled trial*, the intervention is designed by the researchers.[48]

Table 7.2 identified six types of experimental designs based on the number of groups or cohorts and when they are tested. The postprogram-only two-group design (number 1B in Table 7.2) consists of two groups that are observed only after the intervention. The post-program only with propensity matched groups (number 1C) is similar, but makes comparisons within groups of people who were equally likely to be exposed (propensity matching). In the pre- and postprogram two-group design (number 4), both groups are pretested prior to the intervention. In the pre- and postprogram two-group design with post-only treatment group (number 5), an additional, un-pretested treatment group is added to control for the sensitization effect of testing. The Solomon four-group design (number 6) adds an additional un-pretested control group and is the strongest design, because it controls for all potential threats to internal validity.

---

[46] Author interview with Charlotte Cole, May 29, 2013.

[47] Rossi, Lipsey, and Freeman, 2004, p. 237.

[48] Author interview with Devra Moehler, May 31, 2013.

Box 7.3
The Use of Experimental Designs for Evaluating IIP Activities: The Impact of Partisan Radio Stations in Ghana

Researchers at the Annenberg School for Communication at the University of Pennsylvania recently completed a field experiment in Ghana regarding the impact of exposure to partisan radio stations on the attitudes and political behaviors of citizens riding in public transportation (*tro tros*—privately owned minibus share taxis). The design and methods were innovative and could potentially be adapted to measuring the effects of DoD IIP activities. The study, which was led by Jeffrey Conroy-Krutz and Devra Moehler, used a four-group posttest-only design with two controls. *Tro tro* riders in Ghana were randomized to one of four conditions: a partisan radio station supporting the government, a partisan station supporting the opposition, a neutral political talk show, or no radio station. Subjects were interviewed after they departed the *tro tro*.[a]

A *tro tro* carries passengers and goods in Accra, Ghana. Creative Commons photo by Eileen Delhi.

The researchers were interested in the impact of partisan radio on four measures: (1) attitudes toward politicians of other parties; (2) ethnic discrimination; (3) support for electoral malfeasance; (4) participation and engagement. In addition to survey questions designed to elicit relevant attitudes and behavioral intentions (stated preference), Conroy-Krutz and Moehler used *behavioral measures* to assess how respondents actually behave (revealed preferences) by testing how they respond to certain scenarios. Behavioral measures used by the researchers included: (1) giving the participants money for participating and then asking them to donate a portion of that money to a cause associated with one side or the other of the partisan split; (2) giving them a choice of key chains, each associated with a different party of the government; and (3) asking them to join a petition about transportation policy by texting a number, which would measure political efficacy and engagement. These behavioral measures provide an innovative and cost-effective technique for addressing the bias inherent in self-reported attitudinal measures when measuring IIP effects.[b]

As of May 2013, the researchers had found that exposure to partisan stations made riders more sympathetic to opposing viewpoints, which countered the researchers' expectations and intuition. The researchers also included demographic and psychographic measures to validate randomization (e.g., wealth, age, ethnicity, reported partisanship) and to subdivide the population to test whether there were differential effects for people with higher education or political engagement.[c]

The study was supported by the research team at BBC Media Action, which sponsored the study to explore the utility and feasibility of using field experiments to measure effects of its programming.[d]

[a] Author interview with Devra Moehler, May 31, 2013. The results of the Ghana experiment have been documented in a working draft (Conroy-Krutz and Moehler, 2014).

[b] Author interview with Devra Moehler, May 31, 2013.

[c] Author interview with Devra Moehler, May 31, 2013.

[d] Author interview with James Deane, May 15, 2013; interview with Kavita Abraham Dowsing, May 23, 2013; interview with Kavita Abraham Dowsing, May 23, 2013.

The Appropriateness of Experimental Designs for IIP Evaluation: Causal Inference Is Costly

The use of experimental designs for IIP evaluation faces several challenges:

- *Contamination of the control group.* It is difficult to control assignment or limit exposure to the program because subjects share information and move.[49]
- *Resistance to deliberately limiting the reach of the program.* In addition to challenges with inadvertent contamination of the control group, program managers are often reluctant to construct a control group, because programmers want as broad of an audience as possible, and isolating exposure to the treatment group limits the reach of their program. Pamela Jull, president of Applied Research Northwest, a consulting firm that focuses on social marketing efforts in the Pacific Northwest, stated that most program managers think that everyone getting the treatment is worth more than ensuring that there is a comparison group for research purposes.[50] Andrew Hall, deputy country representative in the Office of Transition Initiatives at USAID, explained that the nature of USAID's work does not permit it to identify a place in advance where it will deliberately not implement a program.[51] In such cases, a time-lagged control (providing the intervention to the control group after a delay) might be suitable, as discussed above.
- *Feasibility of constructing a control group.* When it comes to evaluating geostrategic decisions and impacts, controlled experiments are impossible. Nicholas Cull, historian and director of the master's program in public diplomacy at the University of Southern California, illustrated this point with the example of trying to prove that the military intervention in Haiti owed its success to the communication campaign: "What are you going to do? Run a controlled experiment where you invade a Caribbean island without explaining it to everyone?"[52]
- *Low external validity.* When randomization occurs at the level of the individual, researchers typically have to encourage or deliberately ask the subjects to watch the media. Thus, they cannot observe how the subjects would engage the media "in the wild," when exposure is self-selected. This problem can be minimized by randomizing at the group level, where capacity to be exposed is the treatment (e.g., living within a region where the media are shown), but there is a high risk of contamination with those designs.[53]

---

[49] Author interview with Thomas Valente, June 18, 2013.

[50] Author interview with Pamela Jull, August 2, 2013.

[51] Author interview with Andrew Hall, August 23, 2013.

[52] Author interview with Nicholas Cull, February 19, 2013.

[53] Author interview with Charlotte Cole, May 29, 2013; interview with Marie-Louise Mares, May 17, 2013; interview with James Deane, May 15, 2013.

- *Time and resource requirements.* Several SMEs noted that experimental designs are cost prohibitive in most cases due to costs associated with designing and implementing the intervention, providing incentives to participants, and taking the time to recruit a sufficient number of participants.[54]
- *Human subjects concerns.* Victoria Romero, a cognitive psychologist with extensive experience in IIP evaluation across the commercial and government sectors, described the need to protect human subjects and minimize deception as one of the chief challenges to conducting field experiments in places like Afghanistan. Moreover, subjects in these environments are often unwilling to participate in experiments.[55]
- *Ethics.* Paul Hepper and colleagues note that "there are times when it can be unethical to withhold treatment from certain groups of participants."[56]

Box 7.4
The Use of Experimental Designs for Evaluating IIP Activities: The Effectiveness of a Radio Campaign to Reduce Child Mortality in Burkina Faso

In March 2012, Development Media International and the London School of Hygiene and Tropical Medicine began a two-and-a-half-year cluster-randomized trial to test the impact of a radio campaign targeting all causes of child mortality, in Burkina Faso, West Africa. The evaluation design involves broadcasting the behavior change campaign to seven randomized geographic areas across Burkina Faso, and using seven additional clusters as controls. The researchers are able to limit contamination because Burkina Faso has "very localized, radio-dominated media environments" enabling them to use local FM radio stations to broadcast their messages to intervention areas without exposing (contaminating) the control clusters.[a]

The evaluation includes baseline and end line mortality surveys with a sample size of 100,000. According to Development Media International, it is "the most robust evaluation that has ever been conducted of a mass media intervention in a developing country." The study is funded by the Wellcome Trust and the Planet Wheeler Foundation. Full results, including data on child mortality outcomes, are expected to be published in late 2015.[b]

A woman transports water with her baby in Sorobouly, Burkina Faso. Creative Commons photo by Ollivier Girard for the Center for International Forestry Research (CIFOR).

[a] Development Media International, "Proving Impact," web page, undated.

[b] Development Media International, undated.

---

[54] Author interview with Marie-Louise Mares, May 17, 2013; interview with Kavita Abraham Dowsing, May 23, 2013; P. Paul Heppner, Dennis M. Kivlighan, and Bruce E. Wampold, *Research Design in Counseling*, 3rd ed., Belmont, Calif.: Thomas Higher Education, 2008.

[55] Author interview with Victoria Romero, June 24, 2013.

[56] Heppner, Kivlighan, and Wampold, 2008.

Given the difficulties associated with experimental designs and the challenges noted above, the appropriateness and relative value of experimental designs depend on the importance of making causal inference relative to other objectives and constraints.[57] Experimental manipulation is the "only way to truly isolate out differential effects" and provides "the best possible evidence for drawing conclusions about causal inference."[58] Because identifying causal mechanisms and eliminating rival explanations will save resources and improve effectiveness in the long run, experimental designs should be

---

Box 7.5
The Use of Experimental Designs for Evaluating IIP Activities: Matched-Pair Randomized Experiments to Evaluate the Impact of Conflict Resolution Media Programs in Africa

An Oxfam-sponsored radio broadcast on security issues facing the community in Dungu, eastern Democratic Republic of Congo. Creative Commons photo by Oxfam International.

Elizabeth Levy Paluck conducted a group-based randomized experiment to evaluate a reconciliation-themed radio soap opera in Rwanda, and she used matched-pair randomization at the level of listening groups. Communities were sampled to represent political, regional, and ethnic breakdowns and then were matched into pairs with a similar community according to several observable characteristics, such as gender ratio, quality of dwelling, and education levels. Then, "one community in each pair was randomly assigned to the reconciliation program and the other to the health program. This stratification of sites helped to balance and minimize observable differences between the communities ex ante."[a]

Paluck used a related design in eastern Democratic Republic of Congo. The study used randomized pair-wise matching within clusters to evaluate the impact of a radio soap opera when aired in conjunction with a talk show that emphasized conflict reduction through community cooperation. Paluck pair-wise matched regions and randomly chose one treatment and one control region within each pair. The radio program was aired in all of the experiment's regions, but the talk show that followed the radio show, designed to encourage listeners' reactions and discussions, was only broadcast in treatment regions. She found that the listeners who were encouraged by the additional talk show to discuss did discuss more, but they were also more likely to become intolerant and less likely to help outcast community members.[b]

[a] Elizabeth Levy Paluck, "Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda," *Journal of Personality and Social Psychology*, Vol. 96, No. 3, March 2009, pp. 577–578.

[b] Elizabeth Levy Paluck, "Is It Better Not to Talk? Group Polarization, Extended Contact, and Perspectives Taking in Eastern Republic of Congo," *Personality and Social Psychology Bulletin*, Vol. 36, No. 9, September 2010; Marie Gaarder and Jeannie Annan, *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*, New York: World Bank Independent Evaluation Group, June 2013.

---

[57]   William D. Crano and Marilynn B. Brewer, *Principles and Methods of Social Research*, 2nd ed., Mahwah, N.J.: Lawrence Erlbaum Associates, 2002, p. 17.

[58]   Author interview with Devra Moehler, May 31, 2013; interview with Charlotte Cole, May 29, 2013.

used if and when they can feasibly and affordably be conducted. However, experimental designs should be supplemented with quasi-experimental and qualitative research to enhance the generalizability of the findings. While experimental research plays a vital role, researchers should "not be forcing it in circumstances that do not lend itself to it," and "it is an equally big mistake to prioritize RCTs [randomized controlled trials] at the expense of naturalistic studies that look at how people engage the media naturalistically."[59]

While implementing experimental designs is often impossible or cost prohibitive in the summative phase, several SMEs commented on the need for more experimental designs for product and message testing in the formative phase. Valente, for example, argued that randomized controlled trials are "underutilized" in the formative phase and "can be incredibly valuable in testing and guiding the development of messages."[60] This topic is discussed further in Chapter Eight, on formative research methods.

A Note on Survey Experiments

Surveys conduct randomized behavioral experiments on respondents by varying one or more elements of the survey (treatment conditions) across subjects. Survey experiments are cost-effective alternatives to randomized controlled trials or large-scale field experiments where the treatment intervention occurs at a program level, and should be used more often in IIP evaluation. Because these experiments typically do not include pretest measurements, they can be considered a *posttest-only two-group design*.

In a study on voter behavior in Uganda, Devra Moehler and her colleagues used a survey experiment to test the effect of ballot design, such as the inclusion of party names or symbols, on voter behavior. The researchers administered a survey prior to the election featuring a sample ballot that respondents were asked to fill out with four different treatment conditions relating to ballot design. The respondents knew that they were participating in a study but did not know what treatment they were subject to. Other survey questions enabled the researchers to control for demographics and partisan orientation. Survey experiments are relatively easy ways to conduct an experiment in an environment in which you have some infrastructure for administrating a survey and should be considered in the evaluation of the effectiveness of different DoD IIP branding messages or designs.[61]

---

[59] Author interview with Devra Moehler, May 31, 2013; interview with Charlotte Cole, May 29, 2013.

[60] Author interview with Thomas Valente, June 18, 2013.

[61] Author interview with Devra Moehler, May 31, 2013. For more on survey experiments, see Devra C. Moehler, Jeffrey Conroy-Krutz, and Rosario Aguilar Pariente, "Parties on the Ballot: Visual Cues and Voting Behavior in Uganda," paper presented at the International Communication Association annual conference, Boston, Mass., May 26–30, 2011.

Quasi-Experimental Designs in IIP Evaluation

If randomized experiments are not feasible, affordable, or appropriate, quasi-experiment designs are the next best option for making causal inference. These designs, also called *nonequivalent group designs*, are similar to experimental designs, except that the evaluators lack control over who receives the intervention and who does not, often due to qualities of the intervention that make it inherently available to everyone. In these studies, analytic comparisons are made between groups that were not formed through the use of random assignment and therefore may differ on characteristics that are relevant to the outcome of interest. As a consequence, these designs have lower internal validity than true experiments because it is impossible to eliminate the rival explanation that some other characteristic differentiating the treatment and control groups explains the observed outcomes. In terms of internal validity, quasi-experiments are the next best alternative to experimental designs. Because randomized experiments are infrequently feasible, affordable, and appropriate, quasi-experiments are more commonly used in IIP evaluation and could be leveraged considerably in DoD IIP assessment efforts. Quasi-experiments have "the potential to provide insights that would have been lost due to constraints that make it difficult/impossible to research the issues through use of standard true experimental techniques."[62]

Often called "natural experiments," these designs take advantage of natural variation in exposure to the program, such as time, natural variations in treatment, and self-determination of exposure.[63] Quasi-experimental studies can include cross-sectional, panel or cohort, time-series, event-history or survival-analysis, and mixed-method evaluations with qualitative components. In a cross-sectional design, data are collected at one point in time. Cross-sectional studies are typically conducted when generalizability is important, the population is difficult to access, data must be anonymous, or the theory being tested is new. A longitudinal or time-series design collects observations over time and can include panel, cohort, or repeated-measure studies. In a panel design, the same respondents are interviewed repeatedly. A cohort is a single panel.[64] Repeated-measure studies collect cross-sectional data over time, but always from the same subjects.

To produce interpretable results, quasi-experimental or nonequivalent group designs must use a pretest or a proxy pretest and at least two groups. In a posttest-only nonequivalent two-group design, evaluators cannot assess whether differences in observed outcomes are due to the treatment or due to preexisting differences between the groups. Conversely, in a one-group pretest-posttest design, evaluators cannot know whether observed changes are due to the treatment or due to naturally occurring or

---

[62] Crano and Brewer, 2002, p. 150.

[63] Coffman, 2002.

[64] Valente, 2002, pp. 259–261.

exogenous changes over time.[65] Pretests and comparison groups can be constructed post facto through the use of propensity score matching, as discussed below.

Comparing Exposed and Unexposed Individuals

Posttest-only survey-based studies comparing those who have self-reported exposure with those who have no or lower exposure are the primary way that media-effects research is done by M&E practitioners in the field. In the exposed-versus-unexposed design, the comparison group is constructed based on self-reported lack of exposure.[66] Unfortunately, as Moehler points out, in many cases there is not even a comparison group because the researchers will only survey those whom they knew to be exposed, making causal inference very difficult.[67] Kavita Abraham Dowsing, director of research at BBC Media Action, argued that while this design may be considered the weakest of the methodologies in the impact evaluation world, it might be the best for media evaluations due to the challenges associated with isolating a control group.[68]

The chief challenge with exposed-versus-unexposed quasi-experimental designs is selection. Because the exposed treatment group is self-selected, it is likely that it is systematically different from the unexposed control group in ways that might correlate with the behavioral outcomes. For example, individuals that choose to watch a DoD-funded television commercial may be predisposed to supporting the aims of the coalition. As a consequence, it is difficult to know whether differences in self-reported attitudes and behaviors are due to exposure or due to those preexisting differences in dispositions.

The better you can estimate how the exposed group would have responded in the absence of the media, the better you'll be able to determine the media's effects. Moehler summarized several techniques for deriving these estimates, each with varying degrees of reliability and feasibility. These include a baseline survey, making comparisons within groups of people who were equally likely to be exposed (propensity matching), controlling for variables and events that co-occurred with the media, and in-depth qualitative interviews that help determine how the media affected someone.[69] For measuring educational outcomes, the Sesame Workshop uses measures of intrinsic cognitive abilities, such as "digit span," to control for selection bias.[70]

---

[65] Heppner, Wampold, and Kivlighan, 2008.

[66] Author interview with Devra Moehler, May 31, 2013; interview with James Deane, May 15, 2013.

[67] Author interview with Devra Moehler, May 31, 2013.

[68] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[69] Author interview with Devra Moehler, May 31, 2013.

[70] Author interview with Charlotte Cole, May 29, 2013.

Comparing Exposed and Unexposed Communities

Instead of comparing differences between those who self-report exposure and those who do not, evaluators can assess differences between communities that were exposed and those that were not. In Afghanistan, for example, there have been correlations between community-level exposure to advertising and rates of enlistment.[71] Where feasible, this approach is preferable to comparing individual differences within the same target area, because it at least partially controls for selection bias. In this case, the reasons that individuals were not exposed are less likely to be due to predispositions that are correlated with the behavioral outcomes of interest. Sean Aday, director of the Institute for Public Diplomacy and Global Communication at George Washington University, contended that the most feasible quasi-experimental design in an environment like Afghanistan would be to compare outcomes, over the long run, in places where you had a communication intervention with an otherwise similar place without the intervention.[72]

There are nonetheless still several threats to internal validity inherent in these designs. Acknowledging that inferring causality in IO is extremely difficult, Romero noted that "the best you can do is compare outcomes in communities that were exposed to those that were unexposed, but you can't control for contamination or rival explanations, like different levels of poverty or safety," that might influence reported attitudes and behaviors.[73] These threats to internal validity can be minimized through the use of baseline surveys and controlling for confounding variables, but it is often difficult to identify and measure all necessary controls.[74]

Coffman identifies several sources of natural variation in exposure to the intervention that can be leveraged by these quasi-experimental cross-sectional designs. If a campaign is rolled out in different phases with time lags between each phase, "the evaluation can compare areas that were exposed to the campaign in its early stages to those that have yet to be exposed." In other cases, the implementation is bound to fail or not be implemented as intended in certain areas, which can provide useful comparisons. Moreover, some individuals will not be exposed to a campaign because, for example, "they might not have a television or listen to the radio or read the newspaper."[75]

Spillover effects, where awareness spreads from one community to another even when activities do not, can threaten the validity of this approach. When USAID has tried to compare communities that have not had projects with those that have, it has found that individuals in the unexposed communities were aware of the projects.

---

[71] Author interview with Matthew Warshaw, February 25, 2013.

[72] Author interview with Sean Aday, February 25, 2013.

[73] Author interview with Victoria Romero, June 24, 2013.

[74] Author interview with Thomas Valente, June 18, 2013.

[75] Coffman, 2002, p. 28.

This may create bias, because the respondents in the unexposed communities want to answer in a way that either encourage or discourages conducting the intervention in their community.[76]

Split or "A/B" Testing

Split testing, also known as "A/B" testing, is a popular design in marketing and electioneering that allows researchers to examine the relative effectiveness of direct-marketing messages. This design can be implemented as either an experimental or quasi-experimental design. As an experimental design, it is a variant of the two-group pretest posttest design. A/B testing involves sending two variants of a message (A and B, or treatment and control) to two groups of customers that are psychographically and demographically identical, and then measuring the differences in consumer responses to the two messages. The treatment variant of the message should differ only in one respect from the control variant. Split testing has gained tremendous popularity among Internet start-ups and is endorsed as a source of *validated learning* by the *lean start-up* methodology.[77] As a quasi-experimental design, split testing becomes a variant of exposed-versus-unexposed communities or individuals, except that groups or individuals are exposed to different interventions.

Comparing Recently and Less Recently Exposed Communities: Time Lag for Control

Another variant of this design uses a time lag for control. This variant is again similar to exposed-versus-unexposed designs, except that unexposed is changed to "exposed later." So group A receives the intervention at time 1, but group B does not; then, at time 2, group B receives the same intervention. Measures are taken at baseline, between time 1 and time 2, and at end line. The difference between group A and B before time 2 is attributable to the intervention. Such a design is particularly useful in the DoD IIP context, as executors will often want to conduct their efforts across the whole population, not leaving a segment uninfluenced as a control. This design eventually allows everyone to receive the intervention, but still allows for a temporary control (and a certain degree of causal inference).

Propensity Score Matching Posttest-Only Exposed-Versus-Unexposed Designs

A variation on the exposed-versus-unexposed design uses *propensity score matching*. In the context of IIP evaluation, this technique involves comparing those who were exposed with those who were unexposed, within groups that had similar likelihoods, or propensities, to be exposed to the media, as determined by overlapping responses to other survey questions. In this case, both the pretest and the control group are constructed post facto based on their responses to the postintervention survey. The prin-

---

[76] Author interview with Andrew Hall, August 23, 2013.

[77] Eric Ries, *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, New York: Random House Digital, 2011.

cipal benefit of this approach is that it allows the researchers to control for the selection bias that plagues typically exposed-versus-exposed studies without the need for a representative sample or baseline data. For example, survey research may indicate that respondents' likelihood of seeing a DoD-sponsored television commercial is a function of their media-viewing habits, political orientations, ethnicities, religions, and education levels. A propensity-score-matching design to evaluate the efficacy of a particular television-commercial campaign would survey individuals within the target audience and ask questions within three categories: the extent to which they were exposed, their propensity to be exposed as measured by the aforementioned predictors, and outcomes (their knowledge, attitudes, and behaviors). The analysis would then compare the outcomes between individuals who were exposed and those who were unexposed but had overlapping scores on the propensity measures.

Johanna Blakely, managing director of the Norman Lear Center at the University of Southern California's Annenberg School for Communication and Journalism, advocates for the use of propensity-score-matching techniques in media research based largely on her experience with the Measuring Media's Impact project, which used a propensity-score-matching survey instrument to measure the efficacy of the *Food, Inc.* documentary. She stated that propensity-score-matching designs control for selection bias, avoid the need for pretests or baselines, do not require a representative sample, and preserve the intellectual firewall between the programmers and the researchers, because the researchers do not have to be involved with or embedded into the project design from the outset. Separating the evaluators from the programmers avoids "scaring away creators from the evaluation process, . . . because the creative side just wants to make something great and doesn't want it to be engineered from the beginning." This runs contrary to the conventional wisdom that the planners and researchers should collaborate throughout the duration of the intervention.[78]

Limitations to propensity-score-matching techniques include regression to the mean and uncertainty surrounding the similarities of the two groups. If a variable is extreme on the pretest measurement, it will tend to be closer to the mean on the second measurement. This is also known as reversion to the mean and reversion to mediocrity. If two different groups are selected because of extreme scores within their groups that happen to overlap, one can expect those scores to regress toward the mean when measured again. The use of propensity score matching is promising, but the following caveat should be kept in mind: "The fundamental condition of strong ignorability that is necessary for the causal interpretation of treatment effects in the nonequivalent control group design can be probed, but never definitively established. Thus, there is always a degree of uncertainty associated with estimates of causal effects on the basis of this design."[79]

---

[78] Author interview with Johanna Blakley, June 24, 2013.

[79] Stephen G. West, Jeremy C. Biesanz, and Steven C. Pitts, "Causal Inference and Generalization in Field Settings: Experimental and Quasi-Experimental Designs," in Harry T. Reis and Charles M. Judd, eds., *Handbook of Research Methods in Social and Personality Psychology*, New York: Cambridge University Press, 2000, p. 73.

Box 7.6
Quasi-Experimental Designs for Evaluating IIP Activities: Propensity Score Matching to Measure the Impact of *Food, Inc.*
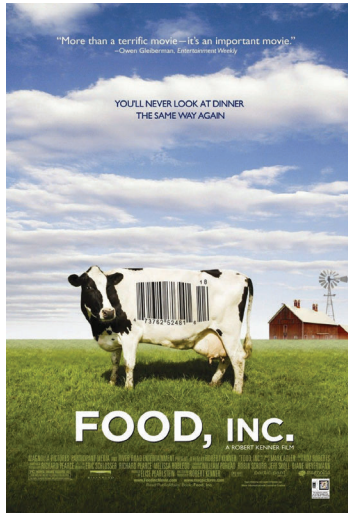
The Measuring Media's Impact project at the University of Southern California's Norman Lear Center developed an innovative survey instrument using the propensity-score-matching technique to measure the impact of the *Food, Inc.* documentary. The survey included questions measuring respondents' propensity to watch the film, and the researchers compared outcomes between those who were exposed and those who were not but had identical propensity scores. The methodology addresses the problem of selection bias, because the large sample size "enabled the researchers to create a detailed profile of likely viewers of the film, and to compare very similar viewers who saw the film with those who did not," allowing the researchers to "construct something similar to a classical study design where individuals are randomly assigned to a treatment and control group."[a] The survey generated approximately 20,000 responses from a sample drawn from email lists and social media. Although the sample was not representative, Johanna Blakely does not see this as a significant issue, because the propensity-score-matching technique can compare the efficacy of a message among people who are not particularly socially engaged.[b]

Food, Inc. viewers were significantly more likely than nonviewers to encourage their friends, family, and colleagues to learn more about food safety, shop at local farmers' markets, eat healthful food, and buy organic or sustainable food. The nonviewers were virtually identical on 17 traits, including their degree of interest in sustainable agriculture and past efforts to improve food safety.[c]

[a] Author interview with Johanna Blakley, June 24, 2013. For more on the Lear Center project, see Norman Lear Center, "Research Study Finds That a Film Can Have a Measurable Impact on Audience Behavior," press release, February 22, 2012. The key findings were first announced by Blakley at TEDxPhoenix; see Johanna Blakley, "Movies for a Change," presentation at TEDxPhoenix, February 12, 2012.

[b] Author interview with Johanna Blakley, June 24, 2013.

[c] Author interview with Johanna Blakley, June 24, 2013.

The Bellwether Method

For campaigns aimed at a small or specific target audience, the *bellwether* method is a cost-effective alternative to large-scale exposed-versus-unexposed quasi-experiments. This method, used principally in advocacy evaluation, measures the extent to which a public-communication campaign is influencing key individuals, typically decision-makers. The method consists of highly structured interviews with two twists. First, the sample is limited to high-profile policymakers or decisionmakers, half of which the researchers are fairly certain were exposed to the messages, and the other half were less likely to be exposed. Second, the researchers are very vague about the subject of

the interview prior to holding it. A major advantage to the bellwether method is that it does not require large sample sizes.[80]

The bellwether method was used by Coffman and colleagues to evaluate the efficacy of a campaign for preschool advocacy, for which they interviewed 40 decision-makers and thought leaders. The researchers told the interviewees that they would interview them about education, but not necessarily about early childhood, allowing the recall measures to be unprompted. The researchers also had to be as vague about their objectives during the interview for as long as possible—though eventually it becomes clear and they can ask specific questions about their exposure.[81]

Longitudinal Designs: Time Series and Repeated Cross Sections or Panels

The designs described here (e.g., exposed versus unexposed) are often implemented as cross-sectional quasi-experiments in that the measurements are taken at one point in time (after the intervention). An alternative or complementary approach is to use a *longitudinal* quasi-experimental design, in which measurements are taken of the same or of different groups over time. According to renowned social psychologist Anthony Pratkanis, for evaluating DoD influence activities, the "most feasible and effective designs track the dependent variable over time and see how it tracks with the intervention."[82]

Longitudinal designs include time-series and repeated cross-sectional designs. In a time-series design, also called a *cohort study*, the same population is observed over time. In a repeated cross-sectional design, different populations are observed over time—e.g., control and treatment or exposed and unexposed populations. However, the individuals sampled from the population are not necessarily the same over time. Panel designs are a variant of the repeated cross-sectional approach that resamples the same individuals at each point in time. Other time-series subtypes include interrupted time series, comparison time series, and regression-discontinuity designs.[83]

A simple variant of the repeated cross-sectional design that is pertinent to DoD IIP uses a time lag as a control so as to avoid deliberately shielding a segment of the target audience from the program. In the first stage, group A is exposed to the intervention, but group B isn't. In the second stage, group B is exposed to the intervention. Measurements are taken at baseline and before the second phase. The difference between the two groups before the second phase represents the contribution of the intervention.

---

[80] Author interview with Julia Coffman, May 7, 2013. For more on the bellwether method, see Julia Coffman and Ehren Reed, *Unique Methods in Advocacy Evaluation*, Washington, D.C.: Innovation Network, 2009.

[81] Author interview with Julia Coffman, May 7, 2013.

[82] Author interview with Anthony Pratkanis, March 26, 2013.

[83] Crano and Brewer, 2002.

For time-series analysis, the data must be periodic, accurate, reliable, consistent, sufficient, and diverse.[84] *Rolling sample surveys* use daily surveys drawn from an independent sample to measure audience exposure and attitudes. These allow evaluators to track the day-to-day shifts in sentiments and behavior, providing opportunities for natural experiments when IIP interventions take place.[85] If possible, researchers should evaluate the dependent variable (the outcome of interest) for several years before and after the intervention. For example, Juan Ramirez and William Crano found little immediate impact of the "three-strikes law" but saw that it had a long-term impact on instrumental crime over time.[86]

Panel studies, in which the same individuals are sampled over time, can be useful with smaller sample sizes and in cases where the researchers want to explore the causal mechanisms for observed outcomes. One data collection method for audience analysis is the use of "people meters" (e.g., the Nielsen families) who agree to have their media consumption tracked with boxes or diaries. This passive mechanism for data collection can provide cost efficiencies and a solution to potential longitudinal survey fatigue but will suffer from selection bias in hostile environments.[87]

Interviewed SMEs have had mixed experiences with panel studies. InterMedia evaluations often use a static panel over time with a control group.[88] Altai Consulting attempted to conduct a panel survey but found that it was too complicated in the conflict environments it operates in, because cell phone numbers change and it is hard to locate the same people again if when door-to-door.[89]

Online panels are becoming increasingly popular; in them case participants agree to be surveyed a number of times throughout the year, typically in exchange for payment. Participants are selected based on their demographics or psychographics. These methods can be biased if conducted in an area in which the online population is not representative of the general population. While that concern is less relevant in the United States, it could present serious bias in conflict environments.[90]

Nonexperimental Designs

At the bottom of the hierarchy of design rigor are those designs that do not involve a control or comparison group, also known as *nonexperimental designs*. These designs

---

[84] Author interview with Thomas Valente, June 18, 2013.

[85] Coffman, 2002, p. 27.

[86] Juan R. Ramirez and William D. Crano, "Deterrence and Incapacitation: An Interrupted Time Series Analysis of California's Three Strikes Law," *Journal of Applied Social Psychology*, Vol. 33, No. 1, January 2003.

[87] Author interview with Maureen Taylor, April 4, 2013.

[88] Author interview with Gerry Power, April 10, 2013.

[89] Author interview with Emmanuel de Dinechin, May 16, 2013.

[90] Author interview with Julia Coffman, May 7, 2013

have the weakest internal validity because, without a comparison group, there is no counterfactual to help determine what would have happened had the intervention not taken place. However, evaluations conducted with these designs can help develop hypotheses regarding likely effects that can be validated with more-rigorous methods. Despite their limitations, there are situations in which nonexperimental designs are the best option because a comparison group cannot be isolated or constructed post facto, as is often the case with complex, multistage interventions aimed at achieving behavioral change over the long term. This section discusses two nonexperimental designs addressed by experts interviewed for this report: case studies and frame evaluations. Expert elicitation, which is addressed in Chapter Eight, may also be considered a nonexperimental design.

Frame Evaluation Research

Framing analysis studies how issues or ideas are discussed in the media or within the target audience by looking for "key themes, expressed as arguments, metaphors, and descriptions to reveal which parts of the issue are emphasized, which are pushed to the margins and which are missing."[91] Taylor suggested using *frame evaluation research* to estimate the causal relationship between the intervention and observed changes in attitudes. These designs assess whether the particular frame used by the intervention has been used or adopted by the target audience, providing a means to estimate the extent to which the audience's change in attitude was due to the intervention or due to something else. For example, instead of simply measuring the population's attitudes toward the coalition forces or to the Afghan government, measures should seek to elicit information about how the audience is framing and rationalizing those attitudes, and compare those frames with the arguments made by the intervention. Content analysis, focus groups, surveys, and other data collection methods can be used to collect framing data that capture the specific frames, standards, or principles being used by the target audience.[92] Frame evaluation designs are a simple and cost-effective tool for estimating whether an intervention is influencing the population, but they have weak internal validity.[93]

Framing analysis is often informally used to measure effects. For example, Paul Bell used a similar logic when discussing the evidence that demonstrated the impact of Information Operations Task Force (IOTF) in Iraq, an operation that he oversaw while he was the CEO of Bell Pottinger. He cited the example of a *New York Times* article that quoted an Iraqi colonel saying he was going to vote because he was not

---

[91] John McManus and Lori Dorfman, "Silent Revolution: How U.S. Newspapers Portray Child Care," *Issue* (Berkeley Media Studies Group), No. 11, January 2002, p. 10.

[92] Author interview with Maureen Taylor, April 4, 2013.

[93] For a primer on frame evaluation research, see Robert M. Entman, "Framing: Toward Clarification of a Fractured Paradigm," *Journal of Communication*, Vol. 43, No. 4, December 1993.

Box 7.7

Quasi-Experimental Designs for Evaluating IIP Activities: International Media and Exchange Efforts to Improve Health and Combat Human Rights Abuses

Impact of BBC Programming on HIV/AIDS-Related Knowledge and Behaviors

In a study sponsored by BBC Media Action, Joyee Chatterjee and colleagues used data from a survey of 834 sexually active men to assess the influence of exposure to a BBC program on HIV/AIDS-related awareness, attitudes, and behaviors. Respondents were matched on gender, age, education, and location. Using structural equation modeling, the researchers were able to show that people exposed to the campaign had higher awareness and knowledge of HIV/AIDs-related issues and that knowledge change predicted attitudinal change. However, the link between attitudinal and behavioral change was mediated by self-efficacy and interpersonal discussion.[a]

Impact of a Reintegration Program on Ex-Combatants in Burundi

Michael Gilligan and colleagues exploited a random disruption in program implementation to construct a control group to evaluate the impact of a reintegration program on ex-combatants in Burundi. Three organizations were given contracts to administer the program, but one delayed providing services for a year for reasons apparently unrelated to predictors of effectiveness. To control for potential systematic differences between individuals in the control group, participants in the treatment and control groups were matched on individual characteristics, community characteristics, and propensity scores.[b]

The Effectiveness of a National Campaign to Promote Family Planning and Reproductive Health in Bolivia

Thomas Valente and colleagues conducted an evaluation of a mass media campaign to promote family planning and reproductive health in Bolivia between 1987 and 1999. Because of the difficulties with isolating a control group for mass media campaigns, the study used a quasi-experimental design to compare those exposed to the campaign with those who were not. The researchers evaluated which parts of the campaign were effective and which were not using two primary data types. First, independent cross-sectional samples provided a broad understanding of whether people were receiving the message. Second, the study tracked a smaller sample over time to identify which people were changing their attitudes and behaviors. Conveniently, three Democratic and Health Surveys were conducted in Bolivia during the study. Those data tracked well with the study data, providing independent validation of the results.[c]

Country-Level Effects of a Student-Exchange Program

Using data on a country's participation in U.S.-hosted military educational exchanges and the number of university students studying in the United States between 1980 and 2006, Carol Atkinson studied the correlational effects between participation in exchange programs and country-level changes in the level of human rights abuse. She used a generalized multilevel longitudinal model and controlled for other country-level predictors of level of human rights abuse. She found support for the hypothesis that U.S.-hosted exchange programs can play a role in disseminating liberal values in authoritarian states.[d]

[a] Joyee S. Chatterjee, Anurudra Bhanot, Lauren B. Frank, Sheila T. Murphy, and Gerry Power, "The Importance of Interpersonal Discussion and Self-Efficacy in Knowledge, Attitude, and Practice Models," *International Journal of Communication*, Vol. 3, 2009.

[b] Michael J. Gilligan, Eric N. Mvukiyehe, and Cyrus Samii, *Reintegrating Rebels into Civilian Life: Quasi-Experimental Evidence from Burundi*, Washington, D.C.: United States Institute of Peace, 2010; Gaarder and Annan, 2013.

[c] For more on the study, see Thomas W. Valente and Walter P. Saba, "Campaign Recognition and Interpersonal Communication as Factors in Contraceptive Use in Bolivia," *Journal of Health Communication*, Vol. 6, No. 4, 2001; Thomas W. Valente and Walter P. Saba, "Mass Media and Interpersonal Influence in a Reproductive Health Communication Campaign in Bolivia," *Communication Research*, Vol. 25, No. 1, February 1998; and Thomas W. Valente and Walter P. Saba, "Reproductive Health Is in Your Hands: The National Media Campaign in Bolivia," *SIECUS Report*, Vol. 25, No. 2, December 1996–January 1997.

[d] Carol Atkinson, "Does Soft Power Matter? A Comparative Analysis of Student Exchange Programs 1980–2006," *Foreign Policy Analysis*, Vol. 6, No. 1, January 2010.

intimidated and repeated the same message that was made in one of the television commercials produced and disseminated on behalf of the IOTF.[94]

Framing analysis is also important in the formative or creative phase of the campaign to determine how the target audience perceives an issue and the opportunities for reframing it.[95] The frame analysis process can use focus groups, surveys, content analysis and interviews.[96]

Case Studies

A case study is an "in-depth description of the activities, processes and events that happened during a program" and can be used both to "inform the program design and to evaluate it."[97] Case studies are conducted via data from behavioral observation, key informant interviews, and literature and document reviews. Case studies are appropriate in four conditions: (1) the program is unique and unrelated to other activities; (2) the program is complicated and other data collection is unfeasible or unwieldy (such as when there are more variables than data points); (3) the program addresses a small or unique population; and (4) the program lacks measurable goals or objectives in the near term.[98] Coffman views case studies as valuable when the researchers are seeking an in-depth understanding of why a particular communication intervention succeeded or failed.[99] Case studies can help explain the factors behind effectiveness, or lack thereof.

Case studies are not conveniently selected anecdotes. Good case studies are thorough and objective, adhering to rigorous research standards. Various criteria can be used for selecting the sample of cases to study. Researchers may use the "success" criteria and trace back success factors or select a failure case and compare with the successful case.[100] In case study evaluations, data are generated from interviews, observations, documentaries, impressions and statements of others about the case, and contextual information.[101] For a thorough treatment of these methods, see *Case Study Research: Design and Methods* by Robert Yin.[102] Qualitative research methods used in the case

[94] Author interview with Paul Bell, May 15, 2013.

[95] Coffman, 2002.

[96] Marielle Bohan-Baker, "Pitching Policy Change," *Evaluation Exchange*, Vol. 7, No. 1, Winter 2001, pp. 3–4.

[97] Valente, 2002, p. 68.

[98] Valente, 2002, p. 68.

[99] Author interview with Julia Coffman, May 7, 2013.

[100] Author interview with Julia Coffman, May 7, 2013.

[101] Michael Quinn Patton, *Qualitative Research and Evaluation Methods*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2002, p. 449.

[102] Robert K. Yin, *Case Study Research: Design and Methods*, 5th ed., Thousand Oaks, Calif.: Sage Publications, 2014.

study research process, such as interviews and expert elicitations, are discussed in Chapter Eight.

Single case studies cannot make causal inference, but combined case studies have a greater capacity to test causal hypotheses, depending on the sample of cases. Cull suggests that DoD IIP programmers and evaluators assemble case study documentation in the form of a wiki casebook to document lessons learned—what worked, what did not, and why those conclusions were drawn.[103] A similar casebook edited by Cull and Ali Fisher, *The Playbook*, was constructed to document public diplomacy successes, failures, and lessons learned.[104]

The Best Evaluations Draw from a Compendium of Studies with Multiple Designs and Approaches

Each design described previously has strengths and weaknesses that vary by environment and circumstance. No single design will be appropriate for all campaigns. And, independent of feasibility, no single design will present a full picture of effectiveness. Thus, the most valid conclusions about program effects are those that are based on results from multiple studies using different designs. From a methodological perspective, this is known as "convergent validity." The Sesame Workshop, for example, advocates for a "compendium of studies," including a mix of qualitative, experimental, and quasi-experimental designs that look at naturalistic versus contributed conditions. As Cole explained, "no single design will tell the full picture, so the key is to have as many studies as possible and build a story when methods converge across multiple studies."[105]

These sentiments were echoed by Moehler, who argues that experimental and quasi-experimental designs do not work for all kinds of questions, and are not appropriate in all circumstances. Even if they are feasible, using the same approaches over and over leads only to a partial answer, which can be a mistaken answer, "so the best way to do research is to approach it from multiple angles—surveys, some experimental work, in-depth interviews, and observational work."[106]

Steve Booth-Butterfield makes that case that triangulation is particularly important in IIP evaluation due to the challenges with data availability and quality.[107] Because there are limitations to each approach, IIP evaluators should look at all evidence from as many different angles that are reasonable, rational, empirical, and feasible, and see whether the evidence is trending in the same direction. While it is relatively easy to identify weaknesses with any single measure, when a collection of measures across dif-

---

[103] Author interview with Nicholas Cull, February 19, 2013.

[104] Nicholas J. Cull and Ali Fisher, *The Playbook: Case Studies of Engagement*, online database, undated.

[105] Author interview with Charlotte Cole, May 29, 2013.

[106] Author interview with Devra Moehler, May 31, 2013.

[107] Author interview with Steve Booth-Butterfield, January 7, 2013.

ferent methods is suggesting the same general trend, you can have much more confidence in your conclusions. He explains that "good evaluations are good because they are complex, rational arguments with several moving parts (clearly defined and organized) with lots of evidence spanning different data types (that qualitative to quantitative range, for example)."[108] However, because of the diversity in perspectives and approaches, effectively implementing this approach requires that one person or group who is familiar with and has the power to affect the whole assessment process be responsible for triangulating disparate approaches. For more on this recommendation, see Chapter Ten.

The Importance of Baseline Data to Summative Evaluations

Given the assessment principle that evaluating change requires a baseline, evaluations, to the extent feasible, should incorporate baseline data that were collected prior to the intervention. If the intervention takes place over a long period of time, data should be collected at midline and other points throughout the campaign. This underscores the importance of building in evaluation design and measures from the beginning of campaign planning: If you just bring in assessors at the end, "you can't expect them to produce meaningful results."[109] In her study of the Japan Exchange and Teaching Programme, Emily Metzgar noted that the program's difficulty in demonstrating impact was principally due to a lack of baseline data.[110] However, baseline data are not always available or feasible to collect. In the absence of baseline data, a baseline should be constructed post facto with techniques like propensity matching.

Baselines can be constructed from surveys or focus group data on the population's familiarity or attitudes toward the issue the intervention is targeting.[111] *The sample frame at baseline and end line should be the same, otherwise it cannot be determined whether observed changes are due to the intervention or changing characteristics of the sample.* Baseline data should be collected immediately before the launch of the program; data collected from the formative process is typically too old by the time the program launches.[112]

Baseline or proxy baseline data should not only capture outcomes; they should also characterize the prior state, system constraints, and intervention inputs that define the system within which the intervention is operating. These system factors should be specified in the logic model and can include measurements of the people, their atti-

---

[108] Author interview with Steve Booth-Butterfield, January 7, 2013.

[109] Author interview with Ronald Rice, May 9, 2013.

[110] Emily T. Metzgar, *Promoting Japan: One JET at a Time*, CPD Perspectives on Public Diplomacy No. 3, Los Angeles, Calif.: University of Southern California Center on Public Diplomacy, 2012.

[111] Author interview with Amelia Arsenault, February 14, 2013; interview with Kavita Abraham Dowsing, May 23, 2013.

[112] Author interview with Charlotte Cole, May 29, 2013.

tudes, the security and economic environment, and institutional and political factors. A key aspect in evaluating a complex campaign is the need to consider, measure, and assess the effect of the major variables that help explain why certain outputs occurred and others did not. System variables that may change over time should be measured with sufficient frequency to capture those changes.[113]

To build flexibility into the assessment process, military IIP program evaluators emphasized that baseline data should be sufficiently broad to capture information that is likely to be of use across changing objectives or commanders. If baseline data collection is tailored to a set of objectives at a particular point in time, the evaluators will have to establish new baselines and start over from the beginning every time objectives and priorities shift.[114]

Summary

This chapter reviewed the three types of IIP evaluations and key concepts governing evaluation design, with a focus on the summative evaluation phase. Key takeaways include the following:

- The best designs are valid, generalizable, practical, and useful. However, there are tensions and trade-offs inherent in pursuing each of those objectives. Evaluators should select the strongest evaluation design from a methodological perspective among those designs that are feasible with a reasonable level of effort and resources.
- Rigor and resources are the two conflicting forces in designing assessment. The rigor and resources associated with an assessment should be proportionate to the potential importance of the results. There should be an allowance for "good enough" assessments.
- Assessment design, processes, and level of rigor and formality should be tailored to the assessment end users and stakeholders. Academic rigor must be balanced with stakeholder needs, appetite for research, and cost considerations.
- Threats to internal validity are controlled by design. Broadly, designs can be classified as experimental (random assignment with a control group), quasi-experimental (comparison group without random assignment), or nonexperimental (no comparison group). The more controlled the design, the higher the internal validity. Thus, the relative value of experimental research depends on the importance of making causal inference.

---

[113] Ronald E. Rice and Dennis R. Foote, "A Systems-Based Evaluation Planning Model for Health Communication Campaigns in Developing Countries," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2013.

[114] Author interview on a not-for-attribution basis, July 31, 2013.

- Quasi-experimental designs are the next best option if an experimental design is not feasible, but they contribute weaker causal estimates due to the challenges with controlling for rival explanations. The most popular quasi-experimental design for IIP evaluation is the nonequivalent group design, in which differences between those who were exposed to the program are compared with those who were not exposed. This design suffers from selection bias (those who were exposed may be predisposed to behavioral outcomes of interest) but is typically the most cost-effective and affordable.

Formative and Qualitative Research Methods for IIP Efforts

This chapter explores qualitative research methods along with other methods that can be used in the formative evaluation phase. While formative and qualitative research often overlap, they are by no means completely equivalent. Formative evaluations can use quantitative methods, and qualitative methods can inform evaluations conducted in each of the three phases.

Formative evaluation consists of the research conducted in the preintervention stage to analyze audience and network characteristics, determine program needs and baseline values, identify campaign strategies, develop and test messages and messengers, and identify the variables that can promote or obstruct the campaign. It is used to specify the logic model and the characteristics of the information environment that the intervention is designed to influence, including barriers to behavioral change.[1] Formative research methods are varied. Classical methods employed include focus groups and in-depth interviews. Increasingly, researchers are relying more on quantitative approaches, such as content analysis and laboratory experiments, to test the cognitive effects of messages and products. Less traditional qualitative methods encountered in our research include community assessments, photojournalism, and temperature maps.[2]

Importance and Role of Formative Research

Several of the SMEs interviewed stressed the importance of formative research and argued that it is systemically undervalued, especially in periods of budgetary cutbacks. An up-front investment in formative research typically saves costs in the long run because it increases the likelihood that the program will be effective, reduces costs associated with program implementation, and minimizes expenses during both the

---

[1] Author interview with Ronald Rice, May 9, 2013; interview with Thomas Valente, June 18, 2013; Coffman, 2002, p. 13.

[2] Author interview with Kavita Abraham Dowsing, May 23, 2013.

process and summative evaluation phases.[3] By demonstrating the likely effects of the effort on targeted audiences, formative research allows researchers to have greater confidence in their conclusions about expected effects of the effort. If an effort has been validated as having a certain effect, campaign effectiveness will then depend principally on the extent of exposure.[4] Likewise, if summative research shows a lack of outcomes, evaluators can more easily isolate the source of program failure if they conducted sound formative research.

DoD should enhance its investment in and focus on the formative evaluation phase. Too often, even "basic" formative research and pretesting "just doesn't happen."[5] In discussing the value that formative research has brought to Sesame Workshop programming, Charlotte Cole urged managers to resist the temptation to cut formative research when budgets are tight.[6] In Simon Haselock's experience, campaign failure can often be traced back to underinvesting in the "understanding phase" (formative research) due to time constraints and other pressures.[7]

Formative research has the additional advantage of helping to demonstrate the value of research to program managers and sponsors.[8] While summative evaluations are often seen as threatening, formative research improves program outcomes and simplifies the planning process, providing tangible and near-term benefits to program managers. Preintervention research also provides an opportunity to collect baseline measures for summative evaluations. However, the data should be collected immediately prior to the launch of the program. Often, data from formative evaluations are too old to serve as an optimal baseline.[9]

Characterizing the Information Environment: Key Audiences and Program Needs

The first component of formative research is to determine the characteristics of the target audience and information environment that shapes the audience's views and behaviors. The first step in the *Joint Information Operation Assessment Framework*, for example, is to characterize the IE, including the "cognitive, informational, and physi-

---

[3]   Author interview with Thomas Valente, June 18, 2013; interview with Charlotte Cole, May 29, 2013.

[4]   Author interview with Mark Helmke, May 6, 2013.

[5]   Author interview on a not-for-attribution basis, January 23, 2013.

[6]   Author interview with Charlotte Cole, May 29, 2013.

[7]   Author interview with Simon Haselock, June 2013.

[8]   Author interview with Julia Coffman, May 7, 2013.

[9]   Author interview with Charlotte Cole, May 29, 2013.

cal domains" to inform campaign planning.[10] "Understand the operational environment" is a key imperative of operational design, and it is a predicate for mission analysis in JOPP, according to JP 5-0. Other traditions may refer to this process as the "needs assessment" or as measuring the "system of influence" that the intervention is operating within. This section explores two key, interrelated analytic tasks associated with this phase: audience segmentation and network analysis.

Audience and network analysis in the formative phase includes several techniques that help researchers and program designers understand and categorize key audiences, including the way they engage with and are influenced by media and each other. This process should help planners understand what media and formats resonate with what audiences, including the target audience and those that influence them. Tony Foleno, senior vice president of research and evaluation at the Ad Council, explains that formative research "helps to ensure the message is tailored to the audience it is supposed to affect and not just the advertisers who developed it.[11] Rebecca Collins, a psychologist at the RAND Corporation who specializes in the determinants and consequences of health risk behavior, encourages program managers to identify and understand the types of media and the key influencers that the target audience gets information from in order to improve the effectiveness and efficiency of message delivery.[12]

Audience Segmentation

Audiences are not a homogeneous group. Recognizing the tremendous diversity in terms of psychographic variables within any given population, planners use audience segmentation techniques to understand how different messages resonate with different segments of the population.[13] IIP interventions should differentiate populations into segments of people who share "needs, wants, lifestyles, behaviors and values" that make them likely to respond similarly to an intervention.[14]

Audience segmentation should shift its focus from demographic differences to psychographic differences (e.g., differences in value priorities). When it comes to message receptiveness, demographic segmentation often poorly reflects diversity within a population. Better approaches segment the audience along psychographic variables and their demographic correlations rather than on just demographic variables alone.[15] Instead of assuming that people of a similar race, gender, or age share similar values,

---

[10] Joint Information Operations Warfare Center, *Joint Information Operations Assessment Framework*, October 1, 2012, pp. 11–12.

[11] Author interview with Tony Foleno, March 1, 2013.

[12] Author interview with Rebecca Collins, March 14, 2013.

[13] Author interview with Gerry Power, April 10, 2013.

[14] Sonya Grier and Carol A. Bryant, "Social Marketing in Public Health," *Annual Review of Public Health*, Vol. 26, 2005, p. 322.

[15] Author interview with Gerry Power, April 10, 2013.

planners should segment the audience according to what is important to them and subsequently determine whether those values correspond to demographic categories.

In *Youth in Iran*, Klara Debeljak used survey data to identify four different psychographic segments of young people: nontraditionalist, mainstream, conservative, and ultraconservative. She found differences in receptiveness to various types of messaging and media formats between these categories that were not significant when looking at demographic differences.[16]

Sonya Grier and Carol Bryant echo this point in the health communication sector, arguing that audience segmentation in public health "is limited by an over-reliance on ethnicity and other demographic variables." In their view, these programs would benefit from a more customized segmentation approach akin to those employed by social marketers. The authors encourage IIP interventions to segment along such variables as lifestyle, personality characteristics, values, life stage, future intentions, readiness to change, product loyalty, propensity for sensation seeking, and interest in changing lifestyles.[17]

Audiences can also be segmented by network characteristics, a technique known as *sociometric segmentation*. Network analysis can optimize a campaign's engagement strategy by identifying key influencers or opinion leaders within a community, as well as those most amenable to the message.[18]

For awareness campaigns, some social marketing experts suggest that audiences should be segmented by self-rated prior knowledge. Andrea Stanaland and Linda Golden have observed that people with higher self-rated knowledge are not receptive to messages, presumably because they do not feel a need for additional information. In this sense, self-rated knowledge may diminish the motivation to process new information, adversely affecting message receptivity.[19]

Social Network Analysis

Social networks mediate the diffusion of information and behavioral change processes. Network analysis, also called "social network analysis," provides quantitative and visual representation of the relationships and information channels among individuals, groups, and organizations within a given target population.[20] By revealing and mea-

---

[16] Klara Debeljak, *Youth in Iran: A Story Half Told: Values, Priorities and Perspectives of Iranian Youth*, Young Publics Research Paper Series No. 1, Washington, D.C.: InterMedia, May 2013.

[17] Grier and Bryant, 2005, p. 332.

[18] Author interview with Thomas Valente, June 18, 2013.

[19] Andrea J. S. Stanaland and Linda L. Golden, "Consumer Receptivity to Social Marketing Information: The Role of Self-Rated Knowledge and Knowledge Accuracy," *Academy of Marketing Studies Journal*, Vol. 13, No. 2, 2009, p. 32.

[20] Maureen Taylor, "Methods of Evaluating Media Interventions in Conflict Countries," paper prepared for the workshop "Evaluating Media's Impact in Conflict Countries," Caux, Switzerland, December 13–17, 2010, p. 1.

suring the characteristics of the audiences' social and information networks, employing network analysis during the formative phase can greatly improve the efficacy and efficiency of the campaign. However, network analysis is underutilized by DoD IO entities and by other government strategic communication entities, largely due to a lack of understanding and familiarity. More should be done to develop and apply these techniques to DoD influence operations.[21]

Network analysis can improve campaign strategy and targeting by identifying key influencers and opinion leaders. Opinion leaders typically have greater exposure to the messages and are more likely to exercise informal influence over the attitudes and behaviors of those in their social networks. Ronald Rice differentiated between a direct effects strategy and a second effects strategy. A direct effect campaign disseminates messages to the target audience; a second effects strategy initiates an indirect or multistep flow by disseminating messages to interpersonal influencers who are positioned to shape the behavior of the target audience. He posited that a second effects strategy could be particularly valuable in counterinsurgency environments. For example, mothers or religious leaders may be particularly well positioned to dissuade a young man from engaging in risky behavior, like implanting improvised explosive devices.[22]

Community-level determinants of behavioral change and confounding or system-level variables that interact with the IIP intervention can also be estimated through network analysis techniques. As discussed in Chapter Five, network analysis techniques can measure innovation thresholds, which define the number of people who need to sign on to something before the individual or community will adopt the change. Innovation thresholds can have significant implications for the design of the campaign. If the focal audience has a high threshold, the campaign may need to be implemented on a community-by-community basis. Thomas Valente points out that "influencers" are not necessarily themselves innovative or low threshold and often have high thresholds to innovation. This can be a source of tension in deciding whom to engage.[23] Network analysis can also be used to measure social capital and other constructs, such as trust in the government or in adversary institutions.[24]

In addition to informing the design of the campaign, network analysis can inform the research process and sample selection strategy. Haselock suggests that IIP planners "take a cue from the intelligence community and journalists" and use network analysis to identify reliable and valuable sources of information and input during the formative phase.[25] Network analysis can also be used in the summative phase to track progress

---

[21] Author interview with James Pamment, May 24, 2013.

[22] Author interview with Ronald Rice, May 9, 2013.

[23] Author interview with Thomas Valente, June 18, 2013.

[24] Author interview with Craig Hayden, June 21, 2013.

[25] Author interview with Simon Haselock, June 2013.

over time. Valente's work provides a thorough treatment of the use of network analysis for communication campaign evaluation.[26]

Audience Issues Unique to the Defense Sector: Target Audience Analysis

The PSYOP community refers to audience analysis as target audience analysis, or TAA, and this is the second of the seven phases in the PSYOP (now MISO) process.[27] According to several SMEs from the defense sector, effective TAA is the "cornerstone" of effective influence because it uncovers "root causes" and identifies the most-effective "levers to pull."[28] The basics of the TAA process are laid out in doctrine.[29] This section discusses some of the issues unique to TAA that were analyzed throughout our research. However, several other methods and tools discussed throughout this report overlap with and can contribute to TAA, such as content analysis and atmospherics (see Chapter Nine) and the qualitative methods discussed at the end of this chapter (e.g., focus groups, interviews, expert elicitation).

The information environment evolves rapidly. To effectively inform campaign planning, TAA should therefore be conceived of as a living process rather than as a static picture of the information environment. TAA should use updates on conversations and sentiments in the target audience to modify products and messages right up to the dissemination moment and as they move throughout the stages of the campaign execution or product cycle.[30]

DoD needs to improve its processes and capabilities for audience analysis and understanding the media environment. Craig Hayden argues that a better understanding of how media circulation amplifies the effects of an event would allow planners to better anticipate effects. This involves "qualitative analysis coupled with quantitative sentiment analysis. You can't just rely on cultural anthropologists, but you need them along with the large-N analysts."[31] A defense-sector SME noted that DoD TAA needs

---

[26] See Thomas W. Valente, *Social Networks and Health: Models, Methods, and Applications*, New York: Oxford University Press, 2010; Thomas W. Valente, *Network Models of the Diffusion of Innovations*, New York: Hampton Press, 1995; Thomas W. Valente, "Network Interventions," *Science*, Vol. 337, No. 6090, July 2012.

[27] See Headquarters, U.S. Department of the Army, 2005, pp. 12–13. Some communication experts, including Valente, argue that DoD should consider moving away from *target* to describe an audience because the term is perceived poorly by the population, particularly in a military context. On the other hand, incorporating audience analysis into the standard DoD targeting process would help integrate IIP activities with other military operations and processes.

[28] Author interview on a not-for-attribution basis, January 23, 2013.

[29] See, for example, chapter 5 in Headquarters, U.S. Department of the Army, and Headquarters, U.S. Marine Corps, *Psychological Operations, Tactics, Techniques, and Procedures*, Field Manual 3-05.301/Marine Corps Reference Publication 3-40.6A, Washington, D.C., December 2003.

[30] Author interview on a not-for-attribution basis, July 30, 2013.

[31] Author interview with Craig Hayden, June 21, 2013. Also see Craig Hayden, *The Rhetoric of Soft Power: Public Diplomacy in Global Contexts*, Landham, Md.: Lexington Books, 2012.

to incorporate more automated tools.[32] For a discussion of automated sentiment analysis, see the section "Content Analysis and Social Media Monitoring" in Chapter Nine.

TAA capabilities are constrained by personnel limitations. An important difference between DoD TAA and audience analysis in other sectors is that TAA is usually conducted by junior enlisted officers with limited formal training. One PSYOP officer described the challenge as follows: "Other organizations that do psychological profiling use personnel with Ph.D.'s or master's [degrees]; we use 20-year-olds. . . . They aren't really up to it."[33]

DoD IIP doctrine could improve TAA by clarifying the tasks and responsibilities associated with defining the information environment. LTC Scott Nelson, who formerly served as the chief of influence assessment at USNORTHCOM, contends that defining the information environment—the first step in the joint IO assessment process—is not achievable because the question is too broad and there are insufficient resources to meaningfully answer it. Defining the information environment requires "the entire intelligence community," but it is "often unavailable to help" even in limited capacities.[34] Doctrine, he suggests, could address this shortcoming by clarifying the questions IO planners need to address and by identifying the intelligence community components or existing data sources that can be leveraged.[35]

DoD IIP planners should consider leveraging the Intended Outcomes Needs Assessment (IONA) methodology and tool developed by the United States Institute for Peace to assist in characterizing the information environment. The tool was built to help planners in the international development community craft media interventions that can address the information-related causes of conflict in a society. It consists of a three-stage interview-based process for collecting and analyzing data on the media, the conflict, and relationship between the two. The framework document and the Excel-based data collection tool (Frame Manager) can be downloaded from the United States Institute for Peace website.[36]

Developing and Testing the Message

After characterizing the information environment, the next major task of formative research is to inform the development of the message or product. To develop effective

---

[32] Author interview on a not-for-attribution basis, July 30, 2013.

[33] Author interview on a not-for-attribution basis, January 23, 2013.

[34] Author interview with LTC Scott Nelson, October 10, 2013.

[35] Author interview with LTC Scott Nelson, October 10, 2013.

[36] Andrew Robertson, Eran Fraenkel, Emrys Schoemaker, and Sheldon Himelfarb, *Media in Fragile Environments: The USIP Intended-Outcomes Needs Assessment Methodology*, Washington, D.C.: United States Institute of Peace, April 2011. The Frame Manager Tool is available for download with the report.

messages, planners and researchers should solicit input from cultural anthropologists, ethnographers, trained participant observers, and trusted local sources who understand the dynamics on the ground. Where possible, voices from the protagonist and antagonist sides should be included.[37]

Network analysis and other techniques used by journalists or intelligence analysts should be leveraged to identify and validate key sources who can inform the research and development process.[38] Joshua Gryniewicz, communication director at Cure Violence, said that his organization relies on neutral groups when adapting its model to local conditions. Neutral groups are not affiliated with a particular militia group or sect and are perceived as credible by all sides in a conflict.[39]

Formative research is typically done in-house. The Sesame model, which has been exported as a best practice, brings together the creative side with the educational specialists and the researchers. Instead of bringing in outside consultants, the teams work together in an iterative process over the course of the entire project. Cole believes that the formative phase needs to be in-house because it is important that the researchers are integrated with the programmers. She cautioned against outsourcing formative research, because it is complex, requires substantive expertise, and must be embedded with the creative process.[40]

Rigorously pretesting messages on representatives of the intended audiences will dramatically improve the likelihood that the message is effective and will mitigate the chance of failure or unintended consequences. One example is Valente's illustration of a message designed to make tobacco use look "uncool" to teens that could easily backfire if it is perceived as manipulation from adults. Likewise, government strategic communication messages must walk a fine line between promoting U.S. interests and being perceived as culturally insensitive. Testing the message in the formative phase is the best way to calibrate the messaging so that it achieves an effect without offending the audience. Unfortunately, this process is often shortchanged by planners,[41] and DoD efforts are no exception.

Piloting the intervention on a small scale can help refine the logic model and preemptively identify sources of program failure. Pilots give researchers more control of the measurements, enabling them to fine-tune the campaign. A successful pilot on a local or regional level is necessary, but not sufficient, evidence that the campaign

---

[37] Author interview with Simon Haselock, June 2013.

[38] Author interview with Simon Haselock, June 2013.

[39] Author interview with Joshua Gryniewicz, August 23, 2013.

[40] Author interview with Charlotte Cole, May 29, 2013.

[41] Author interview with Thomas Valente, June 18, 2013; author interview on a not-for-attribution basis, January 23, 2013.

will be effective on a national level.[42] Despite the rich information provided by pilot programs, planners must keep in mind the different conditions for success at different scales. Kate Fehlenberg emphasizes this point in a paper presented at the American Evaluation Association annual conference, observing that many campaigns are "not designed, monitored or evaluated for . . . performance at scale."[43]

In his *Resource Guide to Public Diplomacy Evaluation*, Robert Banks suggests using competition in the field-testing phase to promote performance-oriented, efficient campaigns. Competing teams of programmers could be tasked with designing an initiative to address a particular issue and could field-test the design in two different countries or regions with similar baseline information environments. Changes in outcomes (e.g., sentiment) could be observed, and the best-performing design, subject to resource constraints, would be selected.[44]

Computer-generated simulations and exercises process "what-if" scenarios by constructing hypotheticals from existing conditions. These simulations can help refine the logic model by identifying sources of failure and validating or invalidating the relevance of various assumptions and causal ties. They are also used to estimate intervention timelines and expected outcomes.[45]

Split or A/B testing (described briefly in Chapter Seven) can be an effective product testing technique in the formative phase if researchers have narrowed the range of potential messages and are interested in the relative effectiveness of a message or associated features. The technique involves employing two variants of a message to two groups within the same audience segment and measuring differences in responses. The treatment variant of the message should differ only in one respect from the control variant.[46]

Importance and Role of Qualitative Research Methods

Given the inevitable challenges associated with collecting valid and reliable quantitative data on IIP effects, evaluators should consider the balance between qualitative and quantitative information at all stages of evaluation. The best quantitative methods are those that generate information that converges with the information produced from qualitative methods, and vice versa. Maureen Taylor recommends, at a minimum,

---

[42] Author interview with Pamela Jull, August 2, 2013.

[43] Kate Fehlenberg, "Critical Juncture: Applying Assessment Tools and Approaches to Scaling-Up: A New Focus for External Validity," paper presented at Evaluation 2013, the annual conference of the American Evaluation Association, Washington, D.C., October 14–19, 2013.

[44] Banks, 2011.

[45] Author interview with Thomas Valente, June 18, 2013.

[46] Ries, 2011.

always pairing a quantitative method with a qualitative method—for example, conducting a survey and a focus group, a survey and in-depth interviews, or content analysis and a focus group.[47]

Military analysts often prefer quantitative data, not because such data are inherently more objective but because they are easier to analyze and they provide, in Jonathan Schroden's words, a "façade of rigor."[48] But, *numeric data* are not the same as *objective data*. Quantitative data are only as valid and reliable as the instruments and processes that generated them, and analysts "should not lose sight of the very qualitative nature of survey questions and administration."[49] Moreover, quantitative data are often *less* useful than qualitative data, because they encourage data customers to view results as countable phenomena, which, in an IIP setting, are more likely to be associated with outputs than with meaningful outcomes.[50] In other words, a numbers-based assessment makes little sense "in the absence of a credible numbers-based theory."[51]

For example, a major limitation to some DoD assessment frameworks is that they discredit the utility and role of qualitative data. In Scott Nelson's view, there is an "ORSA mentality" to "only measure things you can count" that drives these approaches. Because almost all of the data collected in the information environment are qualitative in nature, this mentality is particularly impractical and counterproductive for IIP assessment. While there are challenges with qualitative data, he argued, they should be addressed through social science validation techniques and mixed-method approaches rather than an exclusive focus on quantitative data.[52]

Qualitative methods also help interpret or explain quantitative data, especially unexpected or surprising results. Even where valid and reliable quantitative data are available, qualitative methods such as focus groups and in-depth interviews are needed, because they are better for determining causality and uncovering motivations or the drivers of change.[53] It is often said that quantitative data tell you *what* and qualitative data tell you *why*. Qualitative methods also help develop and improve the survey instruments and content analysis tools by generating hypotheses that can be tested and by identifying the words and phrases used by the target audience to frame issues. Valente describes the process as "iterative" in that qualitative methods help develop, explain, and then refine the quantitative methods.[54]

---

[47] Author interview with Maureen Taylor, April 4, 2013.

[48] Schroden, 2011, p. 99.

[49] Author interview with Jonathan Schroden, November 12, 2013.

[50] Author interview with Simon Haselock, June 2013.

[51] Downes-Martin, 2011.

[52] Author interview with LTC Scott Nelson, October 10, 2013.

[53] Author interview with Matthew Warshaw, February 25, 2013.

[54] Author interview with Thomas Valente, June 18, 2013.

Qualitative data should be generated by rigorous social science methods. As one expert joked, "the plural of *anecdote* is not *data*."[55] Moreover, while qualitative methods add value to quantitative approaches, programmers should avoid making decisions on the basis of a single qualitative method.[56] The following sections discuss the application of several qualitative methods to IIP evaluation, including focus groups, in-depth and intercept interviews, expert elicitation, and narrative analysis.

Focus Groups

Focus groups are "carefully planned discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment."[57] They can "reveal underlying cognitive or ideological premises" that are "brought to bear on interpretation."[58] In the formative phase, focus groups are employed to test products, develop hypotheses, and refine the logic model through the identification of causal mechanisms. In Amelia Arsenault's experience, focus groups are best for identifying "unexpected" causal mechanisms: "You'll hear things you would never think about in your wildest dreams."[59] They are employed in the summative phase to develop and refine survey instruments and to validate data produced by other research methods, and to interpret and explain results—e.g., why the program succeeded or failed. Focus groups are advantageous because they are relatively cost-effective, flexible, and socially oriented and have high levels of face validity.[60]

Focus groups are particularly valuable for testing products and anticipating how the audience will react to various dimensions of a product—message, imagery, language, music, and so forth. Matthew Warshaw recalled a few cases in which planned IO programs were canceled because focus groups showed that the message was "culturally insensitive or that the psychological objective we were seeking was flawed." In one example, a product was designed to make Afghans feel ashamed about their behavior. A focus group uncovered that IO products are particularly bad at instilling a sense of shame and that Afghans may react counterproductively to such attempts.[61]

There are several challenges to implementing focus groups in DoD operating environments. First, they can be difficult to organize and require skilled local facili-

---

[55] Author interview on a not-for-attribution basis, December 15, 2013.

[56] Author interview with Kim Andrew Elliot, February 25, 2013.

[57] Richard A. Kreuger and Mary Ann Casey, *Focus Groups: A Practical Guide for Applied Research*, Thousand Oaks, Calif.: Sage Publications, 1994, p. 18.

[58] Peter Lunt and Sonia Livingstone, "Rethinking the Focus Group in Media and Communications Research," *Journal of Communication*, Vol. 46, No. 2, June 1996, p. 96.

[59] Author interview with Amelia Arsenault, February 14, 2013.

[60] Author interview with Thomas Valente, June 18, 2013; author interview with Rebecca Collins, March 14, 2013.

[61] Author interview with Matthew Warshaw, February 25, 2013.

tators who share demographic characteristics with the focus group sample. Second, responses can be biased due to groupthink and normative pressures of conformity. In Afghanistan, Warshaw found that people tended to agree with each other and would encourage the group to come to consensus. Subjects are very "self-aware" and concerned about the repercussions associated with voicing minority opinions.[62] Finally, outcomes can be unpredictable and results are difficult to standardize ad analyze.[63]

To manage these challenges, SMEs discussed several techniques, best practices, and insights for conducting focus groups to inform and evaluate IIP interventions in conflict environments.

- *Format.* Focus groups typically last one to two hours and are held with four to 12 homogeneous participants. The groups should be moderated by a trained expert, recorded if culturally appropriate, and structured according to an agreed-upon interview guide.[64]
- *Focus group composition.* Groups should be separated by gender, age, education level, and, where relevant or appropriate, ethnicity, religion, or sect. Otherwise, participants may defer to elders or males, engage in groupthink, or feel uncomfortable speaking openly. Facilitator demographics should match those of the group. Because a given focus group is composed of a particular demographic cross section, it has low external validity and cannot be generalized to the population at large. Researchers should hold several focus groups to capture perceptions across different groups.[65] Groups work best when they are composed of strangers.[66]
- *Generating the focus group sample.* The focus group sample frame depends on the target audience. Often, it is better to hold the focus groups with key influencers and representatives of mediating institutions than with representatives of the average citizen.[67] Haselock recommends using network analysis to identify individuals to bring into the focus groups.[68] Arsenault has found that snowball sampling is typically the most feasible option in conflict environments.[69] Local NGO partners can also be valuable in finding focus group participants.[70] If the researchers

---

[62] Author interview with Matthew Warshaw, February 25, 2013.

[63] Author interview with Thomas Valente, June 18, 2013.

[64] Author interview with Thomas Valente, June 18, 2013.

[65] Taylor, 2010, p. 7; interview with Amelia Arsenault, February 14, 2013.

[66] Patton, 2002.

[67] Author interview with Mark Helmke, May 6, 2013.

[68] Author interview with Simon Haselock, June 2013.

[69] Author interview with Amelia Arsenault, February 14, 2013.

[70] Taylor, 2010, p. 7.

intend to use a focus group in the summative phase, they should consider using the same sample for the formative and summative focus groups.[71]

- *Payment.* Paying subjects helps to incentivize participation but can skew results. Payment should be sufficient to compensate participants for their time but should not be perceived as a significant source of income.
- *Facilitators.* Good facilitators are essential to effective focus groups. The facilitator must speak the language and match the demographics of the group. Effective facilitators must have strong social skills, so that they can prevent groupthink and can diminish the influence of a stronger, dominant person.[72] Taylor recommends using two people to facilitate a group, wherein the moderator is assisted by someone who can manage the "people" part of the focus group.[73]
- *Building local capacity to conduct focus groups.* Local research capacity is particularly important for high-quality focus groups, because they depend on good, local facilitators who can be trusted without supervision. Even the presence of an American in the room can "skew the conversation."[74]
- *Open-ended questions.* SMEs had varying opinions on the value of open-ended questions. Anthony Pratkanis suggests erring toward open-ended questions (e.g., "Tell me what you think when you hear or see x construct"), because participants will "just say yes" if the questions are too targeted.[75] On the other hand, the questions need to be specific enough that responses provide relevant information.[76]
- *Recording.* Recording the focus group is ideal, but doing so can skew results or limit the sample in conflict environments, because potential participants may fear being recorded. If participants are hesitant to give permission to be recorded, assign at least two people to take notes.[77]
- *Triangulating.* Focus group answers should be triangulated with data generated by a different research method, such as a survey or content analysis.[78]
- *Analyzing and coding focus group content.* For some questions, you can code answers according to a scale (e.g., somewhat hostile, very hostile, somewhat familiar, very familiar). Intercoder reliability is very important.[79]

---

[71] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[72] Author interview with Maureen Taylor, April 4, 2013.

[73] Taylor, 2010, p. 7.

[74] Author interview with Amelia Arsenault, February 14, 2013.

[75] Author interview with Anthony Pratkanis, March 26, 2013.

[76] Taylor, 2010, p. 7.

[77] Taylor, 2010, p. 7; interview with Amelia Arsenault, February 14, 2013.

[78] Taylor, 2010, p. 7.

[79] Author interview with Amelia Arsenault, February 14, 2013.

Interviews

One-on-one interviews represent "one of the richest sources of information available to researchers."[80] As one SME noted, "Some of this we just can measure until we go and talk to the guy we are trying to influence."[81] Like focus groups, qualitative interviews can be used to test products, identify causal mechanisms, explain program failure, and validate and interpret survey results. Pratkanis and Warshaw believe that one-on-one interviews are better than focus groups for understanding causal mechanisms in conflict environments, because these interviews avoid the challenges associated with groupthink and pressures to conform to social norms.[82]

Qualitative interview methods include in-depth interviews and intercept interviews. In-depth interviews are semistructured interviews between researchers and members of the target audience. Conducting semistructured interviews is still widely recognized as an important qualitative data collection method and is commonly used in policy research, since it is applicable to a broad range of research questions.[83] The interviews should be open-ended, allowing the respondent to express opinions on tangential or unexpected topics, and can last from 30 minutes to two hours. Rapport between the interviewer and the respondent is very important. Interviewers should share characteristics with the subject and should begin the interview with uncontroversial subjects.[84] Interviewers should leverage intelligence-based networks to identify candidate interviewees and should randomly select respondents from the set of candidates.[85]

Intercept interviews, or "person on the street" interviews, are solicited at public places, such as a bazaar, and are useful for gauging public perceptions about a product or an issue. This technique is commonly used to assess the progress of MISO efforts. For example, Marine E-5s will go into a village and ask trusted sources or confidants about their attitudes, and their perceptions of the attitudes of others. This technique suffers from response and selection bias but in some cases is perceived to be the only available data collection method at the unit level.[86]

To get the most out of intercept interviews, researchers should pretest the instrument and vary the days, times, and interviewers.[87] While it is difficult to impose a

---

[80] Valente, 2002, p. 58.

[81] Author interview on a not-for-attribution basis, December 5, 2012.

[82] Author interview with Matthew Warshaw, February 25, 2013; interview with Anthony Pratkanis, March 26, 2013.

[83] Margaret C. Harrell and Melissa A. Bradley, *Data Collection Methods: Semi-Structured Interviews and Focus Groups*, Santa Monica, Calif.: RAND Corporation, TR-718-USG, 2009, p. 1.

[84] Valente, 2002, p. 58.

[85] Author interview with Simon Haselock, June 2013.

[86] Author interview on a not-for-attribution basis, December 15, 2013.

[87] Valente, 2002, p. 60.

formal sampling strategy, the sample of respondents should be as random as possible given the circumstances. For example, respondents can be selected based on walking patterns, where a subject is asked to participate after x number of steps in a certain direction, and so forth.[88] Where possible and tolerated, these interactions should be recorded, transcribed with text-recognition software, and coded.

The bellwether methodology is an emerging interview-based method used in advocacy evaluation to measure the extent to which a media or public communication campaign is influencing key decisionmakers. The method consists of highly structured interviews with high-profile policymakers or decisionmakers, half of whom were likely to have been exposed to the campaign and half of whom were unlikely to have been exposed. To ensure that recall questions are unprompted, the researchers are very vague about the subject of the interview prior to holding it. For example, for their project evaluating preschool advocacy, Coffman and colleagues told interviewees that they would interview them about education but not about early childhood. Coffman believes that a major advantage to the method is its cost-effectiveness due to not requiring large sample sizes. In the preschool advocacy example, they interviewed only 40 individuals.[89]

Narrative Inquiry

A narrative is a "system of stories that share themes, forms and archetype" and "relate to one another in a way that creates a unified whole that is greater than the sum of its parts." When these stories are widely known and consistently retold, these systems are considered *master narratives*.[90] *Narrative inquiry*, or narrative analysis, involves techniques for identifying these narratives to determine how members of the target audience create meaning in their lives through storytelling. It typically involves coding qualitative data collected through content analysis and qualitative methods (e.g., interviews and focus groups) using a standardized index. NATO's JALLC identifies narrative inquiry as a technique for evaluating public diplomacy based on the underlying theory that behavioral change can be assessed by analyzing "the stories people tell and how these stories shift over time."[91]

Cognitive Edge Inc. has developed the SenseMaker software package for narrative inquiry; the company claims that the software is able to identify which attitudes have the potential to be changed and which do not. The tool uses a large volume of

---

[88] Author interview with Amelia Arsenault, February 14, 2013.

[89] Author interview with Julia Coffman, May 7, 2013. For more on the bellwether method, see Coffman and Reed, 2009.

[90] Steven Corman, "Understanding Extremists' Use of Narrative to Influence Contest Populations," paper prepared for the Workshop on Mapping Ideas: Discovering and Information Landscape, San Diego State University, San Diego, Calif., June 29–30, 2011.

[91] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 42.

micronarratives collected voluntarily from subjects or participants in natural environments (e.g., "around the watercooler"), who then interpret, categorize, and tag their stories into abstract categories.[92] While this method produces less valid and generalizable results than a large, formal survey, it is less expensive and quicker, capable of providing real-time content directly from the target audience.[93]

Anecdotes

Anecdotes are widely used to communicate the effectiveness of IIP programs. Sometimes, anecdotes are used because a more rigorous measurement system was not in place. In other cases, measures are not perceived as necessary, because the effect is supposedly evident. Cull provided the example of Japan's response to the U.S. tsunami assistance, which demonstrated the effectiveness of the assistance in promoting the U.S. image abroad.[94]

But anecdotes are often used to demonstrate effect even when more-rigorous measures are available. For example, despite spending approximately $10 million per year on assessment, a leader of the IOTF offered two pieces of anecdotal evidence when asked why he knew "it worked." First, in a letter from Ayman al-Zawahiri to Abu Musab al-Zarqawi, then head of al Qaeda in Iraq, Zawahiri told Zarqawi that he had to "cool it" because atmospherics were becoming increasingly difficult for al Qaeda. Second, a *New York Times* article quoted an Iraqi colonel saying that he was going to vote because he was not intimidated, repeating the same rationale that was made in one of the IOTF television commercials.[95]

Anecdotes are not just easier to generate than experimental evidence; they are often more powerful. A study by Deborah Small, George Loewenstein, and Paul Slovic showed that people are more likely to donate to a cause if shown a picture of a victim than if presented with statistics demonstrating the extent of the problem. The researchers concluded that "people discount sympathy towards identifiable victims but fail to generate sympathy toward statistical victims."[96] Rice explained that research "is not part of our DNA. It's only been a phenomenon since the enlightenment."[97] Stories, rather, are how we make sense of the world.

---

[92]   To read more about SenseMaker software, see SenseMaker, homepage, undated. Also see NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 42.

[93]   NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 42.

[94]   Author interview with Mark Helmke, May 6, 2013.

[95]   Author interview on a not-for-attribution basis, May 15, 2013.

[96]   Deborah A. Small, George Loewenstein, and Paul Slovic, "Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims," *Organizational Behavior and Human Decision Processes*, Vol. 102, No. 2, March 2007.

[97]   Author interview with Ronald Rice, May 9, 2013.

Anecdotes alone are insufficient to empirically demonstrate impact, because there is no counterfactual condition to infer causality and no basis on which to generalize. However, it is good practice to embed stories or narratives into the presentation of the evaluation results to give meaning or color to the quantitative measures.[98] These stories can be elicited informally or from qualitative research methods like interviews and focus groups. Anecdotes are often "surreptitious," says Valente, but can "provide unexpected evidence that may be seen as more credible by policy-makers or outside agencies."[99] This is especially true if the decisionmaker can personally identify with the story or storyteller. For a discussion of structured case study designs, see Chapter Seven.

Expert Elicitation

While eliciting expert judgment is considered methodologically inferior to experimental designs, in many circumstances, structured expert elicitation is the most rigorous method among all feasible and cost-effective options. Given that the information gained from evaluation should be proportional to decisionmakers' needs, resources, and priorities, rigorous, controlled evaluations may be inappropriate, and "properly designed expert evaluations may be cost effective alternatives."[100] Harvey Averch writes that evaluators should consider using expert judgment when

- the program has been in place for many years and there is uncertainty surrounding the extent of historical inputs or activities
- the expected outcomes are highly uncertain
- the expected outcomes will occur far into the future
- the program design and inputs interact in unpredictable ways to produce outcomes.[101]

Eliciting expert judgment can take many forms, from informal "BOGSATs" to interviews with commanders to highly structured, iterative Delphi processes requiring consensus and insulation from personality or authority.[102] This section discusses expert elicitation methods used to inform IIP assessment.

---

[98] Author interview on a not-for-attribution basis, July 31, 2013.

[99] Valente, 2002, p. 70.

[100] Harvey A. Averch, "Using Expert Judgment," in Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds., *Handbook of Practical Program Evaluation*, San Francisco: Jossey-Bass, 2004, p. 292.

[101] Averch, 2004, p. 293.

[102] *BOGSAT* is a nonstandard but common acronym for "bunch of guys sitting around a table," not a particularly rigorous approach to expert elicitation.

The Delphi Method

The Delphi method, originally developed for forecasting trends, aims to generate consensus among experts through an interactive, iterative sequence of questions. After each round of questioning, respondents are encouraged to revise their answers in light of responses from the group. Delphi can be an effective method for characterizing the information environment in the formative phase and for evaluating abstract, long-term, or difficult to measure outcomes in the summative phase. Some IIP programs convene Delphi panels each year to provide annual indicators.[103] They can also inform the research process itself by identifying, for example, expert consensus on the most important constructs to measure or the most valid instruments and techniques. Taylor identifies six steps to conducting a Delphi for media evaluation:

1. Identify experts using a snowball or network sample.
2. Administer the first questionnaire on the topic of interest, consisting of a mix of open, semiopen, Likert, and quantitative questions.
3. Analyze responses for convergence and share anonymized responses with the group.
4. Administer the second questionnaire, encouraging respondents to revise or justify their original responses.
5. As needed, repeat steps 3 and 4, being careful to not incentivize groupthink.
6. Summarize results, highlighting areas of convergence and disagreement.[104]

The Electronic Decision Enhancement Leverager Plus Integrator (E-DEL+I) technique is an electronic real-time variation on the Delphi method that may have elements that may be more appropriate and cost-effective for DoD IIP evaluators. The process has four rounds, described in Figure 8.1, and can be completed in two to three hours. Because it can be completed in a short period of time, it avoids the risk of expert attrition, which can challenge traditional Delphi panels.

Self-Assessment/Interviews with U.S. Commanders

The individuals responsible for a program and its resources are experts. A "crude but rapid" method for assessing effectiveness is therefore to elicit judgment from "operating managers, higher-level administrators and budgetary sponsors." According to Averch, "obtaining judgments from those closest to a program is the most common kind of evaluation."[105] DoD IIP activities are no exception. It is common for IO assessment to be based on interviews with the U.S. commanders responsible for the IO campaign.

---

[103] Taylor, 2010, p. 6.

[104] Taylor, 2010, p. 6.

[105] Averch, 2004, p. 295.

Figure 8.1
The E-DEL+I Process
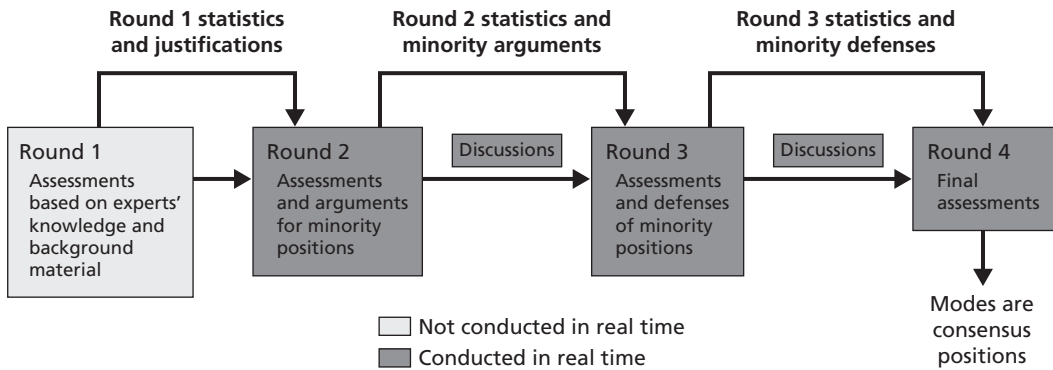
SOURCE: Carolyn Wong, *How Will the e-Explosion Affect How We Do Research? Phase 1: The E-DEL+I Proof-of-Concept Exercise*, Santa Monica, Calif.: RAND Corporation, DB-399-RC, 2003.
RAND *RR809/1-8.1*

The *Commander's Handbook for Assessment Planning and Execution* rationalizes this approach:

> In fast-paced offensive or defensive operations or in an austere theater of operations, a formal assessment may prove impractical. To assess progress in those cases, commanders rely more on reports and assessments from subordinate commanders, the common operational picture, operation updates, assessment briefings from the staff, and their personal observations.[106]

See the discussion on narratives for analysis and aggregation earlier in this chapter (in the section "Narrative Inquiry") and in Chapter Eleven (in the section "The Importance of Narratives").

However, the validity of these data is limited by response bias. Commanders have a strong incentive to emphasize the positive. This is, as RAND's Jason Campbell notes, "understandable and natural, even necessary, but it must be acknowledged so that battlefield commanders' assessments can be treated with a certain care and even skepticism at times."[107] The interviewer can, however, minimize the program managers' incentives to deceive by controlling the way the manager presents information (e.g., elicit specific examples that demonstrate impact) and by imposing direct or indirect penalties if deception is uncovered.[108] Overoptimism can also be controlled for

---

[106]U.S. Joint Chiefs of Staff, 2011c.

[107]Jason Campbell, Michael O'Hanlon, and Jeremy Shapiro, *Assessing Counterinsurgency and Stabilization Missions*, Washington, D.C.: Brookings Institution, Policy Paper No. 14, May 2009, p. 24.

[108]U.S. Joint Chiefs of Staff, 2011c.

to some extent through a formal system of devil's advocacy, in which all positive self-assessments are balanced by a formal and intentional worst-case interpretation of the facts. See the extended discussion of devil's advocacy in Chapter Four.

Despite several limitations, self-assessment data are better than no data and, if analyzed over time and triangulated with other data sources, can inform assessments of trends over time. These elicitations may also be particularly helpful for process- or improvement-oriented evaluation (i.e., determining why things did or did not happen).[109]

Other Qualitative Formative Research Methods

Kavita Abraham Dowsing described three other qualitative techniques that have been employed by the BBC: community assessments, temperature maps, and participatory photojournalism. Community assessments target disadvantaged or vulnerable populations and encourage them to express issues visually or in their own words. One application of this technique is to demonstrate a target audience's understanding of a contentious policy issue. Temperature maps are visual representations of issue saliency across geographic areas. They are generated from focus group questions that ask respondents to assign a level of importance to certain issues. In participatory photojournalism, subjects are asked to take pictures of the things that matter to them to gauge perceptions of governance.[110]

SMEs also discussed the *cultural consensus method*, which measures shared knowledge or opinions within groups. It is used in conjunction with focus groups and in-depth interviews to uncover the core of an issue while attempting to gain an understanding of the atmospherics and perceptions in different provinces.[111] The Darfur Voices project, a joint initiative between Albany Associates and researchers at Oxford, used the cultural consensus method to elicit narratives from both sides of the conflict and determine the points at which the narratives converge or their experiences have been similar.[112]

Summary

This chapter reviewed the data collection methods for formative evaluation and needs assessment and reviewed the qualitative data collection methods, such as interviews

---

[109] Author interview with Steve Booth-Butterfield, January 7, 2013.

[110] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[111] Author interview on a not-for-attribution basis, March 2013.

[112] Iginio Gagliardone and Nicole A. Stremlau, "Public Opinion Research in a Conflict Zone: Grassroots Diplomacy in Darfur," *International Journal of Communication*, Vol. 2, 2008.

and focus groups, that can be used to inform all three phases of evaluation. Key themes and takeaways include:

- DoD should invest more in qualitative and quantitative formative research to improve its understanding of the mechanisms by which IIP activities achieve behavioral change and other desired outcomes. The initial investment will pay off in the long run by reducing the chances of failure, identifying cost inefficiencies, and reducing the resource requirements for summative evaluation. If a program's logic model has been validated through rigorous formative research, programmers can have greater confidence in the effects of the message on exposed audiences.
- Messages and products should be pretested with qualitative techniques (e.g., focus groups) or with more-rigorous, controlled methods. Laboratory experiments are particularly valuable for the development and employment of messages and are underutilized by IIP researchers and planners.
- More psychological and behavioral research is needed to develop and validate the theories of influence that motivate DoD IIP campaigns. Very little research has been done, and the work that has been done was typically conducted on American college student subjects, so the conclusions may not fully generalize to other settings.
- Pilot-testing the intervention on a small scale and using computer-generated simulations can help refine the logic model and preemptively identify sources of program failure.
- Decisionmakers should avoid making decisions on the basis of a single quantitative method; triangulating with qualitative data is essential, given the subjective and complex nature of IIP campaigns. Quantitative data are often overemphasized, because they are easier to analyze and give a facade of rigor. Quantitative data are only as valid as the instruments that produced them and, often, encourage programmers to focus on less important outputs.
- The plural of *anecdote* is not *data*. Qualitative data should be generated by rigorous social science methods. And decisionmakers should not make decisions on the basis of a single qualitative method.
- Interviewing commanders is perhaps the most common method for assessing IIP campaigns. While this method alone is insufficient to determine effectiveness due to response bias, such input can complement other data sources, can inform the assessment of trends over time, and can be useful sources for process evaluation. Response bias can be minimized to some extent if there are known penalties for deception, if interviewers probe for specific demonstrations of impact, or if formal devil's advocacy is used.

Research Methods and Data Sources for Evaluating IIP Outputs, Outcomes, and Impacts

This chapter describes the research methods, measures, and data sources for postintervention evaluation of IIP campaigns, including process and summative evaluations. It describes the methods that help decisionmakers answer one of the core questions motivating this report: "Is IIP working?" The chapter begins with an overview of research methods and a discussion of the importance of the quality and quantity of data. It then describes the methods and data sources for process evaluation. The following sections describe the various components of summative evaluation, including techniques for measuring program exposure and changes in knowledge, attitudes (self-reported and observed), and individual and system behavior. The chapter concludes with a section on aggregation, analysis, and modeling for IIP evaluation. While this chapter provides an overview of the types of measures that are populated with survey research, the actual survey research methods are discussed in Chapter Ten.

Overview of Research Methods for Evaluating Influence Effects

The primary research methods and data sources for evaluating IIP effects are surveys; content analysis, including traditional media monitoring, web analytics, and social media monitoring and frame analysis; direct observation, or atmospherics; network analysis; direct response tracking; and qualitative methods, including focus groups, in-depth interviews, narrative inquiry, and Delphi panels. Secondary and aggregate data, such as data on economic growth or casualties, can also inform summative evaluations. Anecdotes and self-assessment, in which commanders evaluate progress made by subordinate units, are commonly used informal methods for gauging effectiveness.

NATO's framework for assessing public diplomacy summarizes several of these methods in a table that maps each method to the resources required and a time frame for results. A modified version of this *menu of research methods* is presented in Table 9.1. Detail on each method can be found in subsequent sections of this chapter or in the section on qualitative research methods in Chapter Eight.

Table 9.1
Menu of Research Methods for Assessing Influence Activities

Research Method	Role in Preintervention Evaluation	Role in Postintervention Evaluation	Resources Required	Validity	Time Frame for Results	Manpower Requirements	Limitations
Representative survey	Characterize IE and baseline	Measure exposure and attitudes	High	High	Immediate to several weeks	Survey research group, locals	Access, nonresponse, and response bias
Content/sentiment analysis: traditional media	Characterize IE	Measure distribution and changes in attitudes and beliefs	Medium	Medium high	Weeks	Outsource, local coders	Unrepresentative samples, difficult to code
Content/sentiment analysis: online and social media	Characterize IE	Measure changes in attitudes and beliefs	Low	Low medium	Immediate	Limited, mainly software requirements	Unrepresentative samples, limited to tech-savvy audiences
Online and social media analytics (of DoD messages)	N/A	Measure exposure and reactions (web-based campaigns)	Low	High	Immediate	Limited, mainly software requirements	Only relevant to web-based messages
Informal surveys/ intercept interviews	Test products and characterize IE	Measure attitudes and beliefs	Low	Low	Near term (weeks)	In-house	Not representative, nonresponse and response bias
In-depth interviews	Develop messages	Interpret quantitative results	Medium	Medium	Near term (weeks)	Local researchers or in-house	
Focus groups	Develop messages and test products	Validate and interpret quantitative results	Medium	Medium	Days to months	Local facilitators, often outsourced	Groupthink, difficult to manage, selection bias
Laboratory experiments	Develop messages and theories of change	N/A	Medium high	High	Months	Academic researchers	Requires planning, results can be hard to operationalize
Direct observation and atmospherics	Characterize IE	Measure change in attitudes and beliefs	Medium high	Medium	Days to months	In-house or outsourced	"Signal in noise," no systematic approach
Secondary data/ desk research	Characterize IE and baseline	Measure exposure (e.g., using process similar to Nielsen ratings)	Low	Medium high	Immediate (weeks)	In-house	No control over research design or questions

> **Box 9.1**
> **A Note on the Importance of Data to IIP Evaluation**
>
> Collecting or arranging for the collection of sufficient quantities of sufficiently high-quality data should be a priority for any IIP assessment team. Data on IO programs are often lacking, irrelevant, or not validated.[a] Even the most sophisticated analytical techniques cannot overcome bad data. While in some contexts modeling and sophisticated techniques may be valuable, "validated data simply do not exist in large-enough quantities to put those models to use."[b] Assessment guidance should prioritize equipping assessment teams with the resources and skills needed to generate and validate appropriate data, and sufficient assessment design skills to identify which measures need to be supported by high-quality data and which can be adequately covered with less rigorous data collection.
>
> Importantly, *good data* is not synonymous with *quantitative data*. Depending on the methods, qualitative data can be more valid, reliable, and useful than quantitative data. As addressed in Chapter Eight in the section on qualitative methods, expressing data numerically does not make them objective, particularly given the highly subjective nature of the instruments often used to generate quantitative data for IIP assessment. The quality of a data set should be judged on the basis of the validity and reliability of the methods used to generate it rather than whether it is quantitative or qualitative.
>
> [a] Author interview with LTC Scott Nelson, October 10, 2013.
> [b] Author interview with LTC Scott Nelson, October 10, 2013.

Before collecting new data, analysts must evaluate the appropriateness of available existing secondary data sources. While the use of secondary data sacrifices control over the data-generating process, it saves substantial effort and cost. In *Fantasy Analytics*, Jeff Jonas proposes guidelines for the prioritization of new data sources. With all else equal, data sources should be added in the following order: data already collected within your organization, external data that can be purchased or otherwise acquired, primary data collection.[1] This holds particularly true in the defense IIP context; if the intelligence community is already collecting something that even partially meets assessment requirements, savings in terms of time and resources can be significant. See the additional discussion of the benefits of using data that are already being collected in the section "Assessment and Intelligence" in Chapter Four.

Several experts we interviewed stressed the importance of leveraging existing surveys conducted by other organizations, particularly in Afghanistan, to save resources, improve the quality of the surveys, and to reduce the risk of survey fatigue. For example, in some operating environments, it may be possible to use Nielsen panel data for measures of exposure and audience.

Measuring Program Processes: Methods and Data Sources

This section describes the methods and data sources associated with process evaluation. Process evaluation, or program implementation monitoring, seeks to determine

---

[1]   Jeff Jonas, "Fantasy Analytics," blog post, *Jeff Jonas*, November 9, 2012.

the extent to which the program accomplished the tasks it was supposed to accomplish. It is therefore principally concerned with measuring things over which program staff have direct or significant control. Process evaluation is particularly important in cases in which the program failed or fell short of expectations. If the process evaluation reveals that the program was implemented as planned, it tells the program designers that the logic model needs to be revisited, as this would appear to be an instance of potential theory failure rather than program failure (see the discussion in the section "Program Failure Versus Theory Failure" in Chapter Five).

As introduced in Chapter Seven, process evaluation can be conducted at several points in the campaign process: message or program *production* and message *dissemination*. *Production* evaluation documents how the message or program was created. *Dissemination* evaluation measures the distribution and placement (including the *volume*, *channel*, and *schedule*) of messages or the number of events and engagements, depending on the type of campaign.[2] While some researchers include measuring exposure as a component of process evaluation, this report addresses exposure measures separately.

Production measures focus on the time it takes to make a product and the extent to which products were made to specification. *Implementation and dissemination* measures depend on the nature of the IIP activities being evaluated. For messaging campaigns, they include measures of dissemination, including message distribution and placement. *Distribution measures* assess the types and numbers of materials disseminated (e.g., public service announcements, news feeds, brochures, op-eds). *Placement measures* assess the *volume*, *channel*, and *schedule* (time and duration) of message distribution, including timing and frequency of broadcasts, amount of publicity received, the number of times an op-ed ran, downloads of a public service announcement, and so forth.[3] If the campaign outputs being evaluated are engagements (e.g., senior-level engagements, student exchanges) rather than messages, process measures should address the frequency, variety, and quality of events.

Brian Cullin, who served as the senior adviser to the under secretary of state for public diplomacy, urges evaluators to be sensitive to how output measures might be perceived by foreign audiences. Metrics such as "the number of articles in support of U.S. policies" can damage the credibility of the campaign if released publicly.[4] Moreover, when it comes to output measures such as "number of engagements," more is not always better. One SME recounted an example in which a partner national was overwhelmed by too many senior-level visits and "had to cry, 'Enough!'"[5]

The primary sources of data for program implementation measures are direct observation or monitoring of program implementers, media monitoring, service record

---

[2]   Valente, 2002, pp. 75–77.

[3]   Coffman, 2002, p. 21.

[4]   Author interview with Brian Cullin, February 25, 2013.

[5]   Author interview on a not-for-attribution basis, December 5, 2012.

Box 9.2
Documenting DoD Actions and Other Program Inputs

Our interviews suggest that DoD needs to improve its processes for documenting and tracking the inputs and outputs associated with its own activities and programs. In Afghanistan, for example, DoD has done a good job of cataloging what the insurgency has done but a very poor job of cataloging what its own forces have done. And the vast amount of what does manage to be collected is lost when units transition. Jonathan Schroden characterized current efforts as "abysmal" and in need of being systematically addressed: "Even if we're tracking outcomes, it's impossible to know what's working if we don't know what we're doing."[a] Others identified this as a limitation to efforts to evaluate other U.S. government strategic communication efforts. Mark Helmke argued the government needs to keep better records of its public diplomacy engagements, documenting the activities that took place, when they happened, and the individuals engaged. These records would allow evaluators to look back and see whether those who were engaged went on to make influential decisions.[b]

[a] Author interview with Jonathan Schroden, November 12, 2013.

[b] Author interview with Mark Helmke, May 6, 2013.

---

data, service provider data (e.g., interviews with program managers), and event participant or audience data. When using direct observations, researchers should be sensitive to the Hawthorne effect, in which subjects are likely to exert extra effort if they are aware they are being observed. Media monitoring should assess message distribution and placement. When analyzing service record data (e.g., data routinely collected by program management or staff), it is better to analyze a few items collected consistently and reliably than a comprehensive set of information of poor quality.[6] Interviews of commanders or program managers are a common data source for DoD IIP evaluation. While these data have poor validity for evaluating program effects, they can be useful sources of information for documenting program implementation and identifying potential sources of failure.[7]

Measuring Exposure: Measures, Methods, and Data Sources

IIP summative evaluations should include a measure of exposure to the campaign and several measures that capture the internal processes by which exposure influences behavioral change. This section discusses methods for capturing exposure. Subsequent sections address methods for measuring the internal processes—knowledge, attitudes, and so forth—affected by exposure.

The first step in assessing the outcome of an IIP campaign is measuring the extent to which the target audience was *exposed* to the program or message. Program exposure is the degree to which an audience *recalls* and *recognizes* the program. *Recall* is measured by unaided or spontaneous questions that ask the respondent in an open-

---

[6]   Author interview with Thomas Valente, June 13, 2013.

[7]   Author interview with Steve Booth-Butterfield, January 7, 2013.

ended manner if he or she had been exposed to the campaign.[8] *Format-specific recall* establishes whether the audience member recalls the information from the campaign (e.g., a public service announcement) or from other sources (e.g., state news bulletin).[9] *Recognition* is measured by aided or prompted questions that provide a visual or aural cue to assist the respondent in recalling the campaign.[10] Recognition measures have greater response bias.[11]

Recall and recognition measures assess exposure along two dimensions: message awareness—measured by reach, frequency, and recency—and message comprehension. *Reach* assesses the number of people who saw or heard the message, and is typically defined as the percentage of the target audience exposed to the message at least once during the campaign. *Frequency* measures how often the individuals saw the message, defined as the average number of times a person in the target audience had the opportunity to view the message.[12] *Recency* measures are common in IIP evaluation and capture the last time the media was viewed. *Comprehension* is the extent to which the audience understood the message.[13] Exposure measures can therefore be conceptualized along two dimensions, as shown in Table 9.2: aided versus unaided and awareness versus comprehension.

Researchers should not make assumptions about exposure based on distribution. Importantly, *reach* is defined in terms of the audience's ability to recall or recognize (e.g., whether they tuned in as opposed to whether the media was playing). While commonly used, experts strongly discouraged the use of media impressions, which gauges the potential audience, as a proxy for measuring audience. Julia

Table 9.2
Two Dimensions of Campaign Exposure

Dimension	Awareness of Message (reach, frequency, recency)	Comprehension
Recall (unaided)	Did you hear or see something? How often? How recently?	What was the message?
Recognition (aided)	Pictorial, video, or aural cues are provided	Themes are read; respondent is asked about those themes

SOURCE: Adapted from interview with Thomas Valente, June 18, 2013.

---

[8]   Valente, 2002, p. 184.

[9]   Gerry Power, Samia Khatun, and Klara Debeljak, "'Citizen Access to Information': Capturing the Evidence Across Zambia," in Ingrid Volkmer, ed., *The Handbook of Global Media Research*, Chichester, West Sussex, UK: Wiley-Blackwell, 2012, p. 263.

[10]   Valente, 2002, p. 184.

[11]   Author interview with Ronald Rice, May 9, 2013.

[12]   Author interview with Thomas Valente, June 18, 2013.

[13]   Power, Khatun, and Debeljak, 2012, p. 263.

Coffman argues that the measure grossly exaggerates true exposure: "It's a numerator in search of a denominator."[14] In Phil Seib's opinion, "reach is a joke" and a "naturally inflated number [that] does not reflect actual audience" when reach is defined in terms of potential audience.[15] Likewise, Gerry Power has observed that far too much of the exposure work in this sector assumes that "just because someone has access to a radio or TV set" he or she will attend to and comprehend the message.[16] What people are actually exposed to is usually a subset of what you put out.[17]

Capturing Variance in the Quality and Nature of Exposure
Exposure should be measured at multiple tiers and along several dimensions to capture variance in the quality and nature of exposure. Measures of amount (how much), frequency (how often), and quality (media engagement) are all important, and evaluations should include survey instruments that can capture those differences in the nature of exposure. Power noted there is a "long journey" between having a radio and being affected by a message: "Having a radio doesn't mean having a signal; having a signal doesn't mean listening; listening doesn't mean listening to the right program; and listening to the right program doesn't mean listening in an engaged manner."[18]

Charlotte Cole argues that the field needs better measures for capturing variation in the quality of engagement, especially in the formative setting. Researchers at Sesame Workshop often use "eyes on the screen" to measure the engagement among children, but this approach has questionable reliability. Often, a subject's eyes are intently focused on the screen, but he or she is thinking about something entirely unrelated.[19] Emmanuel de Dinechin suggests adding questions about how engaged the audience was while watching the message—for example, whether people were cooking or engaged in another activity at the same time.[20]

Measures of exposure do not need to be dichotomous (e.g., have you seen it or have you not?). Evaluations of exposure can use scales, indexes, and multidimensional approaches to build in variance at multiple levels.[21] In their piece about citizen access to information, Power and colleagues propose an "index of exposure" facilitated by post hoc aggregation and analysis of questions, response categories, and scales. Such an

---

[14] Author interview with Julia Coffman, May 7, 2013.

[15] Author interview with Phil Seib, February 13, 2013.

[16] Author interview with Gerry Power, April 10, 2013.

[17] Author interview with Ronald Rice, May 9, 2013.

[18] Author interview with Gerry Power, April 10, 2013.

[19] Author interview with Charlotte Cole, May 29, 2013.

[20] Author interview with Emmanuel de Dinechin, May 16, 2013.

[21] Author interview with Gerry Power, April 10, 2013.

index allows researchers to assess dose-dependent effects of exposure (that is, whether changes in outcomes vary with changes in the degree or extent of exposure).[22]

Because the quality or appropriateness of the audience is more important than its size, the best exposure measures address the extent to which well-defined strategic audiences have been exposed to the message.[23] The measure denominator should therefore be the target audience, rather than the population at large. Overly broad or amorphous audiences create challenges and undermine the cost-effectiveness of both the execution and evaluation of the campaign.[24]

Methods and Best Practices for Measuring Reach and Frequency

This section discusses methods and associated techniques for measuring the many dimensions of campaign exposure. Primary methods include survey research, household panels or "people meters" (e.g., Nielsen families), real-time return path data for monitoring cable and satellite television usage, tracking rumors, web and mobile analytics, social media analysis, and direct response tracking. Depending on the environment and scale of the campaign, measuring exposure is an area in which secondary data sources can be both cost-effective and of higher quality (e.g., commissioned audience survey research, return path data from providers, Nielsen ratings).

Survey-Based Techniques for Assessing Exposure

Exposure is commonly measured with self-reported assessments of exposure captured by surveys.[25] This section discusses techniques unique to measuring exposure associated with survey research. For a detailed discussion of surveys and survey research for IIP evaluation, see Chapter Ten. Experts discussed several best practices for capturing and validating exposure data.

- *Ask about the content of the show rather than whether someone watched it.* Reach should capture the extent to which audiences actually tuned in or engaged the media. To minimize response bias, surveys should avoid questions like "have you watched *x* program?" due to response bias. Better questions ask subjects to recall or recognize characters, themes, or messages from the program.[26]
- *Use "ringers" and other tests to improve the validity of recognition, or aided recall measures.* The best recognition measures are those that include images from the

---

[22] Power, Khatun, and Debeljak, 2012, pp. 263–266.

[23] Author interview with Kim Andrew Elliot, February 25, 2013; interview with Mark Helmke, May 6, 2013.

[24] Author interview with Joie Acosta, March 20, 2013.

[25] Martin Fishbein and Robert Hornik, "Measuring Media Exposure: An Introduction to the Special Issue," *Communication Methods and Measures*, Vol. 2, Nos. 1–2, 2008.

[26] Marie-Louise Mares and Zhongdang Pan, "Effects of Sesame Street: A Meta-Analysis of Children's Learning in 15 Countries," *Journal of Applied Developmental Psychology*, Vol. 34, No. 3, May–June 2013.

actual message as well "ringers," or images that the respondent was very unlikely to have seen at all. This helps researchers weed out response bias by flagging the respondents who will recognize images they have not actually seen.[27]

- *Use context-specific recall and recognition measures for popular or iconic themes and characters.* Unaided and aided recall measures can be biased when the character has become iconic. Respondents may recognize the character even if they were not exposed to the program under evaluation.[28]
- *Use multi-item measures of exposure like scales and indexes.* Exposure is multidimensional and not dichotomous. Dose-dependent effects of exposure can be identified through the use of indexes that aggregate across many questions and scales.[29]
- *Validate exposure measures by analyzing whether responses correlate with how recently the program was aired.* If respondents are more likely to recall recent images than older images, the measure is more likely to be valid.[30]

Off-the-Shelf and Commissioned Viewership Data

Exposure is frequently measured through secondary data. Members of several organizations we interviewed said that they typically measure reach by purchasing consolidated viewership or through commissioned audience surveys when consolidated data are unavailable.[31] In postconflict and developing countries, firms such as Nielsen typically do not have a permanent presence due to insufficient demand from advertisers. Audience research in these environments is therefore typically done through one-off commissioned studies. However, as advertising firms shift their attention to developing markets, audience research capacity will improve, and DoD may have greater access to consolidated viewership data in key operating environments.[32]

Consistent data from an audience research organization with a permanent presence is better and more cost-effective than data from one-off commissioned surveys. It is difficult to do rigorous or in-depth audience analysis affordably without a sustained research presence. According to de Dinechin, moreover, "you cannot trust a single snapshot" to be representative of the media environment, because media share can fluctuate very rapidly in postconflict environments. In Iraq, for example, the top-ten

---

[27] Author interview with Thomas Valente, June 18, 2013. For more on valid recognition measures, see Valente, 2002, chapter 11, "Measuring Program or Campaign Exposure."

[28] Author interview with Charlotte Cole, May 29, 2013.

[29] Power, Khatun, and Debeljak, 2012, pp. 263–266.

[30] Author interview with Thomas Valente, June 18, 2013.

[31] Author interview with Marie-Louise Mares, May 17, 2013; interview with James Deane, May 15, 2013.

[32] Author interview with Emmanuel de Dinechin, May 16, 2013.

channels capture only about a quarter of the total market share. So the top-three channels will often be the sixth through ninth channels in the following month.[33]

Return Path Data Versus People Meters

With the availability of real-time return path data on viewership, some experts question the relevance of the "people meter" model that uses viewer diaries or meters connected to a television to track the viewing habits of household panels (e.g., Nielsen families). The "ubiquity of digital set-top-boxes" is enabling cable and satellite media providers to collect data on audience viewership as a "by-product of their subscriber management processes."[34] *Return path data*, also called "set-top box data," are any data that can be retrieved from the return path or backchannel, providing electronic communication between the subscriber and the platform company.[35] Return path data are becoming available not only from linear television but also from DVR playback, video-on-demand sessions, interactive television applications, the electronic program guide, and remote controls. Return path data sources include digital set-top boxes, Internet and mobile tracing, and other network monitoring tools such as switched digital video, which is being "rapidly deployed . . . to enable more efficient use of bandwidth."[36]

In light of these data sources, Johanna Blakely argues that the Nielsen families model is a clumsy, cost-ineffective, and primitive approach that "needs to die." Return path, Internet, and mobile data tell programmers "exactly who is watching what and when" rather than estimating audience demographics from unrepresentative samples. Depending on the source, these data can paint a rich psychographic profile of real-time viewers. In her view, the Nielsen families model has only endured because the innovations threaten established institutions, and the entertainment industry is reluctant to share and make transparent their audience analysis tools.[37] Because return path data are collected passively as a by-product of the subscription model, this approach is significantly less resource intensive than household panels or commissioned surveys.

Rumor Tracking

Anthony Pratkanis encourages IIP researchers to apply a technique used by R. H. S. Crossman during World War II to measure the influence of the information pamphlets that his units disseminated. Crossman would visit the building that was keeping track of all of the rumors through HUMINT and other means and would check to see if his

---

[33] Author interview with Emmanuel de Dinechin, May 16, 2013.

[34] Ian Garland, "Return Path Data: A 21st Century Business Tool," undated.

[35] Coalition for Innovative Media Measurement, *CIMM Lexicon 1.0, Terms and Definitions: A Common Language for Set-Top Box Media Measurement*, New York, May 2010, p. 132.

[36] Coalition for Innovative Media Measurement, 2010, p. 2.

[37] Author interview with Johanna Blakely, June 24, 2013.

own messages were circulating as rumors in the adversary information environment.[38] This requires, of course, that someone collect the rumors.

Web Analytics

This section discusses the metrics used to assess the exposure (frequency and reach) of web-based content. Additional measures derived from web and social media content, including content-generated measures of sentiment or influence, are discussed in a subsequent section on content analysis and social media monitoring.

Using data from web and mobile sources to assess exposure has several advantages. First, because data collection is built into the dissemination platform itself (e.g., downloads or site visits), these data can be collected at no to minimal cost. Second, depending on the platform, the information can be richer: researchers can assess how viewers behave after being exposed to the message and can construct a detailed psychographic profile of the audience based on web activity. Third, because web behavior is directly observed, these data avoid the response bias or response acquiescence issues that limit the validity of self-report survey measures. While it is difficult to directly assess unaided recall, there are proxy measures for the extent to which the audience is engaged with the media, such as the time spent on a page, comments or shares, "likes," and how people click through content.

However, web analytics can only measure the reach of web-based content, which is not a widely used medium in many DoD IIP campaigns due to the technology use and media consumption habits of target audiences. Moreover, it is often difficult to find the signal in the noise, and doing so requires advanced analytical techniques that may not be accessible to IIP units. These data are also not generated from a representative sample.

Exposure metrics derived from web analytics fall into one of the three broad categories: traffic analysis, navigation analysis, and market-based analytics (e.g., shares and downloads). These and other metrics can be generated by web analytics platforms like Google Analytics Premium.[39]

- *Traffic analysis* assesses awareness of the campaign and the extent to which users are engaging the content. Basic traffic analysis metrics include *page views*, *unique visitors*, and the *average engagement time* that a user spent interacting with the site or app, as well as *bounce rate*, the number of users who exit the site or app before exploring linked elements. Organizations can combine these numbers to produce metrics that paint a more interesting picture of user engagement, such as *churn*,

---

[38] Author interview with Anthony Pratkanis, March 26, 2013.

[39] Author interview with Maureen Taylor, April 4, 2013.

the number of users lost over time divided by total users, and *stickiness*, the time spent viewing all pages divided by the total number of unique visitors.[40]

- *Navigation analysis* shows how users use the platform (website or app) to find information once they enter it in order to assess the extent to which the user interface meets the needs of the audience. Organizations track navigation through "click streams" and the time spent in particular areas of the platform.[41] Many organizations also use *hyperlink analysis* to identify how a user navigated to their websites.

- *Market-based analytics* gauge interest in the material or products and can include downloads, shares, likes, requests for information, and "conversions," such as registering for a website or an event. Some organizations include a *relevance factor* measure, defined as the number of products downloaded or consumed by users divided by the number of available products.[42]

Depending on network characteristics, organizations can combine these metrics with information on user cookies or geographic profiles of users' Internet protocol addresses to paint a detailed picture of exposure and demand, segmented by demographic and psychographic characteristics. These data help researchers understand who is listening and how to best engage strategic audiences.

Often, the signal sent by changes in web analytics can be misinterpreted. For example, longer session times in response to a new website design could signal greater interest *or* the inability to find needed information. To properly interpret web analytics, it is important to triangulate web analytics with other data sources, such as focus groups or user feedback surveys.[43]

A Note on Vanity Metrics

When analyzing web metrics, researchers should avoid overemphasizing *vanity metrics*, which should not drive decisions or be explained by changes in the IIP program. The notion of vanity metrics has been popularized by the lean startup movement. In his book *The Lean Startup*, Eric Reis proposes the *innovation accounting* principle, which

---

[40] For a discussion on web traffic, as well as navigation, indicators, and tools, see Martin J. Eppler and Peter Muenzenmayer, "Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology," *Proceedings of the 7th International Conference on Information Quality*, Cambridge, Mass.: MIT Sloan School of Management, 2002; NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 46.

[41] Arun Sen, Peter A. Dacin, and Christos Pattichis, "Current Trends in Web Data Analysis," *Communications of the ACM*, Vol. 49, No. 11, November 2006.

[42] A. Phippen, L. Sheppard, and S. Furnell, "A Practical Evaluation of Web Analytics," *Internet Research*, Vol. 14, No. 4, 2004.

[43] Michael Khoo, Joe Pagano, Anne L. Washington, Mimi Recker, Bart Palmer, and Robert A. Donahue, "Using Web Metrics to Analyze Digital Libraries," *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, New York: ACM, 2008.

holds that the only metrics a learning organization should invest resources in collecting and analyzing are those that help drive decisionmaking, which implies a move away from vanity metrics and toward *actionable metrics*. Vanity metrics are metrics that are unable to explain what is driving changes in values or provide direction for how an organization should move forward. Examples can include aggregate website traffic or registered users—metrics that are highly volatile and may not correlate with active users or other outcomes of interest. Aggregate website traffic and monthly earnings can both serve as vanity metrics. Actionable metrics, by contrast, are those that were derived from experimental conditions such as split testing and that are capable of assigning causality to changes in observed customer behavior.[44]

Measuring Self-Reported Changes in Knowledge, Attitudes, and Other Predictors of Behavior

Chapter Six introduced the "knowledge leads to attitudes leads to practices" (KAP) model of behavioral change; decomposed the model into a sequence of discrete, measurable steps on the path from exposure to behavioral change; and explained that the more stages that are measured, the better the researchers will be able to understand the effects of the message on behavior. The previous section discussed methods for capturing program exposure. The balance of this chapter discusses self-reported and directly observed methods for capturing the internal processes of behavioral change that occur following program exposure. This section discusses several of the self-reported measures of these constructs, including measures of knowledge or awareness, issue saliency attitudes, self-efficacy, norms, and behavioral intention.

Knowledge or Awareness Measures

Knowledge or awareness measures capture the extent to which the target audience understands or is aware of the position being advanced by the messaging campaign. The best knowledge measures administer actual tests of knowledge of the issue area before and after the intervention—to exposed and unexposed cross sections. However, such measures are resource intensive to administer, burdensome to the respondent, and not always appropriate to the content of the message. Alternatively, organizations can assess self-reported changes in knowledge. For BBC Media Action governance campaigns, a primary outcome measure is knowledge, defined as the percentage of people who report having increased knowledge as a result of exposure. In that instance, knowledge is not narrowly defined in terms of a specific issue. BBC Media Action health

---

[44] Ries, 2011.

campaigns, by contrast, elicit knowledge regarding the specific behavioral change that is sought (e.g., breastfeeding and other infant-mortality-risk behavior modifications).[45]

Some experts argue that there is an overreliance on awareness as a predictor of behavioral change. According to Coffman, programmers often make the mistake of substituting awareness for impact. Awareness is a necessary condition for behavior change but is by no means sufficient.[46]

Measuring Self-Reported Attitudes and Behavioral Intention

After exposure to and comprehension of the campaign have been established, IIP evaluations seek to determine the nature and extent to which comprehension of the message has shaped and will continue to shape target-audience behaviors. Because behavioral outcomes of interest are often unobservable, researchers may measure attitudes and other predictors of behavioral change. This section discusses the validity and use of measures that predict behavioral outcomes.

Attitudes Versus Behaviors

Chapter Five (in the section "Behavioral Versus Attitudinal Objectives") noted that there is a schism in the IIP field over whether attitudinal objectives are valid, or whether planners should only be concerned with behavioral change. Naturally, this debate extends from planning into the realm of measurement. On the one hand, stated preferences are imperfect measures of true preferences because individuals have difficulty introspectively assessing their likes and dislikes.[47] Evidence suggests that people have little or no insight into their own information processing. Richard Nisbett and Timothy Wilson showed that, many times, individuals are unaware that their responses were influenced or of what influenced their responses.[48] Individuals may also deliberately conceal their true preferences. Victoria Romero has found that the challenges with self-reported measures are compounded by challenges of opinion polling in conflict environments because, for example, "it is very difficult to compel individuals to be honest when they think they're always under surveillance."[49] Moreover, there is not an unambiguous causal directional link between attitudes and behaviors. In cognitive dissonance theory, for example, behavioral change can precede attitudinal change.[50]

---

[45] Author interview with James Deane, May 15, 2013.

[46] Author interview with Julia Coffman, May 7, 2013.

[47] Norbert Schwarz, "Attitude Measurement," in William D. Crano and Radmila Prislin, eds., *Attitudes and Attitude Change*, New York: Psychology Press, 2008.

[48] Richard E. Nisbett and Timothy D. Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, Vol. 84, No. 3, March 1977.

[49] Author interview with Victoria Romero, June 24, 2013.

[50] Author interview with Victoria Romero, June 24, 2013.

On the other hand, attitudinal change often precedes behavioral change and may have a more lasting and profound effect than an observed change in a particular behavior. Christopher Rate and Dennis Murphy argue that an exclusive focus on behavior is "myopic" and sells IIP operations short of their full potential. They maintain that "most external influences (e.g., media) do not shape behavior directly, but affect change through processes in the cognitive domain of the information environment."[51] Moreover, many of the behavioral outcomes of interest for DoD IIP are difficult or impossible to observe in the short or intermediate term. Thus, focusing exclusively on behaviors could produce false-negative assessments and lead to the premature termination of otherwise effective programs.[52] Threats to validity associated with self-reported measures can be minimized with large samples and consistent measurement over time (unless those biases are correlated with time or space, it is possible to elicit valid estimates of the average difference in pre- and post- or exposed versus unexposed attitudes).

Good formative research and logic modeling can help determine the relative importance of measuring attitudinal versus behavioral measures. In some instances, only the attitudinal mediator of the behavior matters. If research has demonstrated, for example, that teens are smoking because they think cigarettes are cool, the evaluation only needs to assess whether the campaign is changing perceptions.[53] In other cases, attitudes may not matter at all. For example, if the goal is to prevent opium farming, it is likely much more effective to encourage other crop options than to have an antidrug strategic communication campaign.[54]

While the most valid measures track how people actually behave, they are rarely the most feasible and often not the most useful. Martin Fishbein and colleagues found that, when behavior cannot be observed directly, the most important outcome measures for behavioral-change communication campaigns are *attitudes toward the behavior*, *norms about the behavior*, and *behavioral intention*.[55] Behavioral intention, the likelihood that a person will engage in a specific behavior, derives from the theory of reasoned action and is frequently identified as the best predicator of actual behavior among self-report measures.[56]

---

[51] Rate and Murphy, 2011, p. 10.

[52] Rate and Murphy, 2011, p. 9.

[53] Author interview with Thomas Valente, June 18, 2013.

[54] Author interview with Victoria Romero, June 24, 2013.

[55] Fishbein, Triandis, et al., 2001.

[56] Author interview with Ronald Rice, May 9, 2013; interview with Julia Coffman, May 7, 2013; Icek Ajzen and Martin Fishbein, "A Theory of Reasoned Action," in *Understanding Attitudes and Predicting Social Behavior*, Upper Saddle River, N.J.: Pearson, 1980.

Best Practices for Eliciting Self-Reported Attitudes and Behavioral Intentions

Experts discussed several techniques and best practices for eliciting valid and useful measures of attitudes and behavioral intentions.

- *Questions should be precise, context-specific, and, where appropriate, sequential.* Public diplomacy experts have found that the best questions elicit thematic, subject-specific attitudes. For example, rather than asking generically about attitudes toward the United States, questions should ask about attitudes toward U.S. foreign policy on a particular topic.[57] Likewise, behavioral intention measures in health communication campaigns are the most valid when they are very specific and layered from more generic behavior to specific, contextual behavior. Because behaviors typically consist of many steps, Ronald Rice suggested decomposing measures of behavior intention into several questions about each step in the behavioral sequence.[58]
- *In potentially hostile environments, measures of attitudes toward Western institutions should be reasonable and culturally appropriate.* Questions designed to elicit attitudes about the United States or Western institutions should be carefully worded to avoid setting unreasonable expectations. Phil Seib explained that in places like Egypt, "a positive outcome is not going to mean wearing American flag lapel pins. . . . But if you can, for example, encourage them to rely on more-reasonable or credible sources of information about America, . . . you're succeeding."[59]
- *Use standardized scales and multi-item measures to assess attitudes and behavioral intentions.* Multi-item standardized scales for values and attitudes such as the Schwartz Value Inventory are more robust than single-item measures, because they control for response bias and it is possible to test the reliability of the different items against one another.[60]
- *Scales and indexes should also be used for measuring behavior intention.* The film industry uses a "definite interest" measure to gauge whether an individual will buy a ticket to a movie. But because most people say that they're "definitely interested," these measures have greater predictive validity if combined into an "index of definite interest" that teases out variation in definite interest.[61]

---

[57] Author interview with Gerry Power, April 10, 2013.

[58] Author interview with Ronald Rice, May 9, 2013.

[59] Author interview with Phil Seib, February 13, 2013.

[60] Author interview with Gerry Power, April 10, 2013. For more on the Schwartz Value Inventory, see Shalom H. Schwartz, "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries," in Mark P. Zanna, ed., *Advances in Experimental Social Psychology*, Vol. 25, San Diego, Calif.: Academic Press, 1992; and Shalom H. Schwartz, "Beyond Individualism/Collectivism: New Dimensions of Values," in Uichol Kim, Harry C. Triandis, Cigdem Kagitcibasi, Sang-Chin Choi, and Gene Yoon, eds., *Individualism and Collectivism: Theory, Method, and Applications*, Thousand Oaks, Calif.: Sage Publications, 1994.

[61] Author interview with Vincent Bruzzese, June 7, 2013.

- *Scales need to be adapted to local contexts.* Standardized social science attitudinal and value scales are generally not well formulated for the environments within which DoD is conducting influence operations.[62] The discriminant validity of value scales is culturally dependent. In the film industry example, responses to the definite-interest measure vary widely by cultural context: "In the U.S., everyone says they're definitely interested. But in Japan, no matter what, we get 15-percent definite interest. Nobody wants to commit. So the standard scales don't work, and we have to tailor them."[63]
- *For DoD influence campaigns, attitudinal measures should focus on attitudes toward the adversary.* Because DoD influence activities are more akin to countermarketing, attitudinal measures should address attitudes toward the adversary and adversary institutions rather than just attitudes toward the coalition.[64] Measures should address an adversary's reputation and resonate (e.g., whether an adversary is viewed as a "troll" or whether the adversary has been shamed out of a space).[65]

Measuring Mediators of Behavioral Change

The link between attitudes and behaviors is mediated by several cognitive processes, including self-efficacy, interpersonal discussion, issue saliency, and norms. To improve the validity of self-report measures, these mediators should be measured. Research conducted by Joyee Chatterjee and colleagues using structural equation modeling to evaluate the efficacy of an HIV/AIDS awareness campaign demonstrated that there is a direct link from media to knowledge acquisition and attitudinal change, but that the link between attitudes and behavioral change is mediated by two factors that "bridge" or "catalyze" behavioral change: self-efficacy and interpersonal discussion.[66]

Self-efficacy is a person's belief that he or she has the ability or competency to perform a behavior, and its ability to predict behavioral outcomes derives from the theory of social cognitive learning.[67] Self-efficacy with respect to HIV/AIDS–related behaviors was measured by Chatterjee and colleagues with survey items like, "If I think necessary, I would insist on using a condom with my partner," and "I can communicate freely with my spouse on matters concerning sex."[68]

The extent to which the program or message promotes *interpersonal discussion* has been found to be a strong predictor of behavioral change, particularly when discuss-

---

[62] Author interview with Amelia Arsenault, February 14, 2013.

[63] Author interview with Vincent Bruzzese, June 7, 2013.

[64] Author interview with Victoria Romero, June 24, 2013.

[65] Author interview on a not-for-attribution basis, July 31, 2013.

[66] Chatterjee et al., 2009.

[67] Coffman, 2002, p. 22.

[68] Chatterjee et al., 2009, p. 629.

ing sensitive topics.[69] Chatterjee and colleagues measured the propensity to engage in interpersonal discussion by giving respondents a list of topics and asking them if they have discussed them, and with whom.[70]

Measures of saliency, or perceptions of the importance of an issue, are important predictors of behavior and are often overlooked by evaluators.[71] Research has suggested that saliency is more predictive of behavior than being informed or opinionated. There may be an inverse relationship between awareness levels and changes in saliency—those with less awareness are more likely to show saliency increases.[72] As a consequence, researchers should not make assumptions about issue saliency based on awareness.

Measures of social norms gauge perceptions of acceptable attitudes and behaviors among the respondent's social network. Coffman argues that norms are often "the most critical factor in achieving behavior change" but frequently go unnoticed in the campaign design and evaluation phase due to a myopic focus on the knowledge-attitudes-practices construct.[73]

Self-Reported Impact of Media

As discussed in Chapter Seven, the most rigorous summative evaluations evaluate impact by assessing key outcomes before and after the intervention and/or between exposed and unexposed groups. These experimental or quasi-experimental designs can make causal inferences by observing changes that may be attributable to the intervention. However, such designs are not always feasible. In these cases, evaluations may ask respondents to self-report the impact of an intervention on their attitudes or behaviors. For BBC Media Action governance-related interventions, for example, the key impact measure is the percentage of exposed individuals who state that they believe that the program played a key role in helping them to hold the government to account.[74]

This approach has low validity because, in contrast to a quasi-experimental design, it has no mechanism by which to control for response bias. As discussed earlier, response acquiescence is a particularly significant challenge in DoD operating environments. Thus, one could expect a systemic positive bias. A better approach, at minimal added costs, would be to compare attitudes between those who were exposed and

---

[69] Michael Papa and Arvind Singhal, "How Entertainment-Education Programs Promote Dialogue in Support of Social Change," paper presented at the 58th annual International Communication Association Conference, Montreal, May 22, 2008; Chatterjee et al., 2009, p. 611.

[70] Chatterjee et al., 2009, p. 630.

[71] Coffman, 2002, p. 22.

[72] Gary T. Henry and Craig S. Gordon, "Tracking Issue Attention: Specifying the Dynamics of the Public Agenda," *Public Opinion Quarterly*, Vol. 65, No. 2, 2001.

[73] Coffman, 2002, p. 22.

[74] Author interview with James Deane, May 15, 2013.

those who were not exposed and infer impact based on the differences between those groups, perhaps within propensity matched cohorts. However, self-reported impact measures can provide some information if used in combination with other questions and approaches.

Content Analysis and Social Media Monitoring

This section discusses the use of content analysis, or media monitoring, to assess IIP campaign exposure, influence, and associated changes in attitudes and sentiments. Content analysis involves the systemic observation of traditional press (television, radio, newspaper) and web and social media sources to quantify programs and messages communicated through the media to determine how messages are spreading throughout the target audience. Because media content reflects both the dissemination of and reaction to the campaign, as well as baseline sentiments, it can be used to inform all three phases of evaluation. Within the summative phase, content analysis can be used to measure campaign exposure as well as changes in knowledge, attitudes, and, to some extent, behavior.

Content analysis can include quantitative and qualitative methods to measure the frequency and placement of program material, as well as key words, names, and even narratives that correlate with attitudinal and behavioral outcomes of interest. In this way, content analysis has two broad purposes: to measure ourselves and to measure the audience. Methods associated with content analysis include *traditional press and broadcast media analysis* (television, radio, newspapers, political events, and associated web content), as well as *social media analysis*. Traditional press and broadcast media analysis is considerably more resource intensive than social media analysis but, depending on target-audience characteristics, may generate a more representative sample. Because of the resource requirements, traditional media analysis is typically outsourced to commercial service providers, including Kantar Media, Cision, and Burson-Marsteller.[75]

Depending on the use it is being put to, content analysis must focus on one or both of two issues: the content of interest (e.g., quantity of content and ability to meaningfully categorize it) and the extent to which the sample represents the audience or population of interest. These factors can conflict. For example, social media platforms such as Twitter provide enormous amounts of content that is relatively easy to code. But it is difficult to determine the extent to which the voices generating that content reflect voices within the target audience. Traditional content—newspapers and popular television programs—is more likely to reflect the views of the general population. But the content is less available, harder to analyze, and more neutral in tone.

---

[75] NATO, Joint Analysis Lessons Learned Centre, 2013, pp. 45–46.

Many experts believe that current DoD content analysis capabilities are lacking, in terms of both the tools and DoD's capacity to appropriately apply and interpret them. Media monitoring operations need to be capable of knowing what is being said and where, across the media spectrum—from conventional to social media to political dialogues to chat rooms.[76] This is not merely a technical deficiency, however, and new "widgets and gizmos and gadgets" will not "crank out the magic answer" if the analysts lack a qualitative appreciation for the target audience and how it engages media. With this caveat in mind, IIP programs should invest in the development, acquisition, and application of more-sophisticated techniques that can, when triangulated with qualitative methods, rapidly leverage and make meaning out of the data generated by traditional and social media sources.

Content Analysis with Natural Language Processing: Sentiment Analysis and Beyond

Automated sentiment analysis—also known as "tonality scoring" and "opinion mining"—is an analytic technique using *natural language processing* to extract the sentiment or tone associated with a particular topic or audience from a variety of content sources. Natural language processing can be used to measure several important constructs along the hierarchy of behavioral change, including awareness, attitudes toward and perceptions of friendly forces, perceptions and resonance of adversaries and adversary institutions, issue saliency, media frames, norms, and related cognitive processes, like integrative complexity. These techniques could improve the precision, usefulness, and efficiency of current DoD content analysis methods, such as those that assess saliency by categorizing all media stories into broad headings—e.g., security, politics, diplomacy, and economics.[77]

Steve Corman at Arizona State University's Center for Strategic Communication is currently leading a project to develop a sophisticated content analysis tool capable of isolating, analyzing, and tracking over time the narratives that extremists use in their public statements and blog posts.[78] Once validated, this tool could contribute to the formative and summative evaluation phases by improving DoD's capacity to characterize the information environment and understand how the campaign is influencing the delivery and diffusion of narratives throughout the target audience.

Related techniques can extract the frames used by the media and the target audience to rationalize or explain arguments and concepts. Frames can be used to estimate

---

[76] Author interview with Simon Haselock, June 2013.

[77] Thomas M. Cioppa, "Operation Iraqi Freedom Strategic Communication Analysis and Assessment," *Media, War, and Conflict*, Vol. 2, No. 1, April 2009.

[78] Author interview with Steve Corman, March 2013.

the impact of a campaign by assessing whether the target audience adopts the frames used by the intervention.[79]

Integrative complexity is a measure that assesses the intellectual style used in processing information and decisionmaking. It has two components: *differentiation*, "the perception of different dimensions when considering an issue," and *integration*, the "recognition of cognitive connections among differentiated dimensions or perspectives." Complexity can be scored from most traditional and web-based content.[80] IIP researchers may be interested in this measure because it is correlated with political ideology and political group status. Groups show decreases in complexity immediately prior to surprise attacks.[81]

Experts offered several suggestions for applying quantitative content analysis to IIP evaluation.

- Quantitative content analysis should be paired with qualitative analysis to properly interpret quantitative results and decipher linguistic nuances. For example, the number of times a topic is mentioned is insufficient without pairing those mentions with qualitative analysis of the accuracy of the information and how it is placed within a particular media text or related set of materials.[82] Qualitative analysis is also needed to pick up on sarcasm, jokes, and other linguistic nuances that software-based tools frequently miss.[83] The "human element," said Chris Scully, "is irreplaceable."[84]
- Valid and reliable content analysis requires skilled coders and detailed coding sheets that are developed collaboratively with the local research firm and program managers. When human coding is required, local coders must be well trained, and categories must be defined in simple terms, as some cultures are less detail-oriented.[85]
- Content analysis should be done in the original language and then translated and back-translated to check for errors. Content analysis of content that is translated prior to analysis has been shown to be less valid.[86]

---

[79] Author interview with Maureen Taylor, April 4, 2013.

[80] Peter Suedfeld, "Suedfeld's Integrative Complexity Research," web page, last updated June 7, 2004.

[81] Peter Suedfeld and Susan Bluck, "Changes in Integrative Complexity Prior to Surprise Attacks," *Journal of Conflict Resolution*, Vol. 32, No. 4, December 1988.

[82] Author interview with Amelia Arsenault, February 14, 2013.

[83] Author interview with Geeta Patel, May 21, 2013.

[84] Author interview with Chris Scully, April 18, 2013.

[85] Author interview with Maureen Taylor, April 4, 2013.

[86] Author interview with Amelia Arsenault, February 14, 2013.

- For traditional media, either analyze everything or use a random sample. If analyzing everything is not feasible, decide on a subset by format (e.g., 10 percent of all news broadcasts) and use a random-number generator to select the sample.[87]
- For digital content, consider using *adequately representative samples*. Adequately representative samples involve systematically scraping the web and then analyzing a random sample of the material. The proportion sampled varies by language and population but follows the principle that a larger population requires a proportionately smaller sample, and a small population requires a proportionately larger sample to be representative.[88]

Related to sentiment analysis is the analysis of Google search terms over time and by location using Google Trends. The platform enables users to see the frequency with which other users have searched for words or phrases over time and by region, including subregions and cities. While not a representative sample, Google Trends can allow programmers to quickly gauge awareness of an issue or campaign.

Social Media Monitoring for Measuring Influence

Social media monitoring is an efficient way to assess the influence of messages and messengers for two main reasons. First, as with web analytics, it is inexpensive because data collection is built into the dissemination itself. As Olivier Blanchard observes, "If you can use [social media] channels to spread content and increase reach, you can also use them to seek feedback, measure it, analyze it, and make course adjustments as needed."[89] Second, social media monitoring can provide rich data for constructing psychographic and sociographic profiles of influencers and audience members. While social media may not be applicable in some contemporary DoD operating environments, it is in others, and is likely to be increasingly relevant in the future.

For the purposes of IIP evaluation, social media data can serve two broad purposes. First, they can be used to assess the reach and appeal of DoD social media messages in ways similar to the web analytics discussed in the preceding section on measuring exposure. Second, social media content can be used to assess attitudes, perceptions, and other cognitive processes within the target audience by analyzing the content or reach of other influential or popular social media messages that might reflect the influence of DoD messages. For example, the Common Operational Research Environment (CORE) lab at the Naval Postgraduate School has developed a tool for dynamic tweet analysis that enables programmers working with technically sophisticated audiences to determine if their themes and messages are resonating in real time.[90] Research-

---

[87] Author interview with Maureen Taylor, April 4, 2013.

[88] Author interview on a not-for-attribution basis, July 31, 2013.

[89] Blanchard, 2011, p. 193.

[90] Author interview on a not-for-attribution basis, May 2, 2013.

ers at the University of Vermont have developed a tool known as the Hedonometer that could inform IIP assessment.[91] The tool measures expressions of life satisfaction over time and across locations by analyzing geo-tagged Twitter data.[92] DoD IIP units may also track views and shares associated with videos and other content produced by al Qaeda and affiliate organizations.[93]

A central challenge in extracting meaningful information from social media data is finding the signal in the noise (e.g., determining which data are important to analyze and how to control for bias). Experts from the public relations sector stress that the key is adopting the right advanced analytics and visualization tools to translate big data into insights that can inform decisionmaking.

Experts agree that data should be collected across many platforms—Twitter, Instagram, LinkedIn, Facebook, Pinterest, and so forth. The Barcelona Declaration of Measurement Principles characterized social media as a "discipline, not a tool," and stressed that "there is no signal metric."[94]

Social media metrics commonly tracked include the *number of fans or followers* over time; the *quality of fans and followers*, in terms of their engagement with the platform; the quantity and content of comments; the quantity and depth of *social interactions*, including "shares" of content; and the *performance* of social content, including likes and analyzing content to determine if opinions have changed over time.[95] Several off-the-shelf tools are available that score the influence of messages and individuals. Klout is a particularly valuable indicator of influence that takes into account behavior across all of the major social media platforms.[96] Other social media monitoring options include Google Insights, Google Analytics, Radian6, and Hootsuite.

The volume and velocity of social media data are exciting, but these "social listening" tools often fail to generate a representative sample of target-audience characteristics. First, the sample is restricted to the portion of the target audience that participates in social media. Second, comments are often biased, because people who comment tend to have extreme views (e.g., cult followings). For example, the *Twilight* film had a much higher social listening score than *The Hunger Games*, though the opening of *The Hunger Games* was twice as big.[97]

Gina Faranda, deputy director of the Office of Opinion Research at the U.S. Department of State, notes that social media monitoring tools need a mechanism for

---

[91] Hedonometer, homepage, undated.

[92] Caitlin Dewey, "Measuring Happiness Through Twitter," *Washington Post*, May 6, 2013.

[93] Author interview on a not-for-attribution basis, July 18, 2013.

[94] "Barcelona Declaration of Measurement Principles," 2010.

[95] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 42.

[96] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 42.

[97] Author interview with Vincent Bruzzese, June 7, 2013.

sifting through the "noise of a particularly vocal minority," and evaluators need to be careful to not attribute opinions to a large population that does not necessarily share the minority's viewpoints.[98]

As with web analytics, social media monitoring should avoid overemphasizing vanity metrics that cannot drive decisions or meaningfully gauge performance on key outcomes. "Ghost followers" who click "like" or "follow" but do not actively engage are "empty numbers," according to Blanchard: "Forgetting to tie the easy numbers to something of substance can send your program down the wrong measurement path."[99]

Measuring Observed Changes in Individual and Group Behavior and Contributions to Strategic Objectives

The previous section discussed self-report measures of attitudes and other predictors of behavioral change. This section discusses directly observable data sources that can be used to measure the influence of messages and associated changes in target-audience attitudes and behaviors. Data on behaviors are difficult to collect in a representative fashion. Nonetheless, the most valid and useful IIP assessments include measures of how the population actually behaves, which complement and validate self-report measures. These include observations of the desired behavior (e.g., voter turnout or surrenders), atmospheric indicators of attitudes and sentiments, data on the achievement of the military or political objectives (e.g., changes in casualties, violence, recruitment, economic growth), and direct or indirect behavioral responses to the campaign (e.g., countercampaigns, calling an 800 number).

Observing Desired Behaviors and Achievement of Influence Objectives

IIP assessment should measure changes in the behavior targeted by the influence objective. For example, if the influence objective is to increase voter turnout, the assessment should measure voter turnout. If the objective is to mislead enemy decisionmaking, the assessment should be capable of capturing the enemy's choices on that decision. If the objective is to increase surrenders, surrenders should be tracked over time. Cullin recalled an instance in which a series of surrenders immediately followed an IIP campaign that disseminated pamphlets telling enemy forces that they "had the option to surrender."[100] As explained in Chapter Seven, however, demonstrating a causal role of the program in spurring these observed behavioral changes requires that rival explanations be formally excluded.

---

[98] Author interview with Gina Faranda, June 13, 2013.

[99] Blanchard, 2011, p. 195.

[100] Author interview with Brian Cullin, February 25, 2013.

When the behavior cannot be observed systematically or aggregately, researchers can use the *participant observation* technique, in which a sample of the target audience is deliberately observed. In health communication evaluations, for example, researchers often use cultural anthropologists to do in-home observations of health behaviors.[101] In the DoD context, "secret shoppers" have been used; local Afghan volunteers observe the behavior of a group targeted by the campaign and report their experiences.[102] This technique has also been leveraged for measuring the impact of exchange programs. The USNORTHCOM influence assessment capability, for example, would assign a participant observer to every exchange. This participant observer was responsible for assessing whether the objectives had been met.[103]

The validity of participant observation is limited by several factors. First, the observer or rater may be biased due to pressures to show program effects. Second, the observer effect biases how the subjects behave when under observation, which is amplified in the case of an armed observer. Third, it is difficult to prove that the sample being observed is representative of the target audience.

Often, the behavior of interest cannot be directly observed, but other behaviors that can be observed are validated proxies or predictors of the behavior of interest. Behavioroid measures reflect the intent to engage in the behavior of interest by measuring behaviors that predict or correlate with the unobserved behavior. Pratkanis provided an example from a phone survey that assessed a campaign to reduce senior citizens' vulnerability to fraud. The researchers could not deliberately scam the respondents to measure their likelihood to fall prey to a fraud, but the researchers knew that the longer the seniors remained on the phone with a person trying to scam them, the more likely they were to fall prey to the fraud. Another behavioroid measure used was whether or not the respondent agreed to a follow-on call or agreed to have something sent to them.[104]

Direct and Indirect Response Tracking

In some cases, behaviors can be observed that directly or indirectly gauge the influence of the program, because the behaviors can only be reasonably explained by the fact that the audience was exposed to the program. In evaluation research this method is often called *direct response tracking*. For example, a social marketing ad may ask a viewer to undertake a direct and measurable response, such as calling an 800 number or visiting a website. Depending on the method of response, these measures can provide research-

---

[101] Author interview with Ronald Rice, May 9, 2013; interview with Charlotte Cole, May 29, 2013.

[102] Author interview with John-Paul Gravelines, June 13, 2013.

[103] Author interview with LTC Scott Nelson, October 10, 2013.

[104] Author interview with Anthony Pratkanis, March 26, 2013. A good reference for developing behavioroid measures can be found in Gardner Lindzey and Elliot Aronson, eds., *The Handbook of Social Psychology: Research Methods*, Vol. 2, 2nd ed., Reading, Mass.: Addison-Wesley, 1968.

ers with additional demographic or psychographic information on the responders.[105] *These are often weak indicators of effects, however, unless research has demonstrated a strong correlation between engaging in the direct response and adopting the desired behavioral change.* To strengthen this approach, some evaluations will contact the direct responders for a follow-up evaluation to determine whether and how the information they received shaped their behavior.[106]

There are several analogs to direct response tracking that may be employed to measure the reach and influence of DoD IIP messages.

- *The quantity and quality of intelligence tips given by the local population.* Tracking intelligence tips can serve as both a general atmospheric indicator of the extent to which the coalition narrative is winning and, depending on the message and medium by which the tip is submitted, a direct response indicator of message effects.[107] For example, the message may ask the population to submit information on known or suspected facilities producing improvised explosive devices by calling a number.[108]
- *The existence of a countercampaign.* While not a traditional direct response measure, the existence of a countercampaign initiated by the adversary is a strong indicator that the message is resonating with the right audiences.[109]
- *Information lines or hotlines.* A campaign may try to reduce insurgency recruitment by providing information to the family members of potential recruits about the dangers and by encouraging family members to call a hotline.

Atmospherics and Observable Indicators of Attitudes and Sentiments

If collected and analyzed systematically and rigorously, atmospherics and associated measures can provide more-robust estimates of sentiment than self-reported survey data.[110] Several SMEs felt that atmospherics are currently underutilized in IIP evaluation. However, atmospherics is poorly defined, and associated data sources and collection mechanisms are ad hoc and anecdotal. This section discusses the application of atmospherics to IIP evaluation and provides several suggestions for improving the processes for deciding what to collect and systematizing the collection and analysis processes.

---

[105] Schneider and Cheslock, 2003.

[106] Coffman, 2002, p. 15.

[107] Author interview with Jonathan Schroden, November 12, 2013; interview with Anthony Pratkanis, March 26, 2013; interview with Mark Helmke, May 6, 2013.

[108] Author interview with Steve Booth-Butterfield, January 7, 2013.

[109] Author interview with Anthony Pratkanis, March 26, 2013.

[110] Author interview with Anthony Pratkanis, March 26, 2013.

Characterizing Atmospherics in Current DoD Practices

Atmospherics is a poorly defined but commonly used term by DoD assessment practitioners. The May 2011 version of DoD Directive 3600.01 defines *atmospherics* as "information regarding the surrounding or pervading mood, environment or influence on a given population." It goes on to clarify atmospherics as a "human-derived information gathering activity" that is distinct from HUMINT and can include "polling, surveys, opinion research, spot reports, and consolidation of other information relevant to prevailing moods, attitudes and influences among a population."[111] However, the most recent version of DoD Directive 3600.01 (May 2013) includes no mention of atmospherics.

Informally, *atmospherics* refers to a range of observable indicators that are used or could be used to characterize the prevailing mood or atmosphere of the target audience. It is distinguished from large surveys or formal opinion polling research. Atmospherics can gauge sentiments toward U.S. or friendly forces as well as trust in public institutions and perceptions of security. Examples include

- how the population responds to patrol vehicles rolling through villages (e.g., throwing stones or cheering)
- the extent to which the population engages with friendly forces (e.g., eye contact, exchanging information, letting friendly forces in the door)
- the number of people shopping at the bazaar or the traffic on a road used to go to a market[112]
- people's willingness to leave their homes in the absence of ISAF forces[113]
- number of families sending girls to school or allowing their children to be vaccinated[114]
- the number of intelligence tips given to friendly forces by the target audience[115]
- observable indicators of an adversary influence, such as the attendance of a particular mullah at a mosque and reactions among the target audience[116]
- using local volunteers to serve as secret shoppers observing the behavior of a group or process that the coalition cannot directly observe[117]
- subjective assessment of the mood from trusted local sources through informal interviews.[118]

---

[111] U.S. Department of Defense Directive 3600.01, 2013.

[112] Author interview with Jonathan Schroden, November 12, 2013.

[113] Author interview with John-Paul Gravelines, June 13, 2013.

[114] Author interview with Victoria Romero, June 24, 2013.

[115] Author interview with Anthony Pratkanis, March 26, 2013; interview with Jonathan Schroden, November 12, 2013.

[116] Author interview with Anthony Pratkanis, March 26, 2013.

[117] Author interview with John-Paul Gravelines, June 13, 2013.

[118] Author interview on a not-for-attribution basis, December 15, 2013.

Data for these measures can come from a variety of sources, including direct observation, as reported in after-action reports, and analysis of footage from intelligence, surveillance, and reconnaissance (ISR) assets or news broadcasts.[119] In the field, *atmospherics* is often defined as data gathered from informal interviews or surveys with trusted local sources or confidants. However, this report discusses those techniques separately. See Chapter Eight for a discussion of in-depth and intercept interviews.

Selecting Valid and Useful Atmospheric Measures and Data Sources

Because there are a nearly infinite number of possible atmospheric indicators, a central challenge with atmospherics is determining what data are essential to collect and analyze, or finding the signal in the noise. The key, according to Pratkanis, "is coupling those atmospheric measures to objectives."[120] Doing so requires a sophisticated understanding of the cultural context so that evaluators can reliably interpret the meaning behind what they're observing.[121] Researchers should consider using empirical analysis and the Delphi process to determine which atmospheric variables are worth capturing.

Christopher Nelson encouraged evaluators to use a Delphi or e-Delphi process involving SMEs and experienced practitioners. Through reviewing the literature and surveying practitioners, researchers could develop a list of the top 30 observable atmospheric variables. To cut this list down to the top five to ten indicators, they could then circulate the list of candidate variables to SMEs and experienced practitioners and ask them to rank them until they have 80- or 90-percent convergence.[122]

Steve Booth-Butterfield suggested that atmospheric variables be developed and validated through empirical analysis. For example, researchers could compare atmospherics in a known hostile area with a known friendly area and see whether there are significant differences in the atmospheric measures when controlling for other predictors. Even if the data are incomplete or questionable, "if it's telling bad news in Anbar and good news somewhere else, . . . it's a measure of validity."[123] Alternatively, researchers could compare the performance of candidate atmospheric measures over time in an area where progress was known to have occurred. Assessors should be creative when generating candidate atmospheric measures.[124]

John Matel, a State Department public affairs officer, recalls his use of the "banana index" for measuring perceptions of safety:

---

[119] Author interview with John-Paul Gravelines, June 13, 2013.

[120] Author interview with Anthony Pratkanis, March 26, 2013.

[121] Author interview with LTC Scott Nelson, October 10, 2013.

[122] Author interview with Christopher Nelson, February 18, 2013.

[123] Author interview with Steve Booth-Butterfield, January 7, 2013; Booth-Butterfield, undated.

[124] Author interview with Robert Banks, March 25, 2013.

Bananas . . . have to be imported from somewhere else. It is very hard to get a banana to market exactly at the right time. They will usually be either green or brown. A banana stays yellow for only a short time and if it is mishandled it gets easily bruised. If you see lots of good quality bananas in the market, you know that the distribution system is working reasonably well and that goods are moving expeditiously through the marketplace.[125]

While standardization is important, atmospheric measures and data collection strategies also must be flexible enough so that they can be tailored to the local information environment and security context. Every village is different, and indicators will have different meanings depending on the context.

Improving Atmospheric Data Collection

Systematizing and institutionalizing the collection and analysis of valid and meaningful atmospherics was flagged by several experts as a priority area for improvement.[126] Rigorous atmospherics on meaningful variables are tremendously valuable and overcome many of the limitations to self-report data. Unfortunately, the ad hoc and anecdotal nature of existing sources limits the validity and usefulness of atmospherics in the overall assessment process. The adage that "every soldier is a sensor" only applies if data from these sensors are captured and properly synthesized.

Experts and practitioners provided examples of mechanisms that could improve the collection and application of atmospherics.

- *Recorders in patrol vehicles coupled with text recognition and data mining software:* Reactions by local populations could be captured through the continuous use of GPS-enabled recorders in patrol vehicles. The recordings could be analyzed with text recognition and natural language processing to score the level of hostility or support among the population. Coupled with geocoding, these data could paint a detailed picture of local sentiments over time and across areas. This approach minimizes the burden on troops and would be relatively inexpensive due to minimal manpower costs.[127] However, given the large amount of data this would generate, advanced machine learning tools may be required to "sift through the noise."[128]
- *Mandatory after-action reports, patrol reports, and debriefs capturing atmospherics:* Units returning from patrols should be routinely debriefed to capture their per-

[125] John Matel, "Hidden Prosperity and the Banana Index in Iraq," blog post, *DipNote*, April 8, 2008.

[126] Author interview with Simon Haselock, June 2013; interview with Jonathan Schroden, November 12, 2013.

[127] Author interview with a former employee of a large IO evaluation contractor, February 25, 2013.

[128] Author interview with LTC Scott Nelson, October 10, 2013.

ceptions of local reactions to the patrol, the density of populations gathering in public spaces, and other validated indicators of population sentiments.[129]

- *Validated checklists of key atmospherics for units on patrol:* Pratkanis pointed out that units "have other things on their mind." To facilitate the collection of the right atmospherics, units should be given lists of the top five to ten indicators that they need to be tracking while patrolling a village, for example.[130]
- *Secure the use of ISR assets to collect atmospherics:* Footage from ISR assets provides a rich source of data for atmospherics. Unfortunately, it is often difficult for IIP units to access such assets for IIP evaluation purposes, because the asset "owners" have a kinetic focus.[131]
- *Improve training and doctrine:* There is little to no doctrine or formal training associated with atmospherics for IIP campaigns. As a consequence, existing efforts are ad hoc and difficult to synthesize.[132]
- *Deconflict and leverage intelligence products and sources:* There is significant overlap between atmospherics and HUMINT. Available HUMINT in response to key IIP assessment questions should be leveraged prior to new atmospheric collection efforts.

Even if systematically collected, atmospherics have several limitations. First, force protection issues create challenges for consistently observing target-audience behavior, as there may be too great a distance between military operators and locals.[133] Second, assessors should also be cautious when generalizing, as it is difficult to determine whether the population observed is representative of the target audience. Finally, depending on the collection method, the observer or armed observe effect may alter the behavior of the target population.

Aggregate or Campaign-Level Data on Military and Political End States

Another directly observed data source is aggregate data reflecting the extent to which military or political objectives are being achieved. IIP activities should, if the logic model is valid, contribute to the achievement of military and political strategic objectives and end states. For example, if the IIP MOPs suggest that the influence program is working but other indicators suggest that violence is increasing and that the coalition-supported government is losing legitimacy, IIP planners should revisit the logic model and inspect the validity and reliability of their MOPs and MOEs. To

---

[129] Author interview with Simon Haselock, June 2013; interview with a former employee of a large IO evaluation contractor, February 25, 2013.

[130] Author interview with Anthony Pratkanis, March 26, 2013.

[131] Author interview with John-Paul Gravelines, June 13, 2013.

[132] Author interview with a former employee of a large IO evaluation contractor, February 25, 2013.

[133] Author interview with Simon Haselock, June 2013.

track the achievement of broader military and political objectives, IIP assessors should track casualties, recruitment, levels of violence (e.g., SIGACTS, hospital discharges), surrenders, and economic and governance indicators within their area of operations. The World Bank Worldwide Governance Indicators, which combine data from 31 surveys, provide publicly available measures of governance over time in six dimensions (voice and accountability, political stability and lack of violence, effectiveness, regulatory quality, rule of law, and control over corruption).[134]

Embedding Behavioral Measures in Survey Instruments

Behavioral measures attempt to measure how people actually behave (revealed preferences), in addition to their stated preferences during the administration of surveys, by testing how the participant responds to certain scenarios or prompts introduced by the researcher. This technique was introduced in Chapter Seven, Box 7.3, when discussing the field experiment in Ghana to test the effects of partisan radio on citizens riding in public transportation. After listening to one of four randomly selected options (a partisan radio station supporting the government, a partisan station supporting the opposition, a neutral political talk show, or nothing), participants were asked a series of questions, including several behavioral measures. Behavioral measures used by the researchers included: (1) giving the participants money for participating and then asking them to donate a portion of that money to a cause associated with one side or the other of the partisan split; (2) giving them a choice of key chains, each associated with a different party or the government; and (3) asking them to join a petition about transportation policy by texting a number, which would measure political efficacy and engagement.[135] These behavioral measures provide an innovative and cost-effective technique for addressing the bias inherent in self-report attitudinal measures when gauging IIP effects.

Techniques and Tips for Measuring Effects That Are Long-Term or Inherently Difficult to Observe

The measures and methods discussed in the previous section assume that the outcome has occurred and is observable. However, it is not always the case that the outcome of interest has occurred by the time the assessment must be conducted.

A core challenge in IIP assessment is balancing near-term assessment and reporting requirements with the strategic imperative to focus on long-term change processes that meaningfully and sustainably shape the information environment. On one hand, behavioral change is a phased process that requires time. The most-effective inter-

---

[134]World Bank Group, *Worldwide Governance Indicators*, online database, undated.

[135]Author interview with Devra Moehler, May 31, 2013.

ventions are those that are sustained over time and focus on long-term behavioral change.[136] Student exchange programs, for example, can take decades to show impact, as the largest effects are those that occur when the students assume influential leadership positions later in their careers.[137] On the other hand, "Congress has a short-term perspective." Because budget allocations occur annually, programs have to prove behavioral change over a period of months rather than years.[138] And without near-term or intermediate measures of effect, there is little basis to shut down or redesign failing programs.[139]

The best way to balance these twin objectives is to develop and field "leading indicators, or near-term predictors of long-term effects." These measures can be identified from the logic model and associated theories of persuasion, behavioral change, and diffusion (see Chapters Three and Five).[140] The extent to which these measures predict long-term effects can be validated through formative and empirical research. The *Information Environment Assessment Handbook* calls this a "time-phased process" and instructs assessors to separate the campaign into "manageable segments."[141] It is important, however, that these near-term measures not incentivize "teaching to the test" and that they not divert attention from long-term goals. Professor James Pamment is concerned that the emphasis on annual assessment reports has fundamentally changed the priorities of the British Council. The council "used to work on five-, ten-, 15-year time frames. . . . They were focused on generational change, . . . but in the last few years, their annual reports have become their top priority."[142]

Those responsible for evaluating the effectiveness of long-term influence activities commonly find themselves *wishing* that data had been collected historically and over time. To facilitate future longitudinal evaluations, IIP programs need to collect consistent data over time on a broad range of input, output, and outcome variables. For example, exchange programs should maintain detailed records of participants and engagements. Retrospectively collecting or estimating who was engaged and when is expensive and difficult.[143] Because organizations, priorities, and evaluation research questions change over time, it is important to collect data on a wide range of variables

---

[136] Author interview with Joie Acosta, March 20, 2013; author interview on a not-for-attribution basis, July 30, 2013.

[137] Author interview with Julianne Paunescu, June 20, 2013.

[138] Author interviews on a not-for-attribution basis, December 5, 2012, and July 30, 2013.

[139] Author interview on a not-for-attribution basis, July 30, 2013.

[140] Author interview with Craig Hayden, June 21, 2013; interview with Devra Moehler, May 31, 2013.

[141] The Initiatives Group, 2013, p. 21.

[142] Author interview with James Pamment, May 24, 2013.

[143] Author interview with James Pamment, May 24, 2013.

that may be relevant to future generations of decisionmakers.[144] Collecting data over long periods of time is also beneficial because it allows researchers to identify aberrant or unusual waves of data that might suggest cheating or other errors affecting the data collection process.[145]

Analyses and Modeling in Influence Outcome and Impact Evaluation

This section discusses select insights, concepts, and best practices associated with analyses in support of IIP assessment in conflict areas. It is by no means an exhaustive treatment of the subject. Assessors are encouraged to review texts on statistical analysis for social and behavioral sciences, including books by Joseph Healey and James Paul Stevens.[146] In his chapter on statistical analysis, Thomas Valente provides an excellent summary of the major statistical analysis conducted in support of communication campaign evaluation.[147] Readers should also review the sections in Chapter Ten on analysis and interpretation of survey data, which address concepts like margins of error.

Prioritize Data Collection over Modeling and Statistical Analysis Tools

While it is important to be familiar with basic analytical techniques, that is not where assessors should be principally focused. For several reasons, the quality and quantity of data are far more important than the statistical technique for analyzing or modeling them. First, you cannot take advantage of data analysis tools if you do not have anything to analyze. LTC Scott Nelson sees this as a challenge with current IO assessment guidance: "The assessment guidance emphasizes modeling, but the validated data simply do not exist in large enough quantities to put these models to use. . . . [Assessors] need to take a step back and dive into the data generating process."[148] Because of data limitations, analytical methods are typically very restricted to the monitoring and evaluation field. Analysts are "primarily using means and percentages and simple t-tests" and are "not even doing logistic regression," because "the data are too bad to make those techniques viable."[149]

Moreover, an overreliance on new assessment "widgets and gizmos and gadgets" may hinder effective assessment by distracting attention and resources from more-

---

[144] Author interview with James Pamment, May 24, 2013.

[145] Author interview with Katherine Brown, March 4, 2013.

[146] Joseph F. Healey, *Statistics: A Tool for Social Research*, 9th ed., Belmont, Calif.: Wadsworth/Cengage Learning, 2012; James Paul Stevens, *Applied Multivariate Statistics for the Social Sciences*, 5th ed., New York: Routledge, 2009.

[147] Valente, 2002, pp. 163–180.

[148] Author interview with LTC Scott Nelson, October 10, 2013.

[149] Author interview with Maureen Taylor, April 4, 2013.

important challenges and by instilling a false sense of confidence that software can solve the complex problems inherent in IIP assessment. As one SME explained, "People make a good living building tools, . . . and DoD likes tools," but the tools fielded for assessment mislead assessors into thinking that they can "put this information in and crank it through and get the magic answer."[150]

Analysts should use the simplest analytic method that is appropriate. Douglas Hubbard punctuates this point with a rhetorical question: "Are you trying to get published in a peer-reviewed journal, or are you just trying to reduce your uncertainty?"[151] InterMedia's policy is to use sophisticated analytic techniques, such as structured equation modeling, only when it is working with a client that can appreciate, understand, and interrogate analyses derived from those techniques.[152]

The Perils of Overquantification and Junk Arithmetic

Often, the arithmetic applied to measures is inappropriate to the nature of the assessment data. Stephen Downes-Martin calls this "junk arithmetic." For example, many of the assessments he observed in theater tried to average ordinal values (ordered or ranked numbers where the distances between ranks are not necessarily the same). Because those codes are not ratio-scale numbers, "by the laws of mathematics . . . functions such as averaging cannot be performed on them. . . . It is nonsensical."[153] Because of this tendency, he encourages analysts to push back on calls for a quantitative metric: "You need to ask what mathematical calculations that metric will be subject to."

Moreover, assessments should express results in the form of qualitative statements about trends and movements toward end states. Using numerical scales to report progress is often unhelpful and distracting to decisionmakers. It is "extremely unhelpful for an information consumer to get hung up on why an assessment is a 2 as opposed to a 3, something forgotten by organizations that operate on ratings such as 3.24."[154]

Aggregation Across Areas, Commands, and Methods

A core component to operations assessment is aggregating assessments across areas and up through hierarchal layers of command structure. This section discusses aggregation best practices as they relate to analysis. More on aggregation can be found in Chapter Eleven (in the section "Aggregated Data"). Principles and best practices for aggregation endorsed by SMEs included the following:

---

[150] Author interview on a not-for-attribution basis, July 30, 2013.

[151] Hubbard, 2010, p. 35.

[152] Author interview with Gerry Power, June 2013.

[153] Downes-Martin, 2011, p. 109.

[154] Upshur, Roginski, and Kilcullen, 2012.

- *DoD needs validated and documented models for aggregating assessments vertically and horizontally.* Current aggregation approaches, centering mainly on color coding, are ad hoc at best and damaging at worst. Downes-Martin recalls seeing some regional commands "color averaging" as an aggregation technique.[155]
- *Aggregation requires consistent measurement over time and across areas. Consistent, mediocre assessments are better than great, inconsistence assessments,* because the latter cannot be aggregated or used to inform trends over time.[156] Consistency in measurement is undermined by turnover and "type A" leadership that wants to entirely revise the process. Leadership—and others driving the design of assessments—needs to be more willing to inherit assessment practices that are "good enough" to preserve consistency.[157]
- *It is not always possible or desirable to aggregate the same metrics across different sites or levels of command.* If the theory of victory is different at the national or regional level, the metrics have a different meaning. As one SME put it, "The whole may be more than the sum of its parts, and aggregation may not answer the mail at the higher level of analysis."[158]
- *MOEs should be weighted.* Analysts should determine the relative value of an MOE to the overall assessment and assign weights accordingly.[159]
- *Identify measures from mixed data sources that are trending together.* Because there are limitations to each measurement approach, the most-valid measures of success are those that converge across multiple qualitative and quantitative data items.[160]
- *The best evaluations triangulate many measures from different methods and data sources. Use many methods and have a single-point synthesis.* To synthesize the disparate results from a mixed-method approach, one person or group who is familiar with and has the power to affect the whole assessment process should be responsible for triangulating disparate approaches.[161]

Narrative as a Method for Analysis or Aggregation

One way to make sense of disparate data or to aggregate across programs, activities, and analyses of different types is to tell a compelling story. This method of analysis and aggregation is referred to as a *narrative approach* and has been strongly advocated for

---

[155] Downes-Martin, 2011, p. 112.

[156] Author interview on a not-for-attribution basis, December 15, 2013.

[157] Author interview on a not-for-attribution basis, December 15, 2013.

[158] Author interview on a not-for-attribution basis, March 1, 2013.

[159] The Initiatives Group, 2013, p. 18.

[160] Author interview with Steve Booth-Butterfield, January 7, 2013.

[161] Author interview with Steve Booth-Butterfield, January 7, 2013.

aggregate campaign- and operational-level assessments by our RAND colleague Ben Connable.[162] Compiling information into a narrative can be viewed as a sort of holistic triangulation, interpreting all available data and making a compelling argument for their interpretation.

Such analyses can be quite useful, but they are vulnerable in several respects. First, like all assessments, where underlying data are suspect, resulting narratives can be suspect. Of course, if the analyst/narrator is aware of weaknesses in the underlying data, that can become part of the narrative and thus an analytic strength. Second, like self-assessment of any kind, narratives are vulnerable to bias and overoptimism (see the discussion related to expert elicitation in Chapter Eight).

If a narrative analysis is conducted within the context of an explicit theory of change/logic of the effort, it can be an important contribution to assessment. For a narrative to have such a connection, it need not ever say "theory of change," but it must make a clear statement about how the various operations and activities being analyzed are supposed to connect to desired end states, describe progress toward those end states, and offer an explanation of any shortfalls in progress toward those expected end states.

For assessment, narratives offer an array of advantages, including: They allow variations and nuances across the area of operations to be captured and appreciated; they remind people of the context and complexity of the operation; they force assessors to think through issues and ensure that their assessment is based on rigorous thought; and they are the only way to ensure that a proper balance is struck between quantitative and qualitative information, analysis and judgment, and empirical and anecdotal evidence.[163] See the additional discussion of narrative as a means of presentation of assessment in Chapter Eleven.

Analyze Trends over Time

The most valid and useful assessments are those that assess trends over time and across areas.[164] First, trend data are more useful than a snapshot, since IIP progress is defined in terms of change over time. Second, analyzing data over time controls for the biases that limit the validity of the quantitative and qualitative data sources. In essence, it is reasonable to assume that those biases are constant over time. Longitudinal analysis also allows researchers to *verify* assessment data by facilitating the identification of unexpected deviations and *validate* assessment methods by assessing whether results exhibit expected relationships with external events. Trend analysis is also addressed in Chapter Ten in the context of interpreting survey data. The requirement to analyze

---

[162] Ben Connable, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, Santa Monica, Calif.: RAND Corporation, MG-1086-DOD, 2012.

[163] Schroden, 2011, p. 99.

[164] UK Ministry of Defence, 2012, p. 3-28.

trends highlights the importance of consistent measurement and consistent assessment processes over time and through rotations.[165]

Statistical Hypothesis Tests

The principal analytic technique in an evaluation is a statistical hypothesis test. These tests determine whether a relationship exists between the independent variable (representing the IIP intervention) and the dependent variable (representing the desired outcome) by testing the hypothesis that there is no relationship. While there are many possible statistical tests, five techniques "cover 90% of the situations" encountered by evaluators: chi-squared, analysis of variance (ANOVA) *t*-tests and *F*-tests, logistic regression, multinomial logistic regression, and the Pearson correlation coefficient.[166]

The choice of test depends on the nature of the dependent and independent variables. Variables can be continuous (e.g., levels of recruitment, age), ordinal, or categorical. Ordinal variables are ranked but with unknown distances between the rankings (e.g., Likert scales). Categorical variables, also known as "nominal" variables, are unranked and include categories like exposed versus unexposed, gender, marital status, and so forth.

If both the dependent and independent variables are categorical, the Pearson correlation coefficient is appropriate. If they are both noncontinuous (ordinal or categorical), the chi-squared test is appropriate. If the outcome measure is continuous but the intervention variable is categorical or ordinal, the ANOVA *t*-test or ANOVA *F*-test, respectively, is appropriate. If the intervention variable is continuous but the outcome variable is categorical or ordinal, the analysis should use a logistic regression or a multinominal logistic regression. The chi-squared and ANOVA tests are particularly relevant to IIP interventions, because the independent variable is commonly categorical, assuming the value of "exposed" or "unexposed." Valente provides an excellent summary of each of these tests with illustrations of how they have been applied to analyze health communication interventions.[167]

Multivariate Analysis

The previous section discussed techniques for bivariate tests of the relationship between the independent and dependent variable. Evaluations should assess the correlation between the intervention variable and the outcome of interest when accounting or controlling for the simultaneous influence of confounding or alternative explanations (see the sections "Attributes of Good Measures: Validity, Reliability, Feasibility, and Utility" in Chapter Six and "Designing Valid Assessments: The Challenge of Causal Inference in IIP Evaluations" in Chapter Seven).

---

[165] Author interview on a not-for-attribution basis, December 15, 2013.

[166] Valente, 2002, p. 167.

[167] Valente, 2002, pp. 167–177.

If the model is properly specified to account for confounding variables, assessors should have low expectations for observing a statistically significant correlation between the intervention and the outcome of interest. In DoD operating environments, confounding or system-level factors are likely to exert a much higher level of influence on key outcomes than the IIP activity. As Rice explained, "Relative to all other forces impacting Afghani beliefs and behaviors, the communication campaign could be just random noise. . . . Planners should set expectations accordingly."[168] Related to this observation is that the better the model is specified, in terms of accounting for all potential confounds, the less likely it is to show an effect.

Structural Equation Modeling

Structural equation modeling (SEM), a form of multivariate analysis, is a popular analytic technique for both developing a logic model and testing the hypotheses embedded within it. For the relationships and mediators between changes in knowledge, attitudes, and behaviors, Power characterizes SEM as "the most valuable analytic technique for teasing out the relative contributions of the various system-level factors and the intervention."[169] For IIP evaluations, SEM involves mapping out exposure with measures of knowledge acquisition, attitudinal change, behavioral change, and controlling for other covariates. It requires defining variables that correspond to the sequence of steps along the path to behavioral change and establishing the directionality and extent of the direct and indirect relationships between the variables (e.g., mapping the logic model and underlying theories of change in measurable terms).[170] LISREL software is commonly used for SEM analysis.

SEM is particularly valuable when the intervention has to catalyze a sequence of actions or processes arranged in time. For example, if you see a lot more of process A and subsequently a lot more of process B, and your theory of change says that A predicts or causes B, "it gives you much more confidence in relating the communication intervention to the communication outcome it was aimed at." Booth-Butterfield described the structural equation model as the model of causality motivating the intervention: "Even if you don't formally specify the model, everyone has an implied structural equation model." The more detailed and theory based you can make the model, "the more successful you'll be at execution and at measurement." For DoD IIP, the structural equation model is built around quantitative measure of the end states and intermediate objectives, enabling planners to "fill in the blanks between dropping pamphlets and desired behavioral change."[171]

---

[168] Author interview with Ronald Rice, May 9, 2013.

[169] Author interview with Gerry Power, April 10, 2013.

[170] Author interview with Gerry Power, April 10, 2013.

[171] Author interview with Steve Booth-Butterfield, January 7, 2013.

For an illustration of the use of structural equation modeling for IIP evaluation, see InterMedia's report on citizen access to information in Papua New Guinea and the study by Chatterjee and colleagues on the impact of a BBC program on HIV/AIDS-related attitudes and behaviors.[172] In the Papua New Guinea study, researchers used SEM to evaluate the impact of the Mothers Matter campaign on people's attitudes and knowledge of maternal health; the researchers also sought to understand the relationship between household media access and recency of media use, how this relationship influenced exposure to the Mothers Matter campaign, and, in turn, "the impact of people's attitudes and knowledge about women's health during pregnancy."[173] In the evaluation of the HIV/AIDS-awareness campaign, Chatterjee and colleagues used SEM to show that exposure directly influences knowledge acquisition, which directly influences attitudinal change, but the link between attitudes and behavioral change is indirect, mediated by self-efficacy and interpersonal discussion.[174] As these examples demonstrate, SEM is valuable in that it directly feeds back into the development and refinement of the theory of change.

Summary

This chapter reviewed the measures, data collection methods, and analytic techniques used to inform postintervention (process and summative) evaluation of IIP campaigns. Key takeaways included:

- The quality of data is important to assessment, and data on IO programs are often lacking, irrelevant, or not validated. Rather than focusing on modeling, assessment guidance should prioritize equipping assessment teams with the resources and skills needed to generate and validate appropriate data and to recognize where data quality is and is not important.
- *Good data* is not synonymous with *quantitative data*. Depending on the methods and the research question, qualitative data can be more valid, reliable, and useful than quantitative data.
- DoD needs to systematically improve (in terms of extent and consistency) how it documents its own activities and inputs in order to conduct process evaluations and generate data for the independent variable in summative evaluations.
- Good formative research can help determine the relative importance of measuring attitudes versus behaviors, because it identifies the extent to which attitudes predict behaviors.

---

[172] Klara Debeljak and Joe Bonnell, *Citizen Access to Information in Papua New Guinea*, Washington, D.C.: Inter-Media, June 2012; Chatterjee et al., 2009.

[173] Debeljak and Bonnell, 2012, p. 56.

[174] Chatterjee et al., 2009.

- The existence of a countercampaign is a strong indicator of the extent to which a message is resonating.
- Useful examples of atmospheric indicators include the number of people shopping at the bazaar or the traffic on a road used to go to a market;[175] the number of families sending girls to school or allowing their children to be vaccinated;[176] intelligence tips; the "banana index," in which the color of bananas represents the health of the distribution system (in conflict areas, by extension, the security of roads and other shipping routes); and observable indicators of an adversary influence, such as the attendance of a particular mullah at a mosque and reactions among the target audience.[177]
- An overreliance on new "widgets and gizmos and gadgets" for assessment may hinder effective assessment by distracting attention and resources from more-important challenges and by instilling a false sense of confidence that software can solve the complex problems inherent in IIP assessment.
- The arithmetic applied to measures is often inappropriate for the nature of the assessment data. Assessment should not, for example, try to compute averages from ordinal measures.[178]
- Aggregation requires consistent measurement over time and across areas. Consistent, mediocre assessments are better than great, inconsistent assessments. Leadership—and others driving the design of assessments—needs to be more willing to inherit assessment practices that are "good enough" to preserve consistency.

---

[175] Author interview with Jonathan Schroden, November 12, 2013.

[176] Author interview with Victoria Romero, June 24, 2013.

[177] Author interview with Anthony Pratkanis, March 26, 2013.

[178] Downes-Martin, 2011, p. 109.

Surveys and Sampling in IIP Assessment: Best Practices and Challenges

Surveys serve as one of many tools that can be used to collect information for IIP efforts. This chapter provides information regarding different elements that should be considered when developing and administering a survey. It begins with a description of how to determine who should be asked to participate and provides a brief overview of different methods that may be used to collect survey data. It then reviews several considerations for survey instrument design, including question wording and ordering. Following the description of sampling, methods, and instrument design, the chapter moves to a discussion of actions that may be taken to improve data quality and then describes data management considerations. Taken together, the elements in this chapter can assist an IIP assessment planner in the design of high-quality surveys that produce informative results.

Survey Research: Essential but Challenging

Survey research is a useful and efficient method for gathering information regarding the traits, attributes, opinions, and behaviors of people.[1] Surveys can serve multiple purposes. They can be used as part of efforts to describe the characteristics of a population, to explain why people hold certain attitudes or behave in certain ways, and to explore the elements that exist in a certain social context.[2] They can serve as a valuable tool for IIP efforts by providing needed information regarding a population of interest or permitting measurement of the effects (or lack of effect) of an implemented program.

However, surveys are not without limitations, and various sources of error can hinder the collection of reasonably accurate information. For example, error can arise from badly designed survey items, poorly translated surveys, and surveys that have been administered incorrectly by research personnel (i.e., mistakes or cheating during

---

[1] Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian, *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd ed., Hoboken, N.J.: John Wiley and Sons, 2009.

[2] Earl Babbie, *Survey Research Methods*, 2nd ed., Belmont, Calif.: Wadsworth Publishing Company, 1990.

administration).[3] Many surveys seek to better understand a population (individuals of interest in a particular effort), and another source of error can be the collection of survey data from a particular sample, or a portion of the population, that does not adequately represent the whole population of interest.

While these sources of error can arise during any effort, survey research in a conflict environment may make some especially likely. For example, in conflict environments, collection of data from an unrepresentative sample may occur due to lack of information regarding the population (e.g., no credible census), limited access to people in particularly difficult to reach areas, and an inability to request participation from certain individuals, such as women or those who are not the head of a household.[4] Despite the potential difficulties in addressing sources of error in a conflict environment, surveys continue to be used, in part, because they provide information that can be presented to and used by military commanders and Congress.[5]

Sample Selection: Determining Whom to Survey

One important goal of a great deal of survey research is to collect data that provide accurate estimates about a population. In other words, researchers would like their survey assessments to correctly capture the characteristics of the population they survey. Many survey-sampling techniques have been developed in an effort to assist with better meeting this goal.[6] This section provides practical information regarding survey sampling that may help IIP planners obtain representative information from a population of interest, which can be one goal of assessments conducted for IO.[7] Failure to collect a representative sample will mean that proportions or other statistics calculated from survey data will not reflect or approximate the true population values.

Collecting Information from Everyone or from a Sample

A census involves collecting data from all the people in the population of interest. However, most research in the social sciences involves the collection of data from a sample of the population, rather than from every person in the entire population.[8] Results that approximate those that would have been obtained had data been collected from an entire population can be obtained from a small selection of people from the

---

[3]   Taylor, 2010.

[4]   Author interview with Kim Andrew Elliot, February 25, 2013.

[5]   Eles et al., 2012; interview with Mark Helmke, May 6, 2013.

[6]   Crano and Brewer, 2002.

[7]   Arturo Muñoz, *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations 2001–2010*, Santa Monica, Calif.: RAND Corporation, MG-1060-MCIA, 2012.

[8]   Crano and Brewer, 2002.

population, given a reasonable amount of statistical error. Thus, a large amount of money and time can be saved by collecting data from a well-considered sample, rather than by collecting a census.

Sample Size: How Many People to Survey

As noted, some error exists in terms of the extent to which a sample represents the population. In other words, the precision of a sample can vary. All else being equal, as the sample size increases, the potential error (in the extent to which the sample results represent the population results) decreases. A larger sample means less error. In addition, the greater variability on a particular topic or characteristic of interest in the population, the lower the precision of the sample is in estimating the results of the population. Thus, for example, if there are many people within a population who hold very different opinions on a topic, a greater sample size will be needed to better capture the population's opinion on the topic. This suggests that when IIP planners are attempting to obtain a sample that is representative of a population of interest, they should consider how much error they are willing to accept in terms of their survey estimates, and they should consider how much variability there seems to be in the population on the topic or characteristics of interest or how much variability there may be in attitudinal or behavioral change over time.

Another element to consider when determining how many people to collect survey data from is subsequent data analysis. In running statistical analyses, researchers want to be able to observe a relationship between variables. In other words, if there is an association to observe (sometimes there is not), researchers want to have enough statistical power to be able to observe that association and thereby find statistical significance. Several factors influence statistical power, or researchers' ability to observe an extant relationship. In addition to the amount of variability in responses (more variability or inconsistency makes it more difficult to observe a relationship), power can also be influenced by the strength of a relationship, such that larger relationships are easier to observe. Analyses that require more-stringent statistical significance (e.g., $p < 0.001$ versus $p < 0.05$) are associated with lower power. Sample size also influences power, such that larger samples sizes are associated with greater power.[9] Usually, researchers want to have an 80-percent chance of detecting an effect if it is present.

Some individuals have provided rules of thumb regarding sample sizes for different assessment approaches (see Table 10.1).[10] These recommended sample sizes can be inaccurate, so researchers have created tools that allow others to more accurately deter-

---

[9]   David C. Howell, *Statistical Methods for Psychology*, 5th ed., Pacific Grove, Calif.: Duxbury, 2002.

[10]   Mertens and Wilson, 2012.

Table 10.1
Approximate Sample Sizes as Based on Approach

Approach	Rough Approximation of Minimum Sample Size Required
Correlational	82 participants (two tailed)
Multiple regression	At least 15 participants per variable
Survey research	100 participants for each major subgroup: 20–50 for minor subgroups
Causal comparative	64 participants (two tailed)
Experimental or quasi-experimental	21 individuals per group (one tailed)

SOURCE: Adapted from Mertens and Wilson, 2012.

mine the number of people from whom they should collect data. A popular and free tool that may be used is call G*Power.[11]

In conflict environments, groups have collected between 1,500 and 12,000 individuals per survey wave.[12] The selected sizes of samples collected in these environments can vary based on the population characteristics and available resources.

Challenges to Survey Sampling

Thus far, this chapter has addressed sampling designs. When collecting data for a survey, multiple factors must be considered. For example, it can be difficult to obtain an accurate sampling frame, or all of those selected to participate in a study may not respond. We next address some of the challenges that may arise in survey sampling and ways that these challenges can be addressed.

Nonresponse

Rarely do all those who are asked to complete a survey agree to participate. This can lead to differences between the group that was sampled and the group that actually responded, which can keep results from being representative of the population of interest, even if the sample was selected in a representative way. This can occur because there may be systemic differences between those who choose to participate in the survey and those who choose not to participate. For example, those who participate may have more-favorable attitudes toward the government, may be more likely to be male, and may be better educated. Thus, their responses may not represent the total population of interest. This is called *nonresponse bias*. In a conflict environment, nonresponse is

---

[11] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang, "Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses," *Behavior Research Methods*, Vol. 41, No. 4, November 2009.

[12] Author interview with Emmanuel de Dinechin, May 16, 2013.

especially problematic, as many potential participants may be concerned about repercussions resulting from their responses. In determining the extent of nonresponse bias, researchers often calculate and report the response rate, which is the number of completed surveys divided by the total number of people asked to participate in a survey. To reduce nonresponse bias, different strategies may be implemented to promote responses. For example, female survey administrators may assist in promoting response rates among females, and the provision of small incentives may also increase response rates.[13] Keeping surveys at a reasonable length and guaranteeing anonymity of participant responses have also been suggested.[14] To reduce the impact of nonresponse bias, several analytic methods exist.[15] These often involve comparing information about respondents (e.g., location, gender) with known information about nonrespondents to see if nonresponse appears to be systemic (and concerning), or random (and thus less so).[16]

Lack of Access

Another issue that may arise in survey research involves access to areas that have been selected to be included in a study. For example, access may be denied, areas may be too difficult to reach, or areas may be too dangerous to enter.[17] However, these areas are often those of most interest to IIP efforts. Information on accessible and inaccessible areas should be maintained and reported by survey implementers.[18] In other words, if certain areas could not be included due to access issues, this information should be recorded and reported. In addition, it may be necessary to realign the sampling frame, based on areas that are accessible and inaccessible, so that additional data are collected from areas that can be accessed or previously unselected areas may need to be reconsidered.[19]

Collecting Survey Data from the Desired Individuals

Researchers who are interested in the attitudes of a certain group in a country, like the rural population, may incorrectly collect data from the entire country's population (those living in rural and urban areas). National-level polls represent the attitudes and opinions of the entire population of a country. The attitudes and opinions of those

---

[13] Author interview with Matthew Warshaw, February 25, 2013.

[14] Crano and Brewer, 2002.

[15] J. M. Brick and G. Kalton, "Handling Missing Data in Survey Research," *Statistical Methods in Medical Research*, Vol. 5, No. 3, September 1996.

[16] Brick and Kalton, 1996; Joseph L. Schaefer and John W. Graham, "Missing Data: Our View of the State of the Art," *Psychological Methods*, Vol. 7, No. 2, June 2002.

[17] Author interview with Matthew Warshaw, February 25, 2013.

[18] Eles et al., 2012.

[19] Eles et al., 2012.

living in particular areas of interest within a country may not be well understood by reviewing results from these polls.[20] Specifically, the information from those in certain areas of interest may be combined with the information from others in areas of less interest, hindering abilities to utilize the data more effectively. The areas and groups of interest should be determined and clearly described, and samples should be collected as based on this determination.

Interview Surveys: Options for Surveying Individuals

In addition to deciding whom to include in a survey effort, IIP planners must also consider how they are going to collect data from these individuals. There are different options for data collection. Four of the primary options include the following: in-person interviews, phone interviews, computer-based surveys, and mailed surveys. In-person interviews and phone interviews involve interviewers verbally asking each question, providing the response options for each question, and then recording the selected response. Computer-based surveys and mailed surveys, by contrast, tend to be self-administered, such that the participant reads the question and records his or her own answer.

The different data collection methods vary in terms of costs and information quality, and the method used should address the resources and capabilities of the population of interest.[21] Interview surveys can be costly and timely, since interviewers must sit with each person.[22] However, in a conflict environment in which many individuals are illiterate, an interview survey is the only viable option. When a large portion of a predominately illiterate population does not have telephones, in-person survey interviews are the only feasible option.

Conducting Survey Interviews In Person: Often Needed in Conflict Environments

Interview surveys have several advantages over self-administered surveys. They often have higher response rates than self-administered mail surveys, especially in conflict environments.[23] In addition, door-to-door surveys may contribute to more-reliable and less biased results.[24] Administering surveys in person may decrease the number of questions that respondents answer using the "don't know" or "refuse to answer" options, and interviewers can assist in addressing respondents' misunderstandings of survey items (but this must be strictly controlled). Finally, interviewers can record observa-

---

[20] Muñoz, 2012.

[21] Valente, 2002, p. 131.

[22] Author interview with Emmanuel de Dinechin, May 16, 2013.

[23] Author interview with Matthew Warshaw, February 25, 2013.

[24] Author interview with Kim Andrew Elliot, February 25, 2013.

tions regarding the respondents and their surroundings, such as characteristics of the dwelling and reactions of participants to certain survey items.[25]

However, different elements of survey interviews must be carefully considered. In conducting surveys, research personnel must try to ensure that each survey item is interpreted in the same way by different respondents. Thus, in survey interviews, the interviewer's presence and presentation of items should not influence, or should at least have as minimal an influence as possible on, how each respondent interprets and then answers each survey item. The interviewer's tone, nonverbal cues, and characteristics are all elements that may influence participant responses.

To address the influence of interviewer characteristics, some have suggested attempting to match the characteristics of the interviewer and respondent.[26] This may include matching race and ethnicity, first language spoken, religion, and gender.[27] For example, female interviewers may be used for interviewing female respondents, and citizens of a country may be used for interviewing respondents in that country.[28] By matching characteristics, respondents' answers may be less biased.

In addition, the survey interviewers should be well trained on how to administer a survey. Various rules for survey interviewing exist.[29] These rules stipulate that an interviewer's appearance and demeanor should somewhat correspond to those being interviewed—for example, an interviewer should dress modestly when interviewing poorer respondents. Further, interviewers should be very familiar with the questionnaire, such that they can read items without error. They should also read questions exactly as written and record responses exactly as provided. To ensure that interviewers follow these rules and administer surveys as intended, they should be well trained; they should be familiar with these provisions and the details regarding each question, and they should have opportunities to practice survey administration during training. When surveys are being administered in the field, procedures permitting careful supervision of interviewers, including supervisor presence during a certain proportion of each interviewer's surveys, should be established.

Additional Methods of Data Collection

In addition to in-person interviewing, several other options exist. As mentioned previously, telephone interviews may be used for populations that have ready access to telephones. Of note, some populations have moved from the use of landline telephones to greater use of cellular phones, which should be taken into account when determin-

---

[25] Babbie, 1990.

[26] Babbie, 1990.

[27] Author interview with Amelia Arsenault, February 14, 2013.

[28] Author interview with Matthew Warshaw, February 25, 2013.

[29] Babbie, 1990.

ing how to contact individuals.[30] Acknowledging the increased use of cellular phones, some groups have begun to use surveys based on short message service (SMS) or text messaging.[31] However, overreliance on surveys that use new technology may be detrimental when administering surveys to poor populations with limited access to phones or the Internet.[32]

The Survey Instrument: Design and Construction

In designing a survey that will later be administered to a certain sample of individuals, researchers seek to create a survey instrument with easily interpretable items that will not inadvertently bias participant responses.[33] When IIP assessment planners design (or contract for) surveys, they must consider question wording, question ordering, response options, and survey length.

Question Wording and Choice: Keep It Simple

In surveys, questions that are simpler are more likely to be understood by respondents.[34] Complex or vague questions that attempt to indirectly assess a certain topic can contribute to respondent confusion and reduce the utility of responses.[35] Questions should be short and use simple terms.[36] Maintaining the use of short and simple questions or survey items can also reduce the potential for double-barreled questions, which should be avoided.[37] In double-barreled questions, respondents are asked about two concepts in one question and are allowed to provide only one response. As such, researchers cannot determine which concepts that respondents are considering when answering the question. For example, the item "Do you think that certain groups have gone too far and that the government should crack down on militants?" addresses two concepts: the behavior of certain groups and the desired behavior of the government. A response to this question may be addressing either of these two concepts, but which one cannot be determined.

---

[30] Mark Blumenthal, "Gallup Presidential Poll: How Did Brand-Name Firm Blow Election?" *Huffington Post*, March 8, 2013.

[31] Author interview with Maureen Taylor, April 4, 2013

[32] Author interview with Lisa Meredith, March 14, 2013.

[33] Crano and Brewer, 2002.

[34] Crano and Brewer, 2002.

[35] Taylor, 2010, p. 10

[36] Valente, 2002, p. 124.

[37] Eles et al., 2012.

In addition to asking simple questions, IIP planners should keep the overall survey simple by keeping the length as short as possible. Survey fatigue occurs when respondents lose interest in and motivation to complete a survey. Survey fatigue can arise when participants perceive that they have been oversurveyed, or asked to complete too many surveys, and when they perceive that a particular survey is overly lengthy and time-consuming. To avoid survey fatigue, IIP planners should ensure that they are not overloading individuals with survey requests. One way to do this is to determine whether similar data are already being collected by other organizations and asking to share data with these groups.[38] In any single survey, planners should only ask for information that is most needed, thereby keeping the survey length as short as possible.[39] Another way to prevent survey fatigue is to inform participants how long it will take to complete a survey. Respondents may be less likely to experience fatigue when their expectations have been set prior to starting the survey.[40]

Open-Ended Questions: Added Sensitivity Comes at a Cost

Open-ended questions involve asking respondents a question and then allowing them to provide their own answers. For example, an open-ended question would ask, "Who is your favorite presidential candidate?" A closed-ended question would ask the same question and then provide a limited set of response options. By asking open-ended questions, information that would not otherwise have been captured may be collected. In addition, respondents can provide greater explanation regarding their responses.[41]

However, open-ended questions come with costs. It takes respondents longer to provide responses to open-ended questions. This increases the length of survey participation and may increase the likelihood of survey fatigue.[42] In addition, during in-person and phone interviews, interviewers must be able to capture participants' responses as accurately as possible. After surveys have been collected, interpreting and analyzing open-ended responses can be a complex and onerous process, involving the creation of a reliable coding scheme.[43] These questions should be used sparingly, when questions have no clear set of predefined answer options, or when more-detailed responses are needed.

---

[38] Eles et al., 2012.

[39] Author interview with Kim Andrew Elliot, February 25, 2013.

[40] Dillman, Smyth, and Christian, 2009.

[41] Author interview on a not-for-attribution basis, March 1, 2013.

[42] Dillman, Smyth, and Christian, 2009.

[43] Eles et al., 2012.

Question Order: Consider Which Questions to Ask Before Others

When implementing a survey, respondents who feel comfortable with and committed to the research may be more likely to respond to sensitive questions.[44] To establish comfort and build rapport, the least-threatening survey items should be asked at the beginning of the survey. Once respondents have answered these, they may be more willing to respond to later questions that may be perceived as more personal or threatening. One cannot assume that demographic questions are the least threatening. Income, education level, and marital status may all be sensitive topics, and these questions may raise privacy concerns for respondents. Instead, easy-to-answer questions that are relevant to the survey may be best to present first.

In addition, a person's responses to earlier questions can influence his or her responses to later questions. For example, if a number of questions ask respondents about the influence of terrorism on their country and a subsequent open-ended question asks what they believe to be one of the biggest threats to their country, terrorism may be a more likely response than it would have been had the open-ended question been asked first.[45] To address the potential for earlier items to affect responses to later items, the research recommends varying the order of item presentation so that the order varies across different questionnaires.[46] If this technique is implemented though, the full order should not be changed—specifically, the least-threatening items should remain at the beginning of the survey.

Survey Translation and Interpretation: Capture Correct Meaning and Intent

Another survey element to consider involves treatment of the survey after the original instrument has been developed. There is a possibility that survey items that were created in English and then translated to another language lost their original meaning and intent during the translation process.[47] To address this, researchers have utilized back-translation, which involves one person translating a survey from the original language to the target language and another person translating this survey back from the target language to the original language.[48] If the final translation is similar to the original translation, the researchers assume that the survey meaning and intent have been maintained. However, words that are literally equivalent in two different languages may not have equivalent meanings.[49] Further, certain groups may take offense to the

---

[44] Crano and Brewer, 2002.

[45] Babbie, 1990.

[46] Crano and Brewer, 2002.

[47] Eles et al., 2012.

[48] Robert Rosenthal and Ralph L. Rosnow, *Essentials of Behavioral Research*, 3rd ed., New York: McGraw-Hill, 2008.

[49] Martin Bulmer, "Introduction: The Problem of Exporting Social Survey Research," *American Behavioral Scientist*, Vol. 42, No. 2, October 1998.

wording of certain items, such as items regarding the rights of women and perceptions of elders.[50]

To reduce the possibility for questions to be interpreted in unintended ways or questions to be translated incorrectly, surveys should be reviewed by individuals who are local to the area to be surveyed.[51] One way to organize this review process is to conduct focus groups in which different sets of individuals review the questions and discuss their interpretations.[52] In addition, experts in opinion polling and in the cultural context of interest can provide valuable information regarding how participants may interpret the surveys and different issues that should be kept in mind.[53] Further, information from other organizations or groups who have collected data in the area can provide assistance with translation and interpretation issues.[54]

Multi-Item Measures: Improve Robustness

Surveys often seek to address complex concepts. A single survey item may not adequately address a complex concept. For example, to assess religiosity, a researcher may include a survey item addressing frequency of mosque or church attendance. However, those who frequently attend mosque or church may not be perceived as strongly religious if other items were used, like frequency of prayer or strength of different beliefs.[55] As such, it is often worthwhile to utilize more than one item to assess a construct. Together, these items are called a scale or index.[56] If all of the items in a scale are assessing the same construct, these items can be aggregated. Use of scales can provide more-comprehensive and more-reliable measures of complex concepts than use of single items.[57]

There are multiple different options for scale creation, including Thurstone scales, Guttman scaling, Osgood's semantic differential technique, and Likert's method of summated ratings.[58] A thorough description of each of these scale techniques is beyond the scope of this chapter. However, one of the most common techniques is the use of Likert scales.[59] With this method, participants are provided with several items on a topic, presented as a range, and can pick one response to each item. For example,

---

[50] Eles et al., 2012.

[51] Author interview with Amelia Arsenault, February 14, 2013.

[52] Author interview with Matthew Warshaw, February 25, 2013.

[53] Eles et al., 2012.

[54] Author interview with Matthew Warshaw, February 25, 2013.

[55] Babbie, 1990.

[56] Valente, 2002, p. 151.

[57] Author interview with Ronald Rice, May 9, 2013.

[58] Crano and Brewer, 2002.

[59] Ronald J. Thornton, "Likert Scales: An Assessment Application," *IO Sphere*, Summer 2013.

a survey might ask participants the extent to which they agree with the following state-ment: "The national government has had a positive influence on my life." Participants could then indicate their level of agreement using one of five possible response options (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree). Several additional items addressing perceptions of the national government may be asked, and then these items may be summed or averaged together. Before combining responses to items, the extent to which the items are related should be assessed. If items are positively related, that suggests that they are measuring the same construct. One way to assess whether scale items are sufficiently related is through the calculation of an alpha coefficient.

When using scales in surveys, IIP planners should keep at least two points in mind. First, as noted previously, inundating participants with items can contribute to survey fatigue. There must be a balance between comprehensively capturing a con-struct and asking an excess of questions on a topic. To assist in striking this balance, the use of preexisting scales that have been demonstrated to be reliable and valid in previous studies should be strongly considered. However, a second point to keep in mind is that previous studies may have included participants who are unlike those of interest in a particular survey. Specifically, a great deal of research in the social sciences involves use of Western college students.[60] It may be necessary to customize the preex-isting scales to a local context.[61]

Item Reversal and Scale Direction: Avoid Confusion

The simplest surveys consist of items with parallel constructions. That is, questions are posed in a similar way and the response options are the same across all questions. Sometimes, survey developers opt to include questions that follow a different format, solicit a different type of response, or request that respondents relay their responses using a scale that moves in the opposite direction. This is often done for lack of a better approach to collect the information required, but asking the exact question you need to ask to obtain the exact information you require has a downside: Changing formats and scales may confuse participants, increasing the risk that you will receive inaccurate data anyway. A 2009 simulation study that involved administering surveys with tradi-tional Likert-format items to school administrators found that incorrect responses to reverse-coded items can have a statistically significant impact on the resulting data.[62] Further, items that need to be reversed before being combined in indexes or scales with other items risk being reversed more than once between collection and final analysis;

---

[60] Joseph Henrich, Steven J. Heine, and Ara Norenzayan, "Most People Are Not Weird," *Nature*, Vol. 466, No. 7302, July 1, 2010.

[61] Author interview with Amelia Arsenault, February 14, 2013.

[62] Gail Hughes, "The Impact of Incorrect Responses to Reverse-Coded Survey Items," *Research in the Schools*, Vol. 16, No. 2, Fall 2009.

several of the authors were involved in analyses of survey data for a DoD IIP effort in which the raw data were not kept inviolate, and several items were reversed in the raw data, perhaps multiple times. This leads to two suggestions: (1) where possible, avoid reverse-scale items, and (2) always protect and preserve the raw data so that any analytically driven recoding can be tracked and undone, if necessary.

Testing the Survey Design: Best Practices in Survey Implementation

Thus far, most of the discussion has focused on actions that IIP planners can take to address specific challenges that can arise during survey implementation. There are also several broad approaches and actions that can improve the accuracy and utility of the overall survey. These involve the systematic assessment of the survey at every stage in the process and maintenance of consistency after survey administration.

Pretesting

When implementing new surveys or when implementing old surveys in a new social context, pretests should be conducted with these surveys. Before implementing a full-scale survey, survey designers should determine whether people in the context of interest understand the questions, whether these people are able to respond to the questions, and whether interviewers can appropriately administer the questions in this social context.[63] Different avenues that are available for pretesting a survey include focus group discussions regarding the survey and individual interviews in which participants provide survey responses and explain what they were thinking when responding to each item. After critically reviewing responses and using feedback to modify questions, pilot testing in the field should be conducted, which is done by administering a small number of surveys and assessing results from those surveys.[64] Pretesting and pilot testing can help address potential issues with a survey before the costly wide-scale implementation.

Maintaining Consistency

At times, commanders or IIP planners may seek to assess changes in attitudes or perceptions among those living in a certain social context. In these instances, it is important that there is continuity in the surveys that are implemented at different time points. In other words, the maintenance of a core set of items that uses the same wording and same response options is needed in order to permit assessment of changes in responses to these items over time. Utilization of different wording, different response options, or entirely different scales hinders assessment of changes in attitudes, because researchers cannot determine whether changes in responses are due to variation in actual attitudes or whether observed changes are due to variation in the measurement

---

[63] Floyd J. Fowler, *Improving Survey Questions: Design and Evaluation*, Thousand Oaks, Calif.: Sage Publications, 1995.

[64] Fowler, 1995.

of these attitudes within surveys. If command changes and new commanders would like to measure different constructs within a survey, careful consideration should be given to changes, because, again, consistency and continuity of a core set of survey items can permit better assessments of change.[65]

Review of Previous Survey Research in Context of Interest

To assist with development of new survey research in a geographic area, IIP planners should review previous research that has been conducted in the area and previous research that has been conducted on the topics of interest. Multiple examples of survey research are available and may be consulted for this purpose. These include Altai Consulting's assessment of Afghan media in 2010, YouGov data collected in Iraq, and various research efforts conducted by the British Council.[66]

Response Bias: Challenges to Survey Design and How to Address Them

There are a number of factors that may influence participant responses to survey items. As noted previously, these include interviewer characteristics and question ordering. Ideally, researchers would like the characteristics of the survey to have a minimal influence on responses. However, this can be difficult, and those designing a survey should be aware of factors that influence participant responses.

Social Desirability Bias

One potential threat to capturing respondents' true attitudes and perceptions is known as *social desirability bias*. This involves individuals' desires to present themselves in a manner that their society regards as positive.[67] In other words, rather than responding to an item or set of items in a way that reflects their true perceptions or actual attitudes, participants may instead respond based on how they believe their society would like them to respond.

To address this problem, some surveys include a ten-item social desirability scale. If responses to certain survey items are strongly correlated with participants' responses

---

[65] Eles et al., 2012.

[66] See Altai Consulting, "Afghan Media in 2010," prepared for the U.S. Agency for International Development, 2010 (the synthesis report and supplemental materials, including data sets and survey questionnaires, are available for download); UK Polling Report, *Support for the Iraq War*, online database, undated; British Council, *Trust Pays: How International Cultural Relationships Build Trust in the UK and Underpin the Success of the UK Economy*, Edinburgh, UK, 2012.

[67] Robert F. DeVellis, *Scale Development: Theory and Applications*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2012.

on this scale, those survey items should possibly be excluded from analyses.[68] Informing participants that their responses are anonymous and cannot be connected back to the participant may increase candor, reducing the influence of social desirability bias.[69]

Response Acquiescence

Another factor that may distort participant responses is known as *response acquiescence*. Other terms for this same concept include *agreement bias* and *response affirmation*. This occurs when participants agree with survey items, regardless of their content.[70] Given a set of survey items and asked to respond on a scale ranging from 1 (strongly disagree) to 5 (strongly agree), biased respondents will tend to express higher levels of agreement with each item, without thoroughly considering what they are agreeing to.

To address this, survey developers include both positively and negatively worded items within a scale. For example, if assessing self-esteem, they may include items focused on high self-esteem (e.g., "I feel that I have a number of good qualities") and items focused on low self-esteem (e.g., "I feel useless at times").[71] The responses of someone who tends to agree with all items, regardless of content, would be balanced across survey items, revealing response acquiescence. Unfortunately, using positively and negatively worded items may confuse respondents. In addition, analysts must reverse code the negatively worded items (e.g., change a response of 1 to a 5 and a response of 2 to a 4) so that all items move in the same direction for analyses. This process may confuse some analysts. (See the section "Item Reversal and Scale Direction: Avoid Confusion" earlier in this chapter.)

Mood and Season

An additional factor that may influence participants' responses is their mood, which may be associated with the season. For example, previous research has shown that participants respond more negatively when it is raining than when it is a sunny day.[72] Others have noted that participants in conflict environments may have difficulty finding fuel to cook or keep warm in the winter, which may dampen their general outlook.[73]

To address the influence of season and mood on responses, researchers should consider collecting data over different time periods, assessing patterns in responses across these periods. By obtaining a better understanding of trends in responses at dif-

---

[68] DeVellis, 2012.

[69] Crano and Brewer, 2002.

[70] Crano and Brewer, 2002.

[71] DeVellis, 2012.

[72] Norbert Schwarz and Gerald L. Clore, "Mood, Misattribution, and Judgments of Well-Being: Information and Directive Functions of Affective States," *Journal of Personality and Social Psychology*, Vol. 45, No. 3, September 1983.

[73] Author interview with Matthew Warshaw, February 25, 2013.

ferent periods of the year, the influence of season on these responses may be ascertained. Another possibility in addressing the influence of season or weather on responses is to first ask participants questions about the weather. Doing so may decrease the likelihood that participants will incorrectly attribute their negative feelings to their general life situations rather than the bad weather.[74]

Using Survey Data to Inform Assessment

After survey data have been collected, they must be analyzed, triangulated with other data sources, and interpreted so as to meaningfully inform IIP assessment. This section addresses these processes.

Analyzing Survey Data for IIP Assessment

This section offers broad, high-level recommendations for the analysis of survey data in support of IIP assessment in conflict areas. It does not address statistical procedures in detail. There are several texts that provide a thorough treatment of statistical methods for the analysis of survey data in the social and behavioral sciences, including work by Joseph Healey and James Paul Stevens.[75]

To allow for analysis of trends over time, all waves of the survey should be combined into a master data set. The absence of a master data set has complicated efforts to analyze some polls in Afghanistan.[76] Merging multiple waves of survey data, along with other analytical techniques, is facilitated by the use of statistical software like SAS, STATA, and R. Polling programs should use advanced statistical packages but should keep versions of the data sets in standard formats to facilitate sharing and transparency.[77] It is worth reemphasizing here that the quantity and quality of the data are far more important than the analytical technique or software program. Even the most-sophisticated techniques cannot overcome bad data.

The sampling error, often expressed as the margin of error, represents the extent to which the survey values may deviate from the true population values. As discussed in the preceding section on sample selection, the survey error is inversely related to the sample size. In Afghanistan, nationwide surveys have margins of error of plus or minus 3 percent, and district surveys have margins of error closer to 10 percent.[78] Because less is known about the population in operating environments like Afghanistan, survey research should be continuously informing estimates of design effects and associated

---

[74] Crano and Brewer, 2002.

[75] Healey, 2012; Stevens, 2009.

[76] Eles et al., 2012, p. 37.

[77] Eles et al., 2012, pp. 36–37.

[78] Downes-Martin, 2011, p. 110.

margins of error. When data from multiple surveys are available, analysts should examine variation across variables that should be constant (e.g., age, marital status) to revise estimated survey errors.[79]

Surveys in Afghanistan typically use disproportional stratified samples, because proportional sampling is typically not feasible. To make inferences about the general population, analysts must therefore apply population weighting when aggregating survey data across the different strata. In the review of the Kandahar Province Opinion Polling program, "oversights regarding the requirement for weighting of the data" were among the most commonly observed mistakes.[80]

Analyzing and Interpreting Trends over Time and Across Areas

Survey results can be used to shape how decisionmakers perceive trends over time and across areas. The best surveys in support of IIP assessment are therefore those that are conducted in several areas and repeated frequently over time. This is true for several reasons. First, as described previously, surveys in conflict environments are particularly prone to response and nonresponse bias. Analyzing data over time and across areas controls for these sources of bias, presuming that they are not correlated with time or region.[81] Second, repeated measurements provide a means to validate the survey by assessing if observed shifts in attitudes exhibit expected relationships with known or likely triggers of attitudinal change, such as upticks in violence or kinetic operations, civilian casualties, or political turmoil. For example, the quarterly ANQAR survey and the annual Survey of the Afghan People by the Asia Foundation have both been conducted for nearly a decade. These surveys are well respected because they have tracked well with events over time and because the previous waves make it easier to identify errors in the data collection process.[82]

Finally, IIP assessment is typically not concerned with a snapshot of attitudes but rather with whether there are attitudinal or behavioral changes over time that can be traced to IIP activities. However, making these causal inferences is the responsibility of the evaluators and not the survey research group. Survey researchers should avoid assessing causal linkages when presenting results to sponsors.[83]

---

[79] Eles et al., 2012, p. 36.

[80] Eles et al., 2012, pp. 36–37.

[81] Eles et al., 2012, pp. 37–38.

[82] Author interview with Katherine Brown, March 4, 2013; interview with Matthew Warshaw, February 25, 2013.

[83] Eles et al., 2012, p. 38.

Triangulating Survey Data with Other Methods to Validate and Explain Survey Results

Given the large margins of error and challenges posed by nonresponse and response biases, survey data are most valuable to IIP assessment when analyzed over time and in conjunction with other qualitative or quantitative data sources. Evaluators should validate survey result by assessing whether data or indicators produced by other methods—e.g., content analysis, focus groups, Delphi panels, atmospherics—are trending in the same direction or converging with survey data. This point was made by nearly every expert interviewed for this study with experience conducting or using surveys in conflict environments.[84] As Steve Booth-Butterfield explains, "Survey data are one part of the argument, . . . but you are building an argument that depends on more than one piece."[85]

In addition to validating survey results, other methods—particularly qualitative methods—should be used to explain and interrogate survey results, especially if they are unanticipated. It is often stated that the survey data tell you the "what" and the qualitative data tell you the "why."[86] As mentioned in Chapter Eight, the relationship between qualitative methods and survey research can be characterized as an *iterative process*: Qualitative research informs the design of the survey, and the survey generates questions that are probed by a second iteration of qualitative research that feeds into the revision of the survey instrument.[87]

Summary

This chapter provides an overview of several points to consider when designing and implementing a survey to produce informative results for IIP efforts. Poorly designed surveys and poorly implemented data collection efforts can be costly and produce ambiguous or misleading information. As such, time and resources should be reserved for the design of a survey effort. Key takeaways are as follows:

- Those responsible for contracting, staffing, or overseeing the administration of a survey in support of IIP assessment should adhere to best practices for survey management, including engaging experts and local populations in survey design, vetting and tracking the performance of local firms, and maintaining continuity throughout the survey period.

---

[84] Author interviews with Simon Haselock, June 2013; Jonathan Schroden, November 12, 2013; Steve Booth-Butterfield, January 7, 2013; and Maureen Taylor, April 4, 2013.

[85] Author interview with Steve Booth-Butterfield, January 7, 2013; see also Booth-Butterfield, undated.

[86] Author interview with Maureen Taylor, April 4, 2013.

[87] Author interview with Thomas Valente, June 18, 2013.

- IIP planners should consider who they would like to survey, how many people to survey, and what procedure to use to administer the survey. Survey takers should represent the target population as closely as possible.
- When considering the ideal number of people from whom to collect survey data, IIP planners should keep in mind the variability in the attitudes and behaviors of the population of interest. Generally, greater variability warrants larger sample sizes.
- Nonresponse and lack of access are challenges inherent in all survey efforts. This is especially true for survey efforts conducted in conflict environments, where populations may move frequently, people may lack access to telephones or the Internet, and areas are inaccessible.
- Surveys should be designed so that the instrument or collection methods do not greatly influence participant responses. Question wording and overall survey length, question structure, question order, and response options can all affect participants' responses.
- Social desirability bias (a desire to conform to social expectations), response acquiescence (a tendency to agree with questions, regardless of their content), and even the respondent's mood, the season, and the weather can affect responses.
- Best practices in survey design and implementation favor the systematic assessment of the survey at every stage in the process, including after the survey is administered.
- Triangulating survey results, comparing a survey's results with information obtained from other surveys or focus groups, may also assist with survey validation.

Presenting and Using Assessments

By now, the spaghetti graph, as it has come to be known, is infamous for its complexity and overlapping lines. According to a *New York Times* article, when GEN Stanley McChrystal was the leader of American and NATO forces in Afghanistan, he jokingly remarked, "When we understand that slide, we'll have won the war."[1] The moral of the story is that how one presents and uses assessment matters, because assessment supports decisionmaking, and poorly presented assessments poorly support decisionmaking. As Maureen Taylor noted, "The biggest challenge facing assessment is getting information into a form that the people who make decisions on the ground can use."[2]

This chapter builds on the earlier chapters by detailing how assessment results can be best presented and ways that assessment can be utilized effectively. The chapter begins by revisiting the importance of assessment to decisionmaking. After discussing the presentational art of assessment data, it then turns to tailoring presentations to stakeholders and using data visualization and narrative. It concludes with a review of meta-analysis: the process of evaluating evaluations.

Assessment and Decisionmaking

Assessments should be designed with the needs of stakeholders in mind; this fully carries over to the presentation of assessments. Only by having a clear understanding of both the assessment users (stakeholders, other assessment audiences) and the assessment uses (the purposes served and the specific decisions to be supported) can assessment be tailored in its design and presentation to its intended uses and users and thus adequately support decisionmaking. Presenting information will mean nothing unless the data are shared with stakeholders who play a major role in decisionmaking. This

---

[1] Elisabeth Bumiller, "We Have Met the Enemy and He Is PowerPoint," *New York Times*, April 26, 2010.

[2] Author interview with Maureen Taylor, April 4, 2013.

provides an impetus to offer better training in data-driven decisionmaking and to make the results and data more accessible to those not trained in research methods.[3]

Disseminating the findings of the research can be just as important as how the results are presented. According to Thomas Valente, dissemination is "the most important yet most neglected aspect of evaluation," and it is "neglected because we don't know what to do until the findings are known."[4] Dissemination and documentation should follow an agreed-upon framework or plan. Findings should be communicated at several stages throughout the research process, since the communication of findings is a process, not a one-time event or a product.[5]

The Presentational Art of Assessment Data

Deciding how and how much assessment data to present in a report or briefing is a difficult challenge. Too much, and the reader or recipient will drown in the data, fail to see the forest for the trees, or simply ignore the material as being too opaque and not sufficiently accessible. Too little data, on the other hand, and the recipient will lack confidence in the results, question the validity of the findings, or ask important questions that the underlying (but unavailable) data should easily answer.

When presenting data, knowing your audience is paramount. In his work on making statistical presentations more meaningful, Henry May outlines three main principles. The first is understandability. Results need to be reported in a form that can be widely understood, makes minimal assumptions about the audience's familiarity with statistics, and avoids the overuse of jargon. The second principle is interpretability, meaning that the metric or unit of measure on which a statistic is based can be easily explained. Finally, May believes that comparability is critical. Simply put, the statistics that might be compared can be compared directly, obviating any need for further manipulation.[6]

When presenting the data in charts and graphs, consider the most effective way to appropriately communicate the information to the audience. Before constructing charts and graphs, consider their necessity and structure. Reduce "chart junk," including unnecessary graphics. Be thoughtful when ordering data points; for example, figure out whether to rank points in order of priority or whether alphabetical order is appropriate.[7] Overall, it is best to present dense and rich data as clearly and simply as

---

[3]   Author interview with Maureen Taylor, April 4, 2013.

[4]   Author interview with Thomas Valente, June 18, 2013.

[5]   Valente 2002, chapter 14.

[6]   Henry May, "Making Statistics More Meaningful for Policy Research and Program Evaluation," *American Journal of Evaluation*, Vol. 25, No. 4, 2004.

[7]   Howard Wainer, "How to Display Data Badly," *American Statistician*, Vol. 38, No. 2, May 1984.

possible to let the research speak for itself. However, do not assume that data speak for themselves; what is obvious to an assessor who has spent hours poring over and analyzing a matrix of data will likely not be obvious to a first-time viewer of even a relatively simple data table. Simple presentation of rich data is often a good idea, and it becomes even better if accompanied by guide marks or a clear statement of what the key takeaway should be.

As the example of General McChrystal's spaghetti graph demonstrates, PowerPoint has its own limitations. In an article titled "PowerPoint Is Evil," the famed researcher on the visual presentation of data Edward Tufte wrote, "The practical conclusions are clear. PowerPoint is a competent slide manager and projector. But rather than supplementing a presentation, it has become a substitute for it. Such misuse ignores the most important rule of speaking: Respect your audience."[8] While many IIP assessment presentations and briefings must still rely on PowerPoint, the takeaway remains clear: Understand and meet the needs of your audience, and respect your audience. Make clear when complicated data support a simple conclusion, and have a more detailed presentation of those data available if needed (perhaps in the backup slides). Again, Tufte's words are instructive:

> Presentations largely stand or fall on the quality, relevance, and integrity of the content. If your numbers are boring, then you've got the wrong numbers. If your words or images are not on point, making them dance in color won't make them relevant. Audience boredom is usually a content failure, not a decoration failure.[9]

One form that can be very effective is quantitative data supported by narrative and qualitative data. Qualitative data are illustrative and provide context to the numbers, while narrative is a strong way to summarize assessments. To be sure, those that explicitly mention a theory of change and how well it is working are even better. All assessments—even narratives—should clarify the underlying data and level of confidence in the result. Presentational art includes finding the right balance in discussing methods and evidence. As one SME concluded, "It is important that you do good science; it is also important that you sell good science."[10]

Tailor Presentation to Stakeholders

We live in a world where we have more access to data than ever before. This is a true double-edged sword because, while access to these data can help us solve problems in

---

8   Edward Tufte, "PowerPoint Is Evil," *Wired*, September 2003.

9   Tufte, 2003.

10 Author interview on a not-for-attribution basis, July 30, 2013.

ways that were previously unimaginable, the signal-to-noise ratio has increased exponentially. In other words, commanders and decisionmakers are inundated with more data than they can reasonably comprehend, so the onus is on those *presenting* the data to tailor their presentations to stakeholders. We have all heard of the elevator speech—the 30-second pitch that perfectly captures the main takeaways from your research. Tailoring presentations to stakeholders is built around this same logic.

Dissemination should adhere to a certain framework, and findings need to be tailored to their intended audiences.[11] Decisionmakers in conflict zones are busy. In terms of reading evaluations, the executive summary is critical: "Often, no one reads anything except the executive summary, so you have to make it count."[12]

To properly tailor the presentation of assessment results to stakeholders, it is crucial to know what they need to know to support the decisions they need to make. Here, it is important to take care when aggregating assessments of individual efforts or programs. In other words, sometimes the whole really is greater than the sum of its parts.

The notion of utilization-focused evaluation was developed by Michael Patton, whose work focuses on multiple ways of communicating with stakeholders throughout an evaluation. Patton believes strongly that a final report should not always be the instrument for providing information for decisionmaking.[13]

On the contrary, utilization-focused evaluation provides information to intended users, and its components include the discussion of potential uses of evaluation findings from the very beginning of a project, not only at the end, when the data are in hand. Patton realizes that encouraging stakeholders to think about what they want to do with evaluation findings before any data are collected is an effective strategy for collecting data that have an increased probability of being used. Another key aspect is the identification of intended users.[14] One interviewee phrased it like this: "It goes to believing the data that we've presented. . . . Set expectations from the get-go."[15]

Closely related to tailoring presentations to stakeholders is the question of how much data to present and in what format. Any effective assessment will include communicating progress (or a lack thereof) through both interim and long-term measures. Some stakeholders will need more hand-holding than others, but the onus is on the research organization to have the data and the ability to provide updates in a meaningful and measurable way.[16] NATO's JALLC framework for the evaluation of public

---

[11] Author interview with Thomas Valente, June 18, 2013.

[12] Author interview with Amelia Arsenault, February 14, 2013.

[13] Oral History Project Team, "The Oral History of Evaluation, Part 5: An Interview with Michael Quinn Patton," *American Journal of Evaluation*, Vol. 28, No. 1, March 2007, cited in Mertens and Wilson, 2012.

[14] Oral History Project Team, 2007, cited in Mertens and Wilson, 2012.

[15] Author interview on a not-for-attribution basis, July 30, 2013.

[16] Author interview with Heidi D'Agostino and Jennifer Gusikoff, March 2013.

diplomacy has devised three separate evaluation products to represent three levels of reporting: dashboards, scorecards, and evaluation reports.

A dashboard provides an overview of monitoring, usually of outputs. It can be used in real time with some media monitoring applications and can be used to produce regular and frequent reports. A dashboard is essentially data with little or no built-in evaluation and limited explanative narrative. A dashboard would typically be updated at least monthly. A scorecard is a display format for less frequent reporting, as it shows progress toward the desired outcomes and desired impacts. A scorecard is essentially data with little or no bulletin evaluation and limited explanative narrative. A scorecard would typically be updated quarterly or biannually.

An evaluation report is a periodic, typically annual, evaluation of results. It presents a balanced view of all relevant results and aims to show what meaningful changes have occurred, how they might be linked to activities, and whether the objectives have been achieved. It should contain narrative answers to the research questions and explain what has worked, what has not, and, whenever possible, why. Evaluation reports can also be published to cover a specific event or program.[17]

Data Visualization

Assessments can be presented in a variety of forms, including research reports, policy memorandums, and PowerPoint briefings packed with a dizzying array of quantitative graphs, maps, and charts. Senior military leaders and policy staffs use these materials for a variety of purposes, including to assess the progress of military campaigns, evaluate resource allocation (or reallocation), identify trends that may indicate success or failure, and discern whether and when it may be necessary to alter a given strategy.[18] It is important to think about different ways to present important data so that they can be visualized properly and have the proper effect (see Table 11.1).

Sometimes, to truly make sense of the data, it is important to visualize them. To really ramp up the productivity of the data, you need a way to ramp up the visualization technology. One tool for doing so is software called Ignite. Such programs allow you to visualize both structured and unstructured data. Using these types of programs can be a great way to demonstrate progress toward your end state.[19] Infographics can also help communicate research results to decisionmakers in the field.[20] A picture is indeed worth a thousand words, if you can generate the right picture.

---

[17] NATO, Joint Analysis Lessons Learned Centre, 2013, p. 12. Illustrations of each type of evaluation product are provided in chapter 3 of the framework.

[18] Connable, 2012, p. iii.

[19] Author interview with LTC Scott Nelson, October 10, 2013.

[20] Author interview with Gerry Power, April 10, 2013.

Table 11.1
A Checklist for Developing Good Data Visualizations

When producing visual presentations, you should think about these things:

☑ **The target group.** Different forms of presentation may be needed for different audiences (e.g., business vs. academia, specialists vs. general population).

☑ **The role of the graphic in the overall presentation.** Analyzing the big picture or focusing attention on key points may require different types of visual presentations.

☑ **How and where the message will be presented** (e.g., a long, detailed analysis or a quick slide show).

☑ **Contextual issues that may distort understanding** (e.g., experts vs. novice data users).

☑ **Whether textual analysis or a data table would be a better solution.**

Accessibility considerations:

☑ Provide text alternatives for nontextual elements, such as charts and images.

☑ Don't rely on color alone. If you remove the color, is the presentation still understandable? Do color combinations have sufficient contrast?
Do the colors work for color-blind users (e.g., red/green)?

☑ Ensure that time-sensitive content can be controlled by the users (e.g., the pausing of animated graphics).

☑ **Consistency across data visualizations.** Ensure that elements within visualizations are designed consistently, and use common conventions where possible (e.g., blue to represent water on a map).

☑ **Size, duration, and complexity.** Is your presentation easy to understand?
Is it too much for the audience to grasp in a given session?

☑ **Possibility of misinterpretation.** Test your presentation on colleagues, friends, or some people from your target group to see whether they get the intended messages.

SOURCE: Modeled on Mertens and Wilson, 2012.

The Importance of Narratives

While visual representations of data can help communicate key points to an audience, to avoid losing the nuance of assessment results, it is pertinent to place metrics in context and frame these visual representations within broader explanatory narratives. This means balancing quantitative metrics with probability and accuracy ratings and also identifying and explaining gaps in the available information. To remain transparent, all information should be clearly sourced and should be retrievable by those seeking an in-depth understanding of specific subjects. Quantitative reports should be presented as part of holistic, all-source analysis.[21]

---

[21] Connable, 2012, p. xix.

It is a solid and acceptable plan to move away from slide shows and stoplight charts as the products of operations assessment and toward narrative formats for products. Narratives can be more time-consuming to produce and consume but might be better suited to communicating progress than a colored map or a series of red, yellow, and green circles. The latter typically invites questions requiring further explanation, particularly when dealing with a subject as complex as warfare.

For assessment, narratives offer an array of advantages. For example, they allow variations and nuances across an area of operations to be captured and appreciated; they remind people of the context and complexity of the operation; they force assessors to think through issues and ensure that their assessment is based on rigorous thought; and they are the only way to ensure that a proper balance is struck between quantitative and qualitative information, between analysis and judgment, and between empirical and anecdotal evidence.[22]

Narratives would be even more effective if they made explicit reference to a theory of change, and included discussion of critical nodes and assumptions that need to be called into question. In short, narratives should not just tell a story; they should tell a consistent and logical story. To accomplish this requires stating assumptions and expectations up front; why you hold these assumptions and expectations; what you observed (and why you think you observed this); what, if any, progress you made; and what you are planning to do differently going forward (and why you think that will make things different, if you do).

Part of communicating a narrative includes relying on anecdotes. What works in communicating effectiveness to donors, Congress, and others is combining quantitative data with anecdotes to color and provide context to the numbers.[23] How do you demonstrate that any organization is important to those who might be skeptical? One way is to use stories and find ways to empower voices of experience—those who have personally benefited from a communication campaign—especially foreign audiences.[24] Sometimes it is unnecessary to measure, because the results are evident, such as Japan's response to U.S. assistance in the wake of the 2011 tsunami.[25]

Depending on the audience, the use of strong anecdotes, such as adversary messages that illustrate awareness of and concern about your efforts, can be a potent demonstration of the effectiveness of a campaign. These attention grabbers are what prompt Capitol Hill audiences' interest in assessment work, which is essential when it comes to securing funding.[26] The following sections address the benefits of narratives in increas-

---

[22] Schroden, 2011, p. 99.

[23] Author interview with Maureen Taylor, April 4, 2013.

[24] Author interview with Nicholas Cull, February 19, 2013.

[25] Author interview with Mark Helmke, May 6, 2013.

[26] Author interview on a not-for-attribution basis, July 18, 2013.

ing understanding, which facilitates the translation of aggregated data into terms that best support decisionmaking and the process of soliciting valuable feedback from end users of assessment results.

Aggregated Data

Transparency and analytic quality might enhance the credibility of aggregated quantitative data.[27] Using numerical scales for aggregate assessment (e.g., "The campaign gets a 4") is often unhelpful to decisionmakers:

> It is extremely unhelpful for an information consumer to get hung up on why an assessment is 2 as opposed to 3—something forgotten by organizations that operate on ratings such as 3.24. The important messages to communicate were movements, projections, and gaps against a defined end state, which spoke directly to the planning process. The scale, therefore, was a distraction.[28]

When aggregating, it is important to remember that ordinal scales can be aggregated and summarized with narrative expressions but not (accurately) with numbers. The simple statement, "All subordinate categories scored B or above except for reach in the Atlantica region, which scored a D," is much more informative than "The Atlantica region scored a 2.1 for reach."

Because a whole really can be greater than the sum of its parts, one must take great care when aggregating assessments of individual efforts or programs to avoid junk arithmetic. Ordinal scales cannot be meaningfully averaged, but because they are represented with numbers, they can be and are often subjected to inappropriate calculations. Ordinal scales are better represented as letter grades than as numbers; it is harder to inappropriately average C, C, and A than it is to inappropriately average 1, 1, and 4. Ordinal scales can be aggregated and summarized with narrative expressions, but not with junk arithmetic (see the discussion in Chapter Nine).

Report Assessments and Feedback Loops

Presenting assessment results is a way of disseminating research findings that directly inform decisionmaking. But disseminating findings is just one piece of the puzzle. To generate valuable feedback loops, those presenting the research must receive feedback from the end user. This, in turn, enables evaluators to know about the utility of the method, improvements going forward, and general feedback about what was successful and what was less successful.

Efforts to improve transparency should include making data and results public and stressing the importance of feedback, both from individuals who have a broad

---

[27] Connable, 2012, p. xix.

[28] Upshur, Roginski, and Kilcullen, 2012, p. 99.

understanding of the issue of interest and from those who have an understanding of specific circumstances and audiences.

Evaluating Evaluations: Meta-Analysis

With all of the time, effort, and resources dedicated to conducting evaluations, how do we know whether an evaluation is sound? By stepping back and conducting research about research, we are, in essence, conducting a form of meta-analysis. In the evaluation context, this means using metaevaluation to assess the assessment. Metaevaluation is the extent to which the quality of the evaluation itself is assured and controlled. Its purpose is to be responsive to the needs of its intended users and to identify and apply appropriate standards of quality. Metaevaluations should be based on adequate and accurate documentation.

Metaevaluation

The term *metaevaluation* was coined by Michael Scriven in reference to a project to help the Urban Institute evaluate the quality and comparability of its evaluations. Metaevaluation is an indicator of an evaluation's quality, and the metaevaluation standard should be used to determine several related issues. These include serving the intended users' needs, the identification and application of appropriate standards of quality, and providing adequate and accurate documentation as a foundation.[29]

Carl Hanssen and colleagues conducted a "concurrent" metaevaluation by evaluating a U.S. federal agency's evaluation technique as it was being developed and implemented. The researchers were involved in the evaluation throughout the entire process and even attended data collection events while verifying the quality of the data collected. Three questions formed their starting point:

- What are the strengths and weaknesses of the evaluation process, including each of its components in terms of producing its intended results? How can the evaluation process be improved?
- How efficacious is the evaluation process for producing its intended results?
- To what degree does the evaluation framework enable an evaluator to produce an evaluation that satisfies accepted program evaluation standards?[30]

Metaevaluation Checklist

The metaevaluation checklist (see Appendix A) is an appropriate tool for summative evaluations or summative evaluations with a process evaluation component. The

---

[29] See Mertens and Wilson, 2012.

[30] Carl E. Hanssen, Frances Lawrenz, and Diane O. Dunet, "Concurrent Meta-Evaluation: A Critique," *American Journal of Evaluation*, Vol. 29, No. 4, December 2008, cited in Mertens and Wilson, 2012, p. 516.

checklist can be used for assessments of actual influence efforts, but not for supporting or enabling efforts that do not have some form of influence as an outcome. The various sections of the metaevaluation checklist include: SMART objectives, the inclusion of an explicit theory of change in the assessment, measurement, use of surveys, analysis, assessment design, presentation, assessment of the process, propriety of assessment, and consistency.

Toward a Quality Index for Evaluation Design

High-quality summative evaluations have certain characteristics. Marie-Louise Mares and Zhongdang Pan conducted a meta-analysis of the evaluations of Sesame Workshop's international coproductions and created a "quality index" that was devised to rate the quality of each study. The quality of a study was determined by the extent to which the study included random sampling or assignment at the individual level, multiple indicators for key variables, reliability assessment for key indexes, quality control in field operations, experimental or statistical controls, and a strong basis for causal inferences (panel design, between-group or pre-post experimental design). Mares and Pan emphasized the importance of multi-item measures that were reliable as essential to good summative research.[31]

Program Evaluation Standards

The Program Evaluation Standards were developed by a joint committee with members from three organizations: the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.[32]

> The Standards are organized according to five main attributes of an evaluation:
>
> - Utility—how useful and appropriately used the evaluation is
> - Feasibility—the extent to which the evaluation can be implemented successfully in a specific setting
> - Propriety—how humane, ethical, moral, proper, legal, and professional the evaluation is
> - Accuracy—how dependable, precise, truthful, and trustworthy the evaluation is
> - Metaevaluation—the extent to which the quality of the evaluation itself is assured and controlled.[33]

---

[31] Mares and Pan, 2013.

[32] See Donald B. Yarbrough, Lyn M. Shulha, Rodney K. Hopson, and Flora A. Caruthers, *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2011, cited in Mertens and Wilson, 2012, p. 23.

[33] Mertens and Wilson, 2012, p. 23.

Summary

This chapter discussed the general principles of presentation. Chief among these is to tailor the presentation of assessment results to the stakeholder. One should be asking, "What do stakeholders need to know to support the decisions they need to make?" Not every stakeholder wants or needs a report, and not every stakeholder wants or needs a briefing. Key takeaways related to the presentation of assessment results include the following:

- Identify, first, how stakeholders will use assessment results, and get them results in a format suited to those needs.
- Quantitative data supported by qualitative data can be very effective: The combination can help illustrate findings and provide context for the numbers.
- Narratives can be an excellent way to summarize assessment results, and those that explain the attendant theory of change and how well it is working in a nuanced context are even better. It is crucial to reemphasize that all assessments should make clear what data form their foundation and how confident stakeholders should be of the results.
- Building on the previous point, narratives also support data aggregation and the process of soliciting feedback from end users of assessment results by increasing stakeholders' and decisionmakers' understandings of what might be complex or opaque approaches to rolling up quantitative data.
- Stakeholders are not the only ones who stand to benefit from assessment data. Input, feedback, and guidance derived from the results should be shared with those who have contributed to the assessment process, as well as, when possible, those who are working on similar efforts.
- Assessors need to take care when aggregating assessments of individual efforts or programs. The lesson here is that, sometimes, the whole really is greater than the sum of its parts. The metaevaluation checklist included in Appendix A can be an effective tool for assessing assessments—specifically for summative evaluations of influence efforts.

Conclusions and Recommendations

This report is substantial, and each chapter has provided useful insights for the practice and planning of assessment and evaluation of DoD efforts to inform, influence, and persuade. Each chapter has its own summary that lists the key insights and takeaways from the discussion. These final conclusions reprise only the most essential of these numerous insights, those that are most intimately connected with the report's recommendations. These key conclusions are followed by recommendations.

Key Conclusions and Insights

Identifying Best Practices and Methods for Assessment

The best analogy for DoD IIP efforts is best practice in public communication (including social marketing), as the finest work in that sector combines the top insights from academic evaluation research but moves away from the profit metrics that appear in business marketing (which are poor analogs for DoD). However, all sectors contribute useful insights that can be integrated within the DoD IIP context—specifically via operational design and JOPP. This is a theme we revisit throughout this report; it is one thing to learn about best practices, but it is quite another to apply them. The best practices revealed the following lessons:

- Effective assessment requires clear, realistic, and measurable goals.
- Assessment starts in planning.
- Assessment requires an explicit theory of change, which is a stated logic for how the activities conducted are meant to lead to the results desired.
- To evaluate change, a baseline of some kind is required.
- Assessment over time requires continuity and consistency in both objectives and assessment approaches.
- Assessment is iterative.
- Assessment is not free; it requires resources.

A key takeaway is that both success and failure provide valuable learning opportunities for DoD IIP efforts and their assessment. (See Chapters One and Four for a full discussion.)

Why Evaluate? An Overview of Assessment and Its Utility

Before considering the process, data collection methods, and theories that underlie assessment, it is important to ask the simple question, "Why evaluate?" Myriad reasons for assessment connect to three core motives: to support planning, improve effectiveness and efficiency, and enforce accountability. These three motives roughly correspond to the three types, or stages, of evaluation: formative, process, and summative. One key insight is that assessment should always support decisionmaking, and assessment that does not is suspect. DoD requires IIP assessment to support planning, improvement, and accountability, but IIP efforts face unique challenges when it comes to meeting these requirements. To best support decisionmaking, assessment must be pursued with these challenges in mind. (See Chapter Two for a full discussion.)

Applying Assessment and Evaluation Principles to IIP Efforts

IIP assessment best practices can be found in all the sectors reviewed for this research (though, again, the best analogy for DoD IIP efforts comes from public communication, including social marketing). Long-term IIP assessment efforts, in particular, may not produce results within the time frame demanded by stakeholders. To resolve this problem, objectives can be nested, or broken into several subordinate, intermediate, or incremental steps. Doing so offers the opportunity to fine-tune the assessment process, identify failure early on, and provide stakeholders with valuable updates on incremental progress. (See Chapter Three for a full discussion.)

Challenges to Organizing for Assessment and Ways to Overcome Them

The research shows that organizations that conduct assessment well usually have an organizational culture that values assessment, as well as leadership that is willing to learn from (and make changes based on) assessment. Furthermore, assessment requires resources; experts suggest that roughly 5 percent of total program resources be dedicated to evaluation. A culture of assessment can facilitate the success of IIP efforts and the implementation of the processes described in subsequent chapters. (See Chapter Four for a full discussion.)

Determining What's Worth Measuring: Objectives, Theories of Change, and Logic Models

Good objectives are SMART: specific, measurable, achievable, relevant, and time-bound. Good IIP objectives specify both the target audience and desired behaviors. Theories of change allow planners and assessors to express assumptions as hypotheses, identify possible disruptors that can interfere with the generation of desired effects,

and, most important, determine where an effort is going awry if it is not achieving its objectives (and provide guidance on how to fix it). A fully explicit theory of change is particularly important in IIP assessment because—unlike kinetic operations—IIP efforts lack commonly held (and validated) assumptions. (See Chapter Five for a full discussion.)

From Logic Models to Measures: Developing Measures for IIP Efforts

The processes and principles that govern the development of valid, reliable, feasible, and useful measures can be used to assess the effectiveness of IIP activities and campaigns. There are two general processes for achieving this end: deciding which constructs are essential to measure and operationally defining the measures. Good measures should consider as many of the confounding and environmental factors that shape the outcome of interest as possible. Feasibility and utility can be in tension, however: Something may be easy to measure, but that does not mean it is useful to measure. (See Chapter Six for a full discussion.)

Assessment Design and Stages of Evaluation

The single most important property of assessment design is that it specifies the way in which the results will (or will not) enable causal inference regarding the outputs, outcomes, or impacts of the effort. The best designs are valid, generalizable, practical, and useful. However, there are tensions and trade-offs inherent in pursuing each of those objectives. Rigor and resources are the two conflicting forces in designing assessment. These two forces must be balanced with utility, but assessment design must always be tailored to the needs of stakeholders and end users. (See Chapter Seven for a full discussion.)

Formative and Qualitative Research Methods for IIP Efforts

Input from the SMEs interviewed for this study strongly suggests that DoD should invest more in qualitative and quantitative formative research to improve understanding of the mechanisms by which IIP activities achieve behavioral change and other desired outcomes. Initial investment in this area would pay off in the long run by reducing the chances of failure, identifying cost inefficiencies, and reducing the resource requirements for summative evaluation. (See Chapter Eight for a full discussion.)

Research Methods and Data Sources for Evaluating IIP Outputs, Outcomes, and Impacts

Good data are important for assessing outputs, outcomes, and impacts; even the most-complicated analytical tools cannot overcome bad data. Furthermore, contrary to prevailing wisdom, *good data* is not synonymous with *quantitative data*. Whether qualitative or quantitative, data should be validated using data from other collection methods whenever possible. (See Chapter Nine for a full discussion.)

Surveys and Sampling in IIP Assessments: Best Practices and Challenges

Despite known limitations, surveys are likely to remain one of the most prominent and promising tools in this area. The survey sample size and sampling methods must be carefully considered and matched to both the target audience and analytic requirements. Any survey effort should adhere to best practices for survey management, including engaging experts and local populations in survey design, vetting and tracking the performance of local firms, and maintaining continuity throughout the survey period. (See Chapter Ten for a full discussion.)

Presenting and Using Assessment

It is vitally important to tailor the presentation of assessment results to the needs of stakeholders and decisionmakers. Particularly central insights for DoD IIP efforts are as follows:

- Assessment needs advocacy, (better) doctrine and training, trained personnel, and greater access to assessment and influence expertise.
- IIP should be broadly integrated into DoD processes, and IIP assessment should be integrated with broader DoD assessment.
- Intelligence and assessment should be better integrated.

Presentation must strike the right balance between offering detailed data and analyses (so that results are convincing) and supporting stakeholder decisions in a way that avoids overwhelming stakeholders with data. Some of the most effective presentations mix quantitative and qualitative data, allowing the qualitative data to provide context and nuance. Summary narratives can be an effective way to synthesize and aggregate information across programs, efforts, and activities to inform efforts at the operational or campaign level. (See Chapter Eleven for a full discussion.)

Recommendations

Based on these conclusions and the more detailed insights and advice distilled throughout this report, we make several recommendations. This report contains insights that are particularly useful for those charged with planning and conducting assessment, but there is also an abundance of information that is relevant to other stakeholders, including those who make decisions based on assessments and those responsible for setting priorities and allocating resources for assessment and evaluation. Because assessment design, data collection, and the analysis and presentation of assessment results are all driven by the intended uses and users of the information produced, our recommendations are organized by stakeholder audience:

- DoD IIP assessment practitioners

- the broader DoD IIP community
- those responsible for congressional oversight
- those who manage DoD IO assessment reporting to Congress.

Although the recommendations presented here are targeted toward specific types of stakeholders, a recurring theme in our discussions of assessment challenges and practice improvement is the need for shared understanding across stakeholder groups. Therefore, points drawn from the experiences of one particular group are likely to prove informative for the others.

Recommendations for DoD IIP Assessment Practitioners

Our recommendations for assessment practitioners echo some of the most important practical insights described in the conclusions:

- Practitioners should *demand SMART objectives*. Where program and activity managers cannot provide assessable objectives, assessment practitioners should infer or create their own.
- Practitioners should *be explicit about theories of change*. Theories of change ideally come from commanders or program designers, but if theories of change are not made explicit, assessment practitioners should elicit or develop them in support of assessment.
- Practitioners should be provided with *resources for assessment*. Assessment is not free, and if its benefits are to be realized, it must be resourced.
- Practitioners must take care to *match the design, rigor, and presentation of assessment results to the intended uses and users*. Assessment supports decisionmaking, and providing the best decision support possible should remain at the forefront of practitioners' minds.

An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*, focuses more specifically on these and other recommendations for practitioners.[1]

Recommendations for the Broader DoD IIP Community

Our recommendations for the broader DoD IIP community (by which we mean the stakeholders, proponents, and capability managers for information operations, public affairs, military information support operations, and all other IRCs) emphasize how advocacy and a few specific practices will improve the quality of assessment across the community, but such efforts cannot be accomplished by assessment practitioners alone.

---

[1] Paul, Yeats, Clarke, Matthews, and Skrabala, 2015.

- DoD leadership needs to *provide greater advocacy, better doctrine and training, and improved access to expertise (in both influence and assessment)* for DoD IIP assessment efforts. Assessment is important for both accountability and improvement, and it needs to be treated as such.
- DoD doctrine needs to *establish common assessment standards.* There is a large range of possible approaches to assessment, with a similarly large range of possible assessment rigor and quality. The routine and standardized employment of something like the assessment metaevaluation checklist in this report (described in Chapter Eleven and presented in Appendix A) would help ensure that all assessments meet a target minimum threshold.
- DoD leadership and guidance need to recognize that *not every assessment must be conducted to the highest standard.* Sometimes, good enough really is good enough, and significant assessment expenditures cannot be justified for some efforts, either because of the low overall cost of the effort or because of its relatively modest goals.
- DoD should *conduct more formative research.* IIP efforts and programs will be made better, and assessment will be made easier. Specifically,
  - Conduct TAA with greater frequency and intensity, and improve capabilities in this area.
  - Conduct more pilot testing, more small-scale experiments, and more early efforts to validate a specific theory of change in a new cultural context.
  - Try different things on small scales to learn from them (i.e., fail fast).
- DoD leaders need to *explicitly incorporate assessment into orders.* If assessment is in the operation order, or maybe the execute order or even a fragmentary order, then it is clearly a requirement and will be more likely to occur, with requests for resources or assistance less likely to be resisted.
- DoD leaders should support the *development of a clearinghouse of validated (and rejected) IIP measures.* When it comes to assessment, the devil is in the details. Even when assessment principles are adhered to, some measures just do not work out, either because they prove hard to collect or because they end up being poor proxies for the construct of interest. Assessment practitioners should not have to develop measures in a vacuum. A clearinghouse of measures tried (with both success and failure) would be an extremely useful resource.

Recommendations for Congressional Overseers

To date, iterations of IO reporting to Congress have not been wholly satisfactory to either side (members of Congress and their staffers or DoD representatives). To foster continued improvement in this area, we offer recommendations for both, beginning with recommendations for congressional overseers.

- Congressional stakeholders should *continue to demand accountability in assessment*. It is important for DoD to conduct assessments of IIP efforts so that those that are not effective can be improved or eliminated and so that scarce resources are allocated to the most-important and most-effective efforts.
- Congressional demands for accountability in assessment must be *clearer about what is required and expected*.
- When refining requirements, DoD representatives must *balance expectations*. Assessment in this area is certainly possible and should be conducted, but assessment should not be expected to fill in for a lack of shared understanding about the psychosocial processes of influence. (Understanding is much more fully shared for kinetic capabilities, such as naval vessels or infantry formations.)

Recommendations for Those Who Manage DoD Reporting to Congress

To those who manage congressional reporting on the DoD side, we make the following recommendations.

- DoD reporting should *strive to meet the congressional desire for standardization, transition from output- to outcome-focused assessments, and retrospective comparison of what has and has not worked*. While these improvements are not trivial or simple, they are possible, and they are part of the congressional requirement that has been made clear.
- DoD reporting *must acknowledge that congressional calls for accountability follow two lines of inquiry, and must show how assessment meets them*. Congress wants to see justification for spending and evidence of the efficacy (traditional accountability), but it also wants proof that IIP activities are appropriate military undertakings. IIP efforts that can be shown (not just claimed) to be contributing to approved military objectives will go a long way toward satisfying both lines of inquiry.

Assessing Assessments: The Metaevaluation Checklist

This appendix provides more detail on the metaevaluation checklist discussed in Chapter Eleven. The checklist is also available as a Microsoft Excel file that accompanies this report on RAND's website.

This metaevaluation checklist is an appropriate tool for summative evaluations or for summative evaluations with a process evaluation component. It can be used for assessments of actual influence efforts, but it is not intended to support or enable efforts that do not have some form of influence as an outcome. Here, we describe the various sections of the checklist, which appears at the end of this appendix (Table A.1).

SMART Objectives

The first section of the checklist addresses SMART objectives, which are operational specifications of goals (as described in Chapter Five). The overarching question is whether the program or activity objectives have been clearly stated. SMART objectives are *specific*, *measurable*, *achievable*, *relevant*, and *time-bound*. In setting SMART objectives, planners should use strong action verbs, state only one purpose or aim per objective, specify a single end product or result, and specify a time frame. DoD needs to develop a broad set of SMART objectives expressed in terms of the specific behavioral change it wants to see, and it should specify tactics for achieving those objectives.

Explicit Theory of Change

The next section of the checklist asks whether the assessment included an explicit theory of change. This includes having a clear, logical connection between activities and expected outcomes, identifying and listing vulnerable assumptions or possible disruptors, and discerning elements of the process (inputs and outputs), outcomes, assumptions, and disruptors.

Measurement

The measurement section of the checklist asks about multiple indicators for most key variables, the validity and reliability of data, and whether measures were in place for changes in knowledge, attitudes, and behaviors.

Surveys

After measurement, the checklist asks whether a survey was used. If so, it attempts to determine whether the scope and scale of the survey matched the scope of the program or activity, whether the survey was consistent with previous surveys and contributed to trend analysis over time, and whether experts in cultural and social science fields were consulted in the design of the survey.

Analysis

The analysis section of the checklist asks whether key conclusions were supported by more than one method and about time horizons—whether there was sufficient time between action and assessment to expect change. This part of the checklist also asks whether conclusions followed logically from the data and analysis.

Assessment Design

After analysis, the checklist moves on to assessment design, which questions whether the actual assessment design involved those who execute the program or activities. This section includes probes about the extent to which assessment sought to assert a causal connection between the activities and the outcome. If it did, the logical follow-up is to identify whether the assessment involved an experimental or quasi-experimental design (with a control group). If it did not seek to assert a causal connection, did it track changes in activities and outcome over time (longitudinally)?

Presentation

The presentation portion of the checklist asks about such factors as the uncertainty or degree of confidence associated with the results, whether the presentation of the results included both quantitative results and narrative and contextual explanation, and whether the presentation avoided inappropriate aggregation or junk arithmetic.

Assessment Process

The assessment process section of the checklist is straightforward. Were practical procedures followed? Did the assessment accomplish what it set out to do? Were resources used efficiently?

Propriety of Assessment

For the propriety of assessment, transparency is an important issue. This section asks about the underlying data being made available to the appropriate parties and whether the measures that led to incentives or rewards for those executing the activities were based on good proxies. This section also connects back to the data collector and assessor roles and asks to what extent they were separated from the validator and integrator roles.

Consistency

The final section of the checklist focuses on consistency. Did the objectives remain the same from the previous reporting period? Were all data and measures collected in the previous reporting period also collected in this period and in the same way? If they were not, did an evolving understanding of the context and the theory of change necessitate a different approach?

**Table A.1
Metaevaluation Checklist**

Caveat: This checklist is appropriate only for summative evaluations or summative evaluations with a process evaluation component. In addition, this checklist is designed for assessments of actual influence efforts only, not supporting or enabling efforts that do not have some form of influence as an outcome.

Objectives

Are program/activity objectives clearly stated?	☐ yes	☐ partly	☐ no

Specific

Are the target actions taken as part of the program or activity clear?	☐ yes	☐ partly	☐ no
Are the target audiences specified?	☐ yes	☐ partly	☐ no

Who are the target audiences? (write in)

[ ]

Are the desired changes in the target audiences specified?	☐ yes	☐ partly	☐ no

What are the desired changes in the target audiences?

[ ]

Are incremental or intermediate objectives specified?	☐ yes	☐ partly	☐ no

What are the incremental or intermediate objectives?

[ ]

Measurable

Can desired outcomes be observed/measured?	☐ yes	☐ partly	☐ no
Can the degree of accomplishment/partial accomplishment or progress toward the goal be measured?	☐ yes	☐ partly	☐ no

Achievable

Are the objectives realistic? Could they actually be met?	☐ yes	☐ partly	☐ no
Is a threshold for success (or incremental success) specified?	☐ yes	☐ partly	☐ no

What is the threshold or criterion for success?

[ ]

Is a threshold for failure specified?	☐ yes	☐ partly	☐ no

What is the threshold or criterion for failure?

[ ]

Relevant

Do the objectives connect to and contribute to higher-level theater or campaign objectives?	☐ yes	☐ partly	☐ no

Table A.1—Continued

Relevant (cont.)			
Why are the desired changes sought? What broader campaign objectives do they connect to?			
Do subordinate objectives connect to and support broader objectives?	☐ yes	☐ partly	☐ no

Time-Bound			
Is an expected timeline for achievement of incremental and overall objectives specified?	☐ yes	☐ partly	☐ no
Is a time limit for completion (or achievement of benchmarks) specified?	☐ yes	☐ partly	☐ no
What are the time constraints/targets?			

Theory of Change/Logic of the Effort			
Does the assessment process include an explicit theory of change/logic of the effort?	☐ yes	☐ partly	☐ no
Is there a clear, logical connection between activities and expected outcomes?	☐ yes	☐ partly	☐ no
Are vulnerable assumptions or possible disruptors identified and listed?	☐ yes	☐ partly	☐ no

Process (inputs and outputs)			
Are activities tracked/measured?	☐ yes	☐ partly	☐ no
Have planned activities been compared with activities actually completed?	☐ yes	☐ partly	☐ no
Has the quality of activities/products been measured?	☐ yes	☐ partly	☐ no
Is there a complete accounting of what funds were spent and how?	☐ yes	☐ partly	☐ no
Is there a cost breakdown, matching outputs to spending?	☐ yes	☐ partly	☐ no

Outcomes			
Was progress toward the ultimate objective or explicit intermediate objectives measured?	☐ yes	☐ partly	☐ no
Were baseline data against which change can be measured collected?	☐ yes	☐ partly	☐ no
What were the baseline estimates?			

Table A.1—Continued

Outcomes (cont.)			
Was change measured against that baseline?	☐ yes	☐ partly	☐ no
How did the outcome of interest change relative to the baseline?			
Was change measured against the previous reporting period?	☐ yes	☐ partly	☐ no

Assumptions and Disruptors				
If progress toward objectives is less than expected, have disruptors/reasons for low yield been identified?	☐ yes	☐ partly	☐ no	☐ N/A
If progress toward objectives lags, can the reason be clearly identified as shortcomings in theory, shortcomings in execution, or a combination of both?	☐ yes	☐ partly	☐ no	☐ N/A
If progress toward objectives lags, did the theory of change/logic of the effort support the identification of challenges?	☐ yes	☐ partly	☐ no	☐ N/A
If not, were the theory of change/logic of the effort and resulting measures updated?	☐ yes	☐ partly	☐ no	☐ N/A

Measurement			
Are there multiple indicators for most key variables?	☐ yes	☐ partly	☐ no
Does the assessment rely on data other than the commander's subjective assessment?	☐ yes	☐ partly	☐ no
Are the data valid (collected through methods known to produce valid results or subject to a validation process)?	☐ yes	☐ partly	☐ no
Are the data reliable (collected through methods known to produce reliable results)?	☐ yes	☐ partly	☐ no

Are measures in place for			
Exposure of the target audiences to the effort?	☐ yes	☐ partly	☐ no
Exposure measures to capture recall and recognition?	☐ yes	☐ partly	☐ no
Changes in knowledge/awareness due to the effort?	☐ yes	☐ partly	☐ no
Changes in attitudes due to the effort?	☐ yes	☐ partly	☐ no
Changes in behavioral intention due to the effort?	☐ yes	☐ partly	☐ no
Changes in behavior due to the effort?	☐ yes	☐ partly	☐ no

Table A.1—Continued

Surveys				
Was a survey used?	☐ yes	☐ partly	☐ no	
Did the scope/scale of the survey match the scope of the program/activity? (If a local effort, was it a local survey?)	☐ yes	☐ partly	☐ no	☐ N/A
Was a representative sample obtained?	☐ yes	☐ partly	☐ no	☐ N/A
Were questions well written, and was their translation confirmed?	☐ yes	☐ partly	☐ no	☐ N/A
Were the questions pretested with locals?	☐ yes	☐ partly	☐ no	☐ N/A
Was there an audit of the survey process to prevent cheating and minimize errors?	☐ yes	☐ partly	☐ no	☐ N/A
Was the survey consistent with previous surveys, contributing to trend analysis?	☐ yes	☐ partly	☐ no	☐ N/A
Were experts in the local culture and in social science methods consulted in the survey design?	☐ yes	☐ partly	☐ no	☐ N/A
Is the survey consistent with previous surveys, contributing to trend analysis over time?	☐ yes	☐ partly	☐ no	☐ N/A
Was the survey firm thoroughly vetted and trained?	☐ yes	☐ partly	☐ no	☐ N/A
Were experts in the local culture and in social science methods consulted in the design of the survey?	☐ yes	☐ partly	☐ no	☐ N/A
Are the survey questions consistent over time?	☐ yes	☐ partly	☐ no	☐ N/A

Analysis			
Are key conclusions supported by more than one method?	☐ yes	☐ partly	☐ no
Does the assessment analyze trends over time?	☐ yes	☐ partly	☐ no
Is there sufficient time between the action and the assessment to expect change?	☐ yes	☐ partly	☐ no
Has the devil's advocate view been formally included in the assessment, allowing the least favorable interpretation of the data?	☐ yes	☐ partly	☐ no
Do conclusions follow logically from the data and analysis?	☐ yes	☐ partly	☐ no
Was optimism in the interpretation formally balanced by an outside audit or internal devil's advocate?	☐ yes	☐ partly	☐ no

Table A.1—Continued

Assessment Design			
Did the design involve those who execute programs/activities?	☐ yes	☐ partly	☐ no
Is the design sound? (If executed properly, it should produce a sound assessment.)	☐ yes	☐ partly	☐ no
Was the evaluation plan developed when activities were planned?	☐ yes	☐ partly	☐ no
Were sufficient resources set aside for adequate evaluation?	☐ yes	☐ partly	☐ no
Does the assessment seek to assert a causal connection between the activities and the outcome?	☐ yes	☐ partly	☐ no
If so, does the assessment involve an experimental or quasi-experimental design (with a control group)?	☐ yes	☐ partly	☐ no ☐ N/A
If not, does the assessment track changes in activities and outcome over time (longitudinally)?	☐ yes	☐ partly	☐ no ☐ N/A

Presentation			
Does the assessment report uncertainty/degree of confidence associated with the results?	☐ yes	☐ partly	☐ no
Does it meet the needs of key stakeholders?	☐ yes	☐ partly	☐ no
Is it presented in a way that is relevant to users?	☐ yes	☐ partly	☐ no
Was it completed and communicated on time?	☐ yes	☐ partly	☐ no
Does the presentation of results include both quantitative results and narrative/contextual explanation?	☐ yes	☐ partly	☐ no
Does the presentation of results include enough methodological information to be credible?	☐ yes	☐ partly	☐ no
Is the presentation of results free of distortion or errors?	☐ yes	☐ partly	☐ no
Does the presentation of results avoid inappropriate aggregation/"junk arithmetic"?	☐ yes	☐ partly	☐ no

Assessment Process			
Does the assessment follow practical procedures? Has it accomplished what it set out to do?	☐ yes	☐ partly	☐ no
Does the assessment use resources efficiently?	☐ yes	☐ partly	☐ no

Table A.1—Continued

Propriety of Assessment

Transparency: Underlying data have been made available to appropriate parties.	☐ yes	☐ partly	☐ no
Disclosure: Findings/results presented without censorship.	☐ yes	☐ partly	☐ no
Human rights and respect: Human subjects protections have been followed (if human subjects are involved).	☐ yes	☐ partly	☐ no	☐ N/A
Conflicts of interest have been avoided.	☐ yes	☐ partly	☐ no
Measures that lead to incentives or rewards for those executing the activities are based on good proxies.	☐ yes	☐ partly	☐ no
The assessment is being conducted outside the executing office.	☐ yes	☐ partly	☐ no
The assessment is being conducted by someone other than the executing contractor.	☐ yes	☐ partly	☐ no
Data collection and assessment roles are separate from validator/integrator roles.	☐ yes	☐ partly	☐ no

Consistency

Objectives remain the same from the previous reporting period.	☐ yes	☐ partly	☐ no
All data and measures collected in the previous reporting period were collected in the current reporting period and in the same way.	☐ yes	☐ partly	☐ no
If not, did evolving understanding of the context and theory of change/logic of the effort necessitate change?	☐ yes	☐ partly	☐ no	☐ N/A

For each box checked "no," provide an explanation:

```




```

Table A.1—Continued

And check all exceptions that apply:

There is a fully validated theory of change/logic of the effort in this context, so there is a minimal need to collect process and outcome data.	☐ yes
Not interested in causation; results based on correlation are sufficient.	☐ yes
The program/activity itself does not have an assessable objective, making assessing progress practically impossible.	☐ yes

Survey Sampling Models and Management

Sampling Models: Balancing Efficiency and Economy

There are multiple sampling models available to IIP planners to select a well-considered sample from a population. When selecting a survey model, researchers often have one of two goals in mind: efficiency or economy.[1] *Efficiency* refers to the goal of balancing the cost of a survey collection effort with the desire to obtain a sample that precisely estimates full population results. Those who prioritize efficiency seek to enhance the survey's cost-data precision ratio. By contrast, *economy* refers to the goal of minimizing the overall expense of data collection, with less concern for the cost-sample precision ratio. In this appendix, we describe a series of sampling models, categorized according to whether they more strongly address one or the other of these two goals.[2]

Sampling Models That Emphasize Efficiency

EPSEMs (equal probability of selection methods) typically emphasize efficiency. These designs include, but are not limited to, simple random sampling, systematic sampling, and stratified random sampling. These designs are based on the principle of probability-based sampling. This principle states that a sample that is selected from a certain population will be representative of that population if all members have an equal chance of being selected.[3] In actuality, samples are rarely perfectly representative of the population, even those drawn using EPSEM designs. However, samples using these designs are more representative of the population than are samples derived by other methods, such as convenience sampling (discussed in the next section).[4]

A *simple random sample* is a sampling model in which every individual in a sampling frame, or every individual in a population, has an equal chance of being selected to participate in a survey. This approach begins with establishing the sample size desired

---

[1] Crano and Brewer, 2002.

[2] Crano and Brewer, 2002.

[3] Babbie, 1990.

[4] Babbie, 1990.

for a particular survey. Individuals are then randomly selected from a population, per-haps drawn from a list of unique numbers assigned to every person in a population. Using a random-number table or a computer program that makes random selections, numbers are chosen that correspond to numbers assigned to individuals in the popula-tion. Simple random sampling requires an accurate and complete sampling frame, or list of population members, that can be used for random selection.[5] Unfortunately, a complete list of population members is often not available, especially in conflict envi-ronments, preventing researchers from obtaining a true simple random sample.[6]

A similar sampling design, known as *systematic sampling*, involves choosing every *k*th person in the sampling frame or from a list of all individuals in the population. For example, if there is a population of 1,000 people and the researchers would like a sample of 100, they might select every tenth person on the list. One limitation of this design is that, like a simple random sample, it requires a complete list of every individ-ual, or subset, in the population. Another limitation is that, if individuals are arranged in a particular order on the list (e.g., in a cyclical pattern), choosing every *k*th person would result in a sample that is not representative of the population. For example, World War II researchers sought to obtain a representative sample of military members by selecting every tenth person on a roster. However, the roster was arranged by squads of ten people, with the first people listed for each squad being sergeants. As such, the sample selected included only sergeants and was not particularly representative of the population.[7] To prevent this kind of error, researchers should examine population lists for patterns.

Stratified sampling represents a modification to the simple random sample and systematic sampling designs. It is used to obtain a greater degree of sample representa-tiveness from a population than the two previously described designs may allow. When collecting a random sample or systematic sample, there is a possibility that, by chance, certain groups or subsets of a population will not be included in the selected sample. To control for this and obtain more-precise numbers of people from certain groups in the population, researchers use a technique called *stratification*. There are different kinds of stratification designs. Generally, these designs involve dividing a population into strata, or groups, such as religious groups or sects. Then, participants are randomly selected from within each stratum to be in the survey sample. See Figure B.1 for a schematic diagram of a stratified random sampling process.

Another sampling model that, depending on how it is implemented, may be con-sidered an EPSEM design is *cluster sampling*. A list of all individuals in a population may not be available, particularly in a conflict environment where there has not been a recent census. However, the population may be grouped into subpopulations or sub-

---

[5]   Mertens and Wilson, 2012.

[6]   Author interview with Steve Corman, March 2013.

[7]   Babbie, 1990.

Figure B.1
Schematic Diagram of Stratified Random Sampling Process
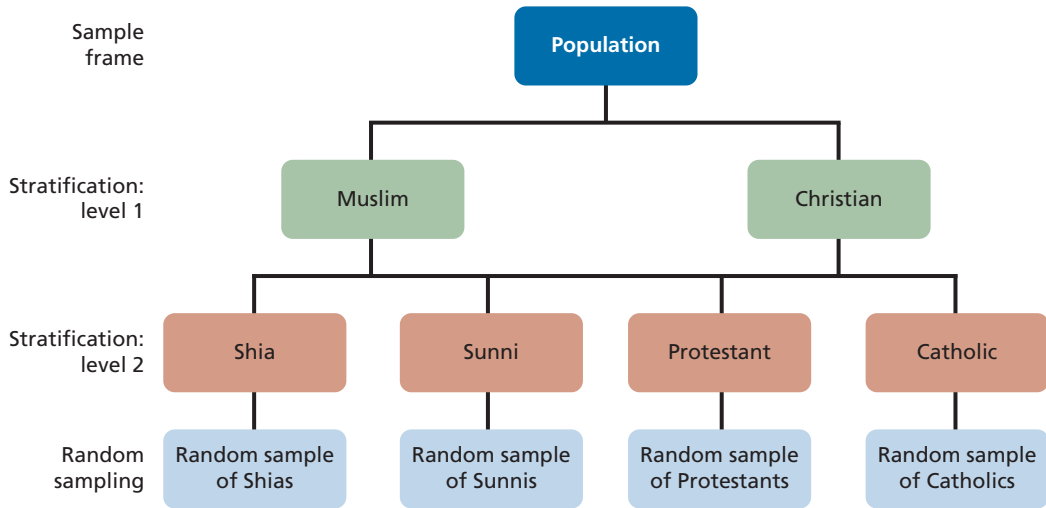
RAND *RR809/1-B.1*

groups, and a list of those subcomponents can be created or obtained. For example, the population of interest may be all individuals living in a particular city. Researchers could create or obtain a list of city blocks, then randomly select blocks to include in the sample and survey all individuals living within the selected blocks. This is considered an EPSEM design if all city blocks contain approximately the same number of people.

A design built from cluster sampling is known as *multistage sampling*. In this design, a cluster is sampled from previously created clusters. For example, one set of clusters consists of city blocks, and certain city blocks have been randomly selected for inclusion in the sample. Rather than collecting data from all individuals in the selected city blocks, researchers may randomly select certain households within the selected blocks to survey. Then, they may continue on to select certain individuals within the selected households. With this kind of sampling design, a Kish Grid approach is commonly used to assist researchers with selecting individuals within households (see Figure B.2). Notably, in certain environments, the researcher may not be able to select specific individuals to participate. For example, in Afghanistan, the head of the household may insist on serving as the survey participant.[8]

In the absence of a complete list of citizens in the country, ACSOR has used a multistage clustering design to collect data from Afghan citizens.[9] For example, ACSOR staff selected districts of interest and, using a list of villages in these districts, randomly

---

[8] Author interview with Emmanuel de Dinechin, May 16, 2013.

[9] Eles, Vasiliev, and Banko, 2012, pp. 11–12.

Figure B.2
Example of Multistage Sampling
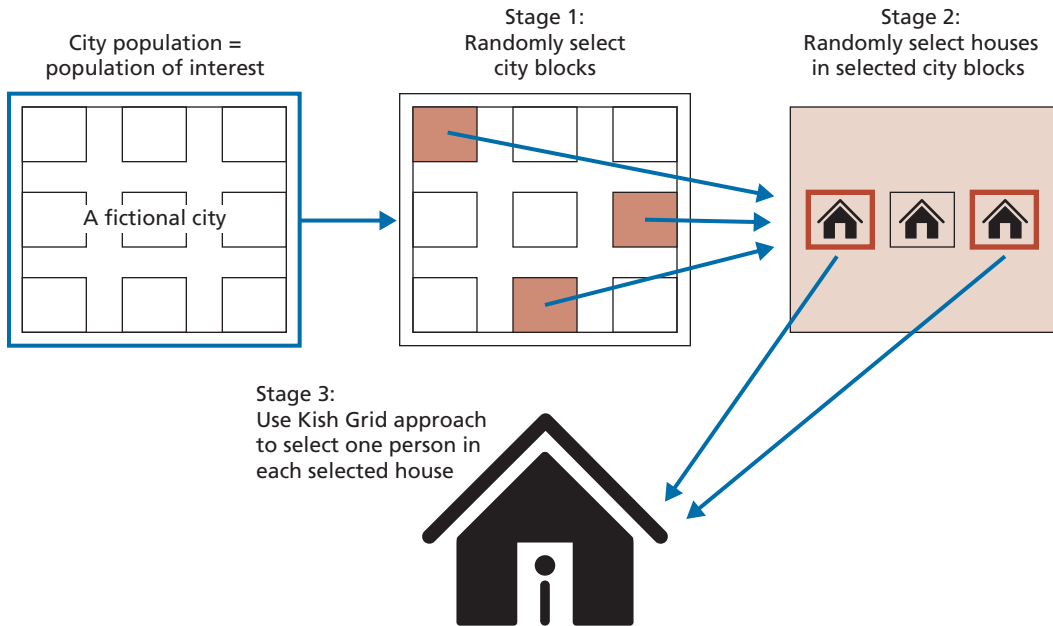
selected villages. Then, they randomly selected households from within the villages and used a modified Kish Grid approach to select one person in each household. To do this, they had to determine the geographic boundaries of districts and villages, as well as acceptable margins of statistical error arising from how data were collected and the number of people included. Table B.1 summarizes the characteristics and requirements of each of the sampling models discussed here.

Sampling Models That Emphasize Economy

Sometimes, obtaining samples that are representative of larger populations is possible or of interest to researchers. For example, it may not be possible to use simple random sampling in a complex conflict environment. As such, researchers may accept or want to collect a sample that is less representative of the entire population. This sample is more likely to be biased but may cost less to collect. Many nonprobability designs emphasize economy. Here, we briefly describe a few of these designs, which are summarized in Table B.2.

Convenience sampling, also known as accidental sampling or opportunity sampling, involves collecting data from individuals who are most available to participate in a survey. Researchers cannot and should not assume that these samples are representative of the full population. However, information from these samples can provide

Table B.1
Sampling Models That Emphasize Efficiency

Strategy	Definition	Requirements
Simple random sampling	Every individual in a sampling frame, or a list of all individuals in a population, has an equal chance of being selected to participate in a survey.	An accurate and complete sampling frame, or a list of population members
Systematic sampling	This involves choosing every kth person in the sampling frame or the list of all individuals in the population.	An accurate and complete sampling frame, or a list of population members, and awareness of any order effects
Stratified sampling	Generally, these designs involve dividing a population into strata. Then, participants are randomly selected from within each stratum.	Determine groups of interest and how to sample from groups
Cluster sampling	Units are randomly selected from within naturally occurring groups, like city blocks. Data are collected from all individuals in selected units.	A full list of clusters of interests
Multistage sampling	These designs use certain sampling strategies at certain levels. For example, start with randomly selected clusters then move to randomly selecting persons in selected clusters.	A full list of clusters of interests

SOURCE: Adapted from Mertens and Wilson, 2012.

Table B.2
Sampling Models That Emphasize Economy

Strategy	Definition
Convenience sampling	A sample is collected from individuals who are the easiest to contact. Information from samples collected using this method cannot be generalized to the larger population.
Quota sampling	Specific groups of interest are targeted, and convenience sampling is used to collect a preestablished number of surveys from these groups. Information from samples collected using this method cannot be generalized to the larger population.
Snowball sampling	A small group of individuals are interviewed and then asked to either recommend others or forward the survey/researcher information to others. Information from samples collected using this method cannot be generalized to the larger population.

initial information that may assist in adjusting a survey instrument or understanding theoretical relationships among variables of interest.

Quota sampling involves setting quotas for groups of individuals and then utilizing convenience sampling to meet these quotas. For example, researchers might want

to collect survey data from 100 Muslims and 100 Christians. Efforts will then be made to obtain these numbers, without randomly selecting individuals. Again, this approach may provide limited, initial information on these groups, but it should not be assumed that data collected using this method are representative of the population.

In snowball sampling, researchers begin by identifying a small group of individuals they would like to survey. After this initial group has completed the survey, the individuals are asked to provide the contact information for others they know and would recommend to participate. The researchers then contact these recommended individuals, and the process continues. Alternatively, initial participants may forward the researchers' information to individuals they know so that these other people can contact the researchers about participating. Again, snowball samples are not representative of the larger population, but they can be useful when trying to reach individuals who are difficult to access or find. Classic studies using snowball sampling have targeted homeless populations, a population that can be difficult to enumerate and survey using a random sampling frame.

Interpreting Results in Light of Survey Error and Sources of Bias

A significant limitation to the use of survey research in support of IIP assessment is that those responsible for assessment are inadequately trained in interpreting and applying survey data. To strengthen the link between research and decisionmaking, DoD should invest in building the capacity of assessors to interpret survey data through improved training in social science research methods. LTC Scott Nelson explains:

> It baffles me that we spend so much money to bring in organizations like Gallup but don't invest in sufficiently training our analysts such that they can interpret and apply and understand the limitations of these polls. They don't have to be experts, but they have to have a basic understanding of concepts like sampling error in order to apply the survey results to a valid and useful assessment of progress.[10]

Analysts must be sensitive to large error margins and sources of bias when interpreting results. District-level polls in Afghanistan have reported margins of error of approximately plus or minus 10 percent. This implies, as Stephen Downes-Martin explains, that if the reported change is less than 20 percent, analysts cannot be certain whether there has been an actual shift in attitudes.[11] When presenting results to decisionmakers, analysts should report survey error, potential sources of bias, and other

---

[10] Author interview with LTC Scott Nelson, October 10, 2013.

[11] Downes-Martin, 2011, p. 110.

limitations and should not present results that are not statistically significant (except to report that no significant change was observed).[12]

Some perceive an overconfidence in survey data in counterinsurgency assessment due to the "ORSA mentality" that quantitative data are inherently less biased and prone to error. Analysts with this mind-set have a tendency to treat polling data as "objective facts," losing sight of the very qualitative nature of survey instruments and the people responding to them.[13] As a result, some analysts tend to view small changes in polling results as indicators of effectiveness "even when not statistically significant or causally related to operations."[14]

Survey Management, Oversight, Collaboration, and Transparency

This section addresses concepts and best practices for the management and oversight of survey research in support of IIP activities, including contracting, quality assurance, and collaboration with organizations within and outside DoD. Survey programs are complex, with many moving parts. Successful implementation requires vigilant oversight throughout the entire process, input from experts and stakeholders, and a willingness to collaborate and be scrutinized. For a full treatment of these and related topics across assessment broadly writ, see Chapter Ten.

Contracting, Staffing, and Stakeholder Engagement in Support of Survey Research

Those responsible for contracting, staffing, or overseeing the administration of a survey in support of IIP assessment should consider the following recommendations.

- As early as possible and throughout the survey process, engage and involve cultural experts, survey research experts, stakeholders, and organizations familiar with the target audience, if possible through a survey working group. These experts should be leveraged in vetting local research firms, designing and testing the survey instrument, selecting the sample, and charting the logistics of the survey's administration. P. T. Eles and colleagues recommend creating a working group that includes local experts, stakeholders, technical advisers, and military planning staff. The group could meet regularly to review progress and make course adjustments.[15]
- Involve locals in the design of the survey instrument. Charlotte Cole could "not emphasize enough" the importance of involving representatives of the target audi-

---

[12] Eles, Vasiliev, and Banko, 2012, p. 38.

[13] U.S. Central Command, 2012.

[14] Author interview with Jonathan Schroden, November 12, 2013.

[15] Eles, Vasiliev, and Banko, 2012, p. 32.

ence in this process.[16] See the section "The Survey Instrument: Design and Construction" in Chapter Ten for more detail on soliciting local and expert feedback.

- Maintain continuity in survey management by repatriating deployed management. Based on their experience with opinion polling in Kandahar Province, Eles and colleagues argue that the costs of survey management turnover exceed the costs associated with reachback management. It is optimal, they conclude, to have reachback management with deployed operational analysts.[17]

- Data collectors must represent the demographics of the respondents. As addressed in the section "Challenges to Survey Sampling" in Chapter Ten, in some regions, female interviewers should interview female respondents to minimize response and nonresponse biases.[18] Depending on the environment, survey personnel may also need to be matched according to religion, ethnicity, age, and local dialect.[19] This is one of several reasons that data collection should be done by local researchers.[20] However, this requirement could be challenging to fulfill in operational environments such as Afghanistan, where it is difficult to find locals who fulfill niche demographic requirements who "can not only read but . . . can read aloud."[21]

- Thoroughly vet local research firms prior to awarding contracts. Contractors should look for "proof of professional qualifications, references, evidence of related work, and membership in relevant professional associations," and contracts should "include options for early termination."[22] According to Katherine Brown, the pressure to give contracts to the lowest bidder has created quality-control challenges.[23]

- Keep records of high- and low-performing research firms to maintain knowledge across contracting officers through staff rotations. Firms that have been caught cheating have been rehired on the same contracts because the incoming contracting officer was not briefed on their prior performance. Poorly performing or fraudulent firms can therefore compete for "the same contracts over again because no one is there in the long term to check quality."[24]

---

[16] Author interview with Charlotte Cole, May 29, 2013.

[17] Eles, Vasiliev, and Banko, 2012, p. 31.

[18] Author interview with Matthew Warshaw, February 25, 2013.

[19] Author interview with Amelia Arsenault, February 14, 2013.

[20] Author interview with Emmanuel de Dinechin, May 16, 2013.

[21] Author interview with Matthew Warshaw, February 25, 2013.

[22] Eles, Vasiliev, and Banko, 2012, p. 31.

[23] Author interview with Katherine Brown, March 4, 2013.

[24] Author interview with Katherine Brown, March 4, 2013.

- Make an up-front investment in building local research capacity so that there are sustainable research institutions when coalition personnel leave. Both the contracting office and the local researchers will benefit in the long run by saving the costs associated with redoing or recommissioning the survey.[25] Building local research capacity is discussed at length in Chapter Four.
- The initial contract with a survey research firm should cover one wave of polling and be flexible. The contract should permit changes to the survey design and should include early-termination clauses to prevent and manage cheating. Those commissioning the study may consider developing a pilot survey to test the instrument and the firm's research capabilities and to demonstrate the usefulness of the survey to stakeholders.[26]
- If the first survey is successful, subsequent contracts should seek to establish continuity in survey design and a long-term relationship between the contracting unit and the local research firm. These contracts should give consideration to building infrastructure for data collection over time (e.g., planning for quarterly surveys and managing a master data set with all survey waves).[27]

Data Verification and Quality Assurance to Manage Cheating and Errors in Data Collection and Reporting

Quality controls and data verification mechanisms are essential to generating credible and usable data from surveys conducted in conflict environments. Those managing contracts with survey research firms must be "very vigilant over the course of the whole process."[28] Eles and colleagues note that, in Afghanistan, entire polling programs have been terminated after issues with the data suggested fraud, cheating, or other errors in the data collection process.[29] Katherine Brown and other experts discussed cases in which it was discovered that interviewees were filling out questionnaires themselves or asking friends and family to respond.[30] ACSOR had to redo approximately 10 percent of its surveys due to suspect data generated by local subcontractors.[31]

This section reviews techniques and best practices for managing cheating, fraud, and other errors in the survey administration and data collection processes. Broadly, there are three approaches: field monitoring and statistical techniques to detect cheating or errors in a specific wave of survey data, external validation, and prevention

---

[25] Author interview with Amelia Arsenault, February 14, 2013.

[26] Eles, Vasiliev, and Banko, 2012, p. 31.

[27] Eles, Vasiliev, and Banko, 2012, p. 32.

[28] Author interview with Charlotte Cole, May 29, 2013.

[29] Eles, Vasiliev, and Banko, 2012, p. 36.

[30] Author interview with Katherine Brown, March 4, 2013.

[31] Author interview with Matthew Warshaw, February 25, 2013.

through the training and vetting of local research firms. Specific techniques include the following:

- Through in-person monitoring, or "spot checks," a certain percentage of survey administration staff are randomly observed to ensure that data are being collected properly.[32]
- GPS tracking and GPS-enabled smartphone or tablet-based survey instruments provide cost-effective, real-time monitoring, though they are not always practical in conflict environments. For example, when ACSOR tried to use GPS trackers to monitor survey administration, the researchers with the devices were detained by the Afghan Directorate of Security.[33]
- Researchers can build test questions built into survey instruments to which the answer is known or can be determined (e.g., telephone numbers). Analysts can then check for cheating by comparing interviewees' responses against the actual value.[34]
- "Customer callback" involves calling respondents and asking the same questions to check for consistency. Altai Consulting calls back approximately 15 percent of all respondents.[35] This approach is not feasible for anonymous surveys, however.
- Various statistical procedures can be used to check for suspicious patterns or responses (e.g., repetition, outliers). If survey administrators are cheating, "their answers tend to look very different."[36] Look for improbable outliers by "plotting raw responses by primary sampling unit."[37] Determine whether the unusual patterns or responses are associated with a single interviewer or are systemic.
- Comparing results on related items in cross tabulations, or "contingency tables," can verify that the data show expected relationships.[38]
- Analyzing the time spent to complete each survey can reveal cheating. If it only took a few minutes to complete a survey that should have taken 40 minutes, this may be evidence of cheating.[39]
- For longitudinal data collection, external validation can provide some assurances against cheating. Multiple waves of data allow researchers to identify unusual

---

[32] Author interview with Charlotte Cole, May 29, 2013; interview with Matthew Warshaw, February 25, 2013; interview with Emmanuel de Dinechin, May 16, 2013.

[33] Author interview with Emmanuel de Dinechin, May 16, 2013.

[34] Author interview with Emmanuel de Dinechin, May 16, 2013.

[35] Author interview with Emmanuel de Dinechin, May 16, 2013.

[36] Author interview with Matthew Warshaw, February 25, 2013.

[37] Eles, Vasiliev, and Banko, 2012, p. 36.

[38] Eles, Vasiliev, and Banko , 2012, p. 36.

[39] Author interview with Matthew Warshaw, February 25, 2013.

results and test the correlation between measured shifts in opinion and probable drivers of those shifts.[40] This approach highlights the value of long-term data collection for assessing progress over time. For example, the Asia Foundation has sponsored the same survey for the past eight years. So, "if something was really awry one year, it would be obvious."[41] External validation is discussed in more detail in the section "Using Survey Data to Inform Assessment" in Chapter Ten.

• Vetting and training local research firms prior to survey implementation and including termination options in contracts can minimize the risk of cheating and errors. Chapter Four provides more detail on this topic; also see the section "Contracting, Staffing, and Stakeholder Engagement in Support of Survey Research" in this appendix.

Depending on the severity of the problem, parts of the sample may need to be removed, or the entire survey may need to be redone. Altai Consulting typically removes 3–5 percent of its total sample for failure to meet its quality-control standards.[42] If cheating is localized—for example, limited to a certain interviewer—researchers can discard only those data.

Collaboration, Sharing, and Transparency

Opinion polling is often poorly coordinated across and within the U.S. government, international organizations, and NGOs. This section discusses challenges and opportunities in collaboration and transparency in survey research. Collaboration and transparency in IIP assessment is discussed more broadly in Chapter Four.

A widely perceived lack of transparency and "aversion to cooperation and sharing" create inefficiencies and duplication in survey research in environments like Afghanistan.[43] Several related and overlapping opinion polls are routinely conducted with no coordination and limited visibility. Rather than leverage work done by others, contracting offices and survey research firms "reinvent the wheel over and over again," because there is no incentive to share and collaborate.[44] It is often difficult to share data even across surveys owned by the same research firm or for the same client.[45] DoD would save resources and improve the quality of assessment if it did more to pool survey research resources within the department and to leverage ongoing survey research conducted by non-DoD agencies and actors. In addition to improving effi-

---

[40] Author interview with Matthew Warshaw, February 25, 2013.

[41] Author interview with Katherine Brown, March 4, 2013.

[42] Author interview with Matthew Warshaw, February 25, 2013.

[43] Author interview with Kim Andrew Elliot, February 25, 2013.

[44] Author interview with Amelia Arsenault, February 14, 2013.

[45] Author interview with Kim Andrew Elliot, February 25, 2013.

ciency, building a close relationship with other polling organizations could improve the quality of survey instruments. Questionnaires could build in complementary items that allow for comparisons between polls.[46]

To improve the quality and cost-effectiveness of survey research in conflict environments, it is worth considering the virtues of an omnibus or consolidated survey that could be used by several organizations with overlapping information requirements. Such an approach would provide an economical and fast way to assess media habits and sentiment on broad topics of interest at the national or regional level.[47] However, consolidated surveys are unlikely to fulfill specific and locally focused information requirements. Acknowledging that there are many surveys, Matthew Warshaw pointed out that each serves a different purpose and has a different sample frame, because different surveys are concerned with different target audiences. For example, "USAID and DoD influence activities are interested in very different things." In his view, a consolidated survey would undercut the quality of an instrument, bias the results, and risk creating a survey research monopoly with inefficient pricing: "To force that high level of coordination might not be worth the cost savings."[48] The Afghan Assessments Group attempted to organize a consolidated survey, but the effort was canceled because the various groups could not come to an agreement regarding which questions should be included.[49]

Where possible, survey data and results should be shared with other organizations.[50] To improve transparency, Maureen Taylor suggested that data be published to a public-use "clearinghouse" and that results be presented at periodic conferences.[51] This may not be wholly feasible in the DoD IIP context due to the sensitivity or proprietary nature of some of the data or the operations they support. However, according to Amelia Arsenault, transparency of any kind holds organizations accountable for performance: "If you can see the mistakes or successes of previous interventions, it goes a long way toward designing more-effective interventions in the future."[52] In her experience, resistance to transparency stems from a desire to "bury" evidence of failure to shield program managers from public scrutiny. Because of this dynamic, the evaluation system "is set up so that honesty is punished and not rewarded."[53]

---

[46] Eles, Vasiliev, and Banko, 2012, p. 33.

[47] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 40.

[48] Author interview with Matthew Warshaw, February 25, 2013.

[49] Author interview with Matthew Warshaw, February 25, 2013.

[50] Author interview with Thomas Valente, June 18, 2013; interview with Amelia Arsenault, February 14, 2013.

[51] Author interview with Maureen Taylor, April 4, 2013.

[52] Author interview with Amelia Arsenault, February 14, 2013.

[53] Author interview with Amelia Arsenault, February 14, 2013.

Evaluating Inform, Influence, and Persuade Efforts: Examples and Additional Resources

Even with an abundance of theories and key concepts available for assistance, the design, implementation, and assessment of an IIP effort can be challenging. Remembering and applying abstract principles may be difficult, so additional resources and concrete examples can help illustrate the major assessment principles discussed in this report.[1] This appendix describes in more detail the current doctrinal publications that guide the assessment of IIP efforts across DoD and provides several examples of similar efforts that have been implemented across disciplines. These examples offer IIP planners potential lessons regarding avenues to pursue and avoid when designing their own IIP efforts.

Assessment in Defense Doctrine

Although they have been criticized for being overly vague, DoD doctrinal publications describe and provide definitions of critical components of operational assessments.[2] That said, they offer helpful background on the reasons for assessment and encourage something of a common vocabulary for assessment that can be particularly useful in joint efforts or in aggregating individual efforts in support of broader campaigns.

A fundamental contribution that the publications listed here have made to the practice of good assessment is their emphasis on the importance of continuous evaluation throughout an IIP effort.

Field Manual 3-53: Military Information Support Operations

FM 3-53 provides guidance for U.S. Army MISO activities. Part of this guidance focuses on assessment, which is considered one of six core components of a MISO program.[3] Specifically, a MISO program should consist of psychological operations, and

---

[1] Chip Heath and Dan Heath, *Made to Stick: Why Some Ideas Survive and Others Die*, New York: Random House, 2007.

[2] Schroden, 2011.

[3] Headquarters, U.S. Department of the Army, 2013b.

planning should identify target audiences for these operations, key themes to promote and avoid, channels for dissemination, concepts that adhere to operational goals and paths to achieving those goals, and appropriate assessment approaches.

As described in FM 3-53, assessment is "the continuous monitoring and evaluation of the current situation, particularly the enemy, and the progress of an operation."[4] Continuous assessment involves MISO planners working with commanders to determine operational goals and establish informative and useful MOEs. This communication and the overall process are informed by current knowledge of target audiences, adversary influence on these audiences, and past and current data collection efforts.

After determining goals and appropriate measures, MISO planners should work to communicate the assessment requirements to those who can support MISO efforts. For example, working in collaboration with soldiers in the field, MISO planners can develop data collection plans that facilitate the assessment of changes in behaviors of interest among target audiences. To collect pertinent data, MISO planners must have a clear understanding of the commander's intent; must have knowledge and skills regarding current operations, general psychological operations, and assessment; and must work closely with multiple parties to ensure the appropriate implementation and assessment of a psychological operation.

Field Manual 3-13: Inform and Influence Activities

MISO serves as just one line of support for inform and influence activities (IIA).[5] Where FM 3-53 focuses on MISO organization and implementation, FM 3-13 focuses specifically on IIA. Although FM 3-53 and FM 3-13 describe overlapping aspects of assessments, FM 3-13 provides more-detailed guidance on the assessment of IIA, including methodologies for selecting high-value entities on which to focus (i.e., targeting).

For example, one methodology for selecting targets to address in IIA that is described in FM 3-13 is known as CARVER, a mnemonic for criticality, accessibility, recuperability, vulnerability, effect, and recognizability. This method involves assigning a value to each of six characteristics of a potential target, such that higher values are indicative of a more suitable target. Values are then assigned across potential targets to inform their selection. The CARVER method involves assigning quantitative values to qualitative characteristics of a target, which may facilitate assessment of the suitability of a target for certain operations, and the values assigned to each of the six aspects of CARVER are ordinal, representing ranked categories. Consequently, it cannot be assumed that the difference between two values is equal to the difference between two other values. This suggests that the intervals between the values assigned to each aspect cannot be interpreted in a clear and easily describable manner; this precludes assessors

---

[4]   Headquarters, U.S. Department of the Army, 2013b.

[5]   Headquarters, U.S. Department of the Army, 2013a. FM 3-13 uses *inform and influence activities* to refer to a particular component of IO, and those activities would fall under our general definition of *IIP*.

from calculating easily decipherable means and standard deviations using these values. (See the discussion of the misuse of the numerical values of ordinal rankings in Chapter Nine in the section "The Perils of Overquantification and Junk Arithmetic.")

In the CARVER rating scale, shown in Table C.1, criticality is the extent to which a particular target does or can harm an adversary's operations (1 = loss would not affect mission performance, 5 = loss would be a mission stopper). Accessibility concerns the ability to gain access to, or reach, a particular target (1 = very difficult to gain access, 5 = easily accessible, and away from security). Recuperability is the length of time an adversary will require to address the damage caused by eliminating or impairing a

Table C.1
CARVER Value Rating Scale

Value	C	A	R	V	E	R	Value
5	Loss would be a mission-stopper	Easily accessible; away from security	Extremely difficult to replace; long downtime (1 year)	Special operations forces definitely have the means and expertise to attack	Favorable sociological impact, neutral impact on civilians	Easily recognized by all, with no confusion	5
4	Loss would reduce mission performance considerably	Easily accessible outside	Difficult to replace with long downtime (< 1 year)	Special operations forces probably have the means and expertise to attack	Favorable impact; no adverse impact on civilians	Easily recognized by most, with little confusion	4
3	Loss would reduce mission performance	Accessible	Can be replaced in a relatively short time (months)	Special operations forces may have the means and expertise to attack	Favorable impact; some adverse impact on civilians	Recognized with some training	3
2	Loss may reduce mission performance	Difficult to gain access	Easily replaced in a short time (weeks)	Special operations forces probably have no impact	No impact; adverse impact on civilians	Hard to recognize, confusion probable	2
1	Loss would not affect mission performance	Very difficult to gain access	Easily replaced in a short time (days)	Special operations forces do not have much capability to attack	Unfavorable impact; ensured adverse impact on civilians	Extremely difficult to recognize without extensive orientation	1

SOURCE: Headquarters, U.S. Department of the Army, 2013a, Table 6-1.

NOTE: For specific targets, more precise target-related data can be developed for each element in the matrix.

target (1 = easily replaced in a short time [days], 5 = extremely difficult to replace, with a long downtime [one year]). Vulnerability addresses the ability of forces to attack the target and includes the resources and expertise of those forces (1 = special operations forces do not have much capability to attack, 5 = special operations forces definitely have the means and expertise to attack). The effects category addresses the impact that actions against a target will have on the populace (1 = unfavorable impact, an ensured adverse impact on civilians, 5 = favorable sociological impact, neutral impact on civilians). Finally, recognizability involves the extent to which a target can be identified easily by multiple entities and under different environmental conditions (1 = extremely difficult to recognize without extensive orientation, 5 = easily recognized by all with no confusion). Thus, the CARVER scale assists with the initial determination of which targets to pursue, and this general method of assessment may assist with initial decisionmaking in other kinds of efforts.

Joint Publication 5-0: Joint Operation Planning

JP 5-0 provides guidance regarding assessment, describing it as "the continuous monitoring and evaluation of the current situation and progress of a joint operation toward mission accomplishment."[6] Like the Army's field manuals described here, JP 5-0 addresses the necessity of ongoing assessment. It also emphasizes the use of assessment to determine current operational effectiveness in comparison with planned operational goals—a comparison that should inform subsequent adjustments to operations.

To design an effective operational approach, JP 5-0 notes the importance of communication between headquarters and commanders, between commanders and subordinate leaders, and between subordinate leaders and their staff. As noted in this doctrine, "While [combatant command commanders] and national leaders may have a clear strategic perspective of the problem, operational-level commanders and subordinate leaders often have a better understanding of specific circumstances that comprise the operational situation."[7] Communication among those involved in different aspects of an effort permits the clarification of objectives and the application of these objectives in a particular context. This informs the initial operational approach.

The initial operational approach is also informed by baseline assessments conducted at the level at which an operation may be targeted. For example, the communication between operational-level commanders and subordinate leaders should be informed by assessments of a local environment. To address a problem, it is important to determine the current state of the environment and root causes. Baseline assessments can identify which variables may help or hinder certain operations, thereby providing guidance in designing an approach. Variables that may be considered include available

---

6 U.S. Joint Chiefs of Staff, 2011a.

7 U.S. Joint Chiefs of Staff, 2011a.

resources and environmental conditions. Again, these data are collected at the potential levels of operations.

However, JP 5-0 also notes that the operational environment may experience significant changes before or during the implementation of an operational approach. These changes may contribute to a change in operational approach or a change in objectives. To determine what environmental changes have occurred and whether operational or objective adjustments are needed, informative assessments regarding the current state are needed, and they should be compared to the initial-state (baseline) assessments and desired end goals. These assessments should include the collection of MOPs that evaluate task performance and MOEs that assess the impact of tasks and operations. As noted in JP 5-0, "Commanders continuously assess the operational environment and the progress of operations and compare them to their initial vision and intent. Based on their assessment, commanders adjust operations to ensure objectives are met and the military end state is achieved." Assessments should inform both the initial approach and modifications to this approach. In other words, assessment should inform recommendations and actions.

Limitations of Current Defense Doctrine

Among DoD-driven efforts, operations assessments have been severely criticized for their inability to provide credible, concise, and cogent information regarding campaign progress. Multiple reinforcing issues may contribute to poor DoD-driven operations assessments. Jonathan Schroden has suggested that available doctrine regarding operations assessments is somewhat vague in its description of *how* to perform operational assessments.[8] In addition, those who must implement these assessments may not have the appropriate training to do so, so they may implement poor assessment procedures and create crude products. Further, commanders expect theoretical benefits that assessment practitioners may not be able to produce, which contributes to commanders losing interest in and reducing support for assessments. Collectively, these issues contribute to a cycle of assessment failure.

Schroden makes several suggestions for ending this cycle of failure. First, he proposes that an assessment advocate is needed in DoD to track current doctrine, collect knowledge, and identify areas of deficiency that must be addressed. Second, he suggests that current doctrine (described earlier in this appendix) should be adjusted to provide greater guidance regarding assessment implementation, not just general concepts and definitions. Third, Schroden advocates for a training pipeline that contributes to a cadre of knowledgeable and experienced assessment practitioners. Finally, he recommends that assessment practitioners shift from attempting to establish purely quantitative assessments and move toward using both quantitative and qualitative

---

[8]   Schroden, 2011.

assessments.[9] These recommendations suggest that poor assessments are not simply the result of poor metrics. Rather, there are systemic components that should be considered and addressed to improve DoD-driven operations assessments.

Assessment in Defense IIP Efforts

In addition to defense doctrine, the methods used can provide valuable information to those designing and assessing IIP efforts. A review of previous efforts can provide insight into best practices and highlight practices to avoid. This section offers some example efforts from a range of geographic locations, with several focused in Iraq and Afghanistan.

Information Operations Task Force in Iraq

The IOTF in Iraq involved various IIP efforts. To determine the impact of these efforts, planners incorporated assessments into the efforts' plans, and more than $10 million per year was dedicated to that purpose.[10] IOTF's program of self-assessment involved surveys, focus groups, atmospheric assessments, and media monitoring, conducted by the program contractor, Bell Pottinger.

Specifically, Bell Pottinger commissioned or conducted three types of surveys: media surveys, surveys assessing MOPs, and surveys assessing MOEs. The media surveys were conducted biannually to determine audience media consumption, or what the target audience was watching. Program performance was contingent on audience exposure to certain messages, so the MOP surveys were conducted each month to assess recall and the impact of certain efforts. Program effects were assessed based on changes in audience attitudes. The MOE surveys were conducted each month to track the political attitudes of the target audience. During these assessments, contractor staff conducted interviews with more than 320,000 audience members via 82 surveys. To complement the quantitatively focused assessments with qualitatively focused assessment, 1,100 focus groups were also conducted, along with in-depth interviews of more than 7,000 target-audience members.[11]

Despite the abundance of quantitative and qualitative information from these self-assessments, Bell Pottinger's client cited anecdotal evidence when describing the effectiveness of the IOTF program. For example, rather than referencing survey analyses or focus group statements, the client cited intercepted communications and quotes from political leaders as evidence. The outcome of the self-assessment program highlights the importance of addressing the needs of key personnel involved with an effort

---

[9]   Schroden, 2011.

[10]   Author interview with Paul Bell, May 15, 2013.

[11]   Author interview with Paul Bell, May 15, 2013.

and ensuring that these personnel know how to interpret potentially valuable assessment results.

Strategic Communication Assessment in Operation Iraqi Freedom

OIF began in March 2003 and supported at least two efforts, the Multi-National Force–Iraq (MNF-I) and U.S. Mission–Iraq (USM-I). Core components of these operations were strategic communication and assessment. COL Thomas Cioppa of U.S. Army Training and Doctrine Command described strategic communication as follows:

> Strategic communication can be defined as a comprehensive orchestration of actions, words, and images and requires monitoring, measuring, analyzing, and assessing. A methodological approach that is adaptable, flexible, and responsive is required to ensure desired effects are being achieved.[12]

In Iraq, strategic communication involved the use of television, radio, and print. This communication was improved through the use of assessments of effectiveness, continuous feedback provided to leadership regarding these assessments, and communication among subordinate and senior commanders regarding goals and acceptable methods. Thus, ongoing monitoring, measurement, analyses, and communication among personnel and leadership guided modifications to the strategic communication efforts.

For example, frequent polling provided information about the media sources most accessed by Iraqis, and media monitoring informed the stories or topics of interest to this group. To assist in making the large amount of information generated more comprehensible and useful to leadership, media monitoring data were organized according to four categories of primary interest: political, economic, diplomatic, and security. Further, messages containing erroneous and harmful information were also tracked and addressed. Finally, the results from media and polling efforts informed OIF messages and the channels through which these messages were disseminated.

Through continuous monitoring, MNF-I and USM-I personnel were able to maintain an understanding regarding the target audience of Iraqis. And by communicating with leadership, personnel were able to determine which communication channels and messages were of greatest interest to decisionmakers, which guided assessments and analyses. In turn, results from these analyses informed additional avenues to pursue and changes to be made to strategic communication efforts.

International Security Assistance Force Strategic Communication Assessment

Various U.S. military–led PSYOP (now MISO) campaigns have been undertaken in Afghanistan. The limited available information about these campaigns suggests

---

[12] Cioppa, 2009.

a lack of coordination among different U.S. military entities, shortfalls in planning and assessment, and low levels of understanding regarding Afghan communication channels—all of which may have hindered the success of these efforts.[13] Further, there is no central repository of data regarding efforts conducted in the country. For this reason, and because many of the efforts are classified, it is difficult to draw general conclusions from efforts conducted in Afghanistan. However, examples from specific efforts do provide some insight into operations in that country.

Like efforts in other locations, ISAF strategic communications assessments involved surveys and focus groups. However, planners also instituted additional assessments on top of these more traditional measures. In the United States, secret shoppers or mystery shoppers are used to monitor customer service and company compliance; individuals are hired to blend in with other customers and conduct certain transactions. This concept was applied to Afghanistan, where ISAF recruited local Afghan volunteers to use checkpoints and then provide information regarding their experiences back to ISAF headquarters.[14] This checkpoint evaluation process was ongoing, and ISAF personnel tracked trends in checkpoint transactions, allowing them to observe overall patterns and changes over time.

In addition, ISR assets were used to observe behavioral patterns among Afghan civilians, such as the number of people at local markets. The purposes of these assessments were to identify changes in behavior patterns and willingness to visit particular areas—an indication of the population's perception of security. One limitation of these assessments was that the owners of the ISR assets were reluctant to collect information on basic behavioral patterns rather than conducting the kinetic-focused assessments with which they were more familiar.[15]

In general, these types of data collection methods may be applied in other locations, thereby improving other efforts. Unfortunately, information regarding these innovative techniques may be lost or forgotten due to the absence of a central repository for assessment methods.

Military Information Support Operations in Libya

Although Iraq and Afghanistan have been the focus of multiple IIP and strategic communication efforts in recent history, several other campaigns have utilized similar approaches in other countries. For example, the organization and coordination of U.S. MISO activities began approximately one month prior to the March 2011 bombing campaign in Libya. This early start allowed for greater integration of MISO with the campaign. Thus, MISO messages could be produced in different languages and dis-

---

[13] Muñoz, 2012.

[14] Author interview with John-Paul Gravelines, June 13, 2013.

[15] Author interview with John-Paul Gravelines, June 13, 2013.

semination in coordination with kinetic operations.[16] During the 12 days of bombing operations in Libya, MISO personnel disseminated 50 messages.

MISO activities continued in the months following the 12-day campaign, with personnel disseminating approximately 200 additional messages. Many of these messages targeted the efforts of the regime. Anecdotal evidence suggests that these efforts may have been successful: The Libyan regime responded to several MISO messages by refuting their claims. However, quantitative assessments of overall effectiveness are not available. This may be because the MISO effort in Libya took the form of a series of distinct messages rather than a coordinated effort. Further, a lack of data on changes in attitudes and behaviors of interest certainly hindered the assessment of the initiative's effects.[17]

U.S. Northern Command Influence Assessment Capability

Key IO tasks at USNORTHCOM involve building partner capacity.[18] To address the need for assessments of the effectiveness of these activities, the command established an assessment team, with a director, deputy, branch chiefs, research staff, and analysis staff. This team was tasked with evaluating SME exchanges. To facilitate this task, USNORTHCOM provided the team with a guiding methodology—a participant observation methodology—and topics of interest. Military objectives also guided the subgoals and research design created with the team. Thus, assessment of the program involved collecting information that could specifically address topics of interest and whether subgoals were achieved.

This design and assessment process was used to train Mexican military personnel in IO, with the general objective to assist the Mexican military in addressing transnational criminal organizations. To assess the effort's effectiveness, the team studied interactions among SMEs each day of the training. These data provided insight into how the audience's knowledge level may have changed during the training and how the Mexican military staffed and planned IO processes. These observations were complemented by surveys. As evidence of the utility and broad applicability of this assessment process, other commands have begun to use this process.[19]

---

[16] Geoffrey Childs, "Military Information Support to Contingency Operations in Libya," *Special Warfare*, Vol. 26, No. 1, January–March 2013.

[17] Childs, 2013.

[18] Author interview with LTC Scott Nelson, October 10, 2013.

[19] Author interview with LTC Scott Nelson, October 10, 2013.

Assessment in Business and Marketing Efforts

In the realm of business and marketing, several principles and approaches may assist with IIP efforts. These efforts often focus on increasing profits, rather than creating social change. Therefore, these should be considered carefully before being applied, as appropriate, to defense IIP efforts.

The Barcelona Principles: Communicating and Assessing Public Relations Efforts

Public relations efforts can involve communicating specific messages to target audiences, changing audience knowledge or attitudes, or meeting company or client objectives.[20] To meet these objectives and determine the effectiveness of an effort, it is common practice in public relations to establish specific goals, develop a theory of change, and utilize high-quality research methods and assessments.[21] The international public relations community developed the Barcelona Declaration of Measurement Principles to formalize these practices and to guide planning and measurement.[22]

The Barcelona Principles consist of seven voluntary guidelines:

1. Importance of goal setting and measurement: A public relations effort should set clear goals that account for audience reach, audience awareness, audience comprehension, and whether audience behaviors change. Measures could include the number of articles on a topic of interest (reach), audience recollection (awareness and comprehension), brand loyalty (attitudes), and purchase decisions (behavior).

2. Measuring the effect on outcomes is preferred to measuring outputs: While it is important to examine how outputs affect outcomes, this guidance does not go far enough for defense IIP efforts and therefore should be approached with caution. Measuring outcomes is important, but if outcomes are not those desired, measuring outputs may help identify why. Assessments of IIP efforts benefit greatly from an understanding of *how* a goal was met and whether changes can be attributable to certain aspects of an effort.

3. The effect on business results can and should be measured where possible: In the industry, organizational impact assessments address such issues as market share and stock price. In the IIP context, assessments show how individual efforts provide value to an overall campaign.

---

[20] Author interview with David Rockland, March 2013.

[21] David Michaelson and Sandra Macleod, "The Application of Best Practices in Public Relations Measurement and Evaluation Systems," *Public Relations Journal*, Vol. 1, No. 1, October 2007.

[22] See Ketchum Global Research and Analytics, undated.

4. Media measurement requires quantity and quality: In this context, *quantity* refers to the number of messages, and *quality* refers to the characteristics of these messages (e.g., negative, positive, or neutral).

5. Advertising value equivalents are not the value of public relations: Common advertising value equivalents are the cost of a centimeter of print space in a newspaper or a second of radio talk time, but they do not capture message variety or placement, and they cannot measure impact in newer communication channels, such as social media.[23]

6. Social media can and should be measured: Social media analysis should be treated similarly to conventional media analysis, with a focus on both quantity and quality. There are limitations to social media assessment, however, and it should be used in conjunction with other assessment approaches.

7. Transparency and replicability are paramount to sound measurement: Clearly documenting the assessment process will increase perceptions of validity, and making the results available will allow others to learn from and build on the data.

Advertising Analytics: Assessing Performance and Effects

New options for measuring the influence of various marketing efforts are available but underused by business marketers. These approaches could hold value for a variety of IIP efforts.

Traditionally, marketers have used a "swim-lane" approach to analyzing the performance of their advertising activities, considering the amount of money spent and advertising and the amount of revenue earned through individual advertising channels.[24] For example, markets might compare the cost of an email campaign and the amount of revenue generated from people clicking on a link embedded in the message. These numbers would then be compared with similar numbers on the amount of money spent on television advertisements and the amount of revenue generated. One problem with this approach is that it assumes that advertising effects can be isolated, such that exposure to television commercials would not influence email clicks. Another problem with this approach is that it fails to consider the actions of competitors and the channels that competitors are using to communicate information to the same potential customers.

Wes Nichols, cofounder and CEO of MarketShare, recommends a new form of advertising analytics involving three types of activities.[25] The first, attribution, involves gathering quantitative data on the amount spent on each advertising activity (by the

---

[23] For more on evaluating dissemination approaches, see the next section, "Advertising Analytics: Assessing Performance and Effects."

[24] Wes Nichols, "Advertising Analytics 2.0," *Harvard Business Review*, March 2013.

[25] Nichols, 2013.

firm responsible for the message and by competitors) and tracking customer behaviors over time. Companies are often unaware that they already have the information needed to conduct these analyses because it is maintained in different databases managed by different departments, such as customer service or sales. The second type of activity is optimization, which involves a kind of war-gaming in which different variables are adjusted (e.g., money spent on certain advertising channels internally and by competitors) and the expected impact on customer behaviors is determined. The final activity in new advertising analytics is allocation, in which the optimization analyses are applied to business behaviors so that appropriate funds are allocated to appropriate markets on a regular basis. Under this model, further analysis incorporates the new data, helping analysts track the need for changes in a timely manner.

For this new approach to analytics to be successful, it must be embraced by a firm's staff and leadership. This means that the firm must be organized for this type of assessment, with, for example, a staff member in place to spearhead the effort, communication across departments to consolidate data already being collected, a common understanding of assessment goals, and a process in place to test approaches on a small scale before they are implemented.

Assessment in Public Communication and Social Marketing Efforts

Public communication, or social marketing, builds on the concepts of traditional, or commercial, marketing. Public communication has been defined in many different ways, but the following is a commonly used definition:

> Social marketing is the application of commercial marketing technologies to the analysis, planning, execution, and evaluation of programs designed to influence the voluntary behavior of target audiences in order to improve their personal welfare and that of their society.[26]

Thus, the behavior change that social marketing seeks to induce differs from that of traditional marketing. Rather than encouraging individuals to purchase a product, for example, social marketing seeks to foster prosocial behaviors.[27] Consequently, social

---

[26] Alan R. Andreason, "Social Marketing: Its Definition and Domain," *Journal of Public Policy and Marketing*, Vol. 13, No. 1, Spring 1994.

[27] Stewart I. Donaldson, "Theory-Driven Program Evaluation in the New Millennium," in Stewart I. Donaldson and Michael Scriven, eds., *Evaluating Social Programs and Problems: Visions for the New Millennium*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2003.

marketing efforts may be more informative than commercial marketing efforts for IIP planners.[28]

Social marketing efforts can vary greatly by location and desired behavior change, but they share some characteristics. For example, the Joint United Nations Programme on HIV/AIDS (UNAIDS) states that "research is fundamental to effective social marketing and behaviour change." In this case, knowledge of the target population's needs, values, and desires permits the identification of areas that may be addressed by social marketing—and the necessary data collection can be bolstered by working in collaboration with local organizations.[29] Researchers have also emphasized that successful social marketing campaigns are based on established theories of persuasion, not untested beliefs.[30] In this section, we briefly present a series of concrete examples of social marketing efforts across sectors.

Sesame Workshop International Coproductions: Designing and Assessing Efforts

Sesame Street is a children's educational television program that is broadcast in more than 130 countries. The specific version of *Sesame Street* shown in a particular country may be tailored to the local context. However, they share common themes and aim to improve children's cognitive outcomes in the areas of literacy and numeracy, geography and cultural knowledge, the environment and health, social reasoning, and prosocial behavior and attitudes toward members of different groups. Recent analyses suggest that exposure to *Sesame Street* programming is connected with increases in these cognitive outcomes. These results were derived from meta-analyses of different studies conducted by different groups using different methods in different countries.[31] This suggests that the program's effects are not limited to a particular context or research design.

In describing the approach used by *Sesame Street* to design and assess its programming, Charlotte Cole, senior vice president of global education at Sesame Workshop, noted that much of the program's success was due to collecting relevant data and using these data to inform subsequent programming. Specifically, formative research is conducted in-house, with educational specialists and researchers evaluating products as they are being developed. Then, multiple research efforts are conducted after program implementation. According to Cole,

---

[28] Tim A. Clary, *USAID/Haiti: Social Marketing Assessment, 2008*, Washington, D.C.: Global Health Technical Assistance Project, 2008.

[29] Joint United Nations Programme on HIV/AIDS, *Condom Social Marketing: Selected Case Studies*, Geneva, Switzerland, 2000.

[30] William D. Crano, *Lessons Learned from Media-Based Campaigns, or, It Takes More Than Money and Good Intentions*, Claremont, Calif.: Claremont Graduate University, 2002.

[31] Mares and Pan, 2013.

At Sesame, we advocate for a "compendium of studies," including a mix of qualitative, experimental, and quasi-experimental designs, that look at naturalistic versus contrived experimental conditions. No single design will tell the full picture. The key is to have as many studies as possible and look across studies to see patterns emerge. You can build a story when you have multiple methods converge.[32]

BBC Media Action: Using a Theory of Change

BBC Media Action uses social marketing to address international poverty by focusing its efforts in three areas: governance and rights, health, and resilience and humanitarian response. These themes are addressed at four different levels of change: systems (e.g., social and political), organizations (e.g., nonprofit and commercial), practitioners (e.g., medical professionals), and people. The approaches to addressing these themes include traditional mass media channels, interpersonal communication, and social media. As such, BBC Media Action has a broad theory of change, which can be tailored to specific contexts.

According to Kavita Abraham Dowsing, director of research at the BBC Media Action, "Research is ingrained in the DNA of the organization."[33] Focusing on the three themes of interest, BBC Media Action collects self-report data on knowledge, attitudes, and behaviors. The specific measures used are based on the logistical frameworks of specific efforts, and the organization employs local research staff in the countries in which it operates. These individuals may be mentored by research personnel from London, but the local researchers collect the data. Although the model requires intensive monitoring, it serves to promote local research capacity.

Afghan Media in 2010: Understanding the Local Context

USAID is a U.S. government agency that works to address poverty and other issues in multiple countries. One component of USAID's efforts is the use of media to assist with social marketing, and one country on which the agency has focused its efforts is Afghanistan. To better understand the country context and inform subsequent efforts, USAID contracted with Altai Consulting to study Afghan media and public perceptions.[34] Using research tools developed in collaboration with USAID, Altai Consulting collected both qualitative and quantitative data at the national level and from high-priority districts.

Emmanuel de Dinechin, founder and lead partner at Altai Consulting, briefly summarized the results:

---

[32] Author interview with Charlotte Cole, May 29, 2013.

[33] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[34] Altai Consulting, 2010.

The research demonstrated that well-designed programs using locally trusted experts, placed on the right local media, were likely to have good local buy-in and have some impact on communications, individual opinions, and collective decisionmaking processes.[35]

According to Altai Consulting's 2010 report, the results showed a large increase in the number of media outlets in Afghanistan over a period of several years, many of which were created to promote certain religious or political interests. The study also revealed that most media efforts had focused on urban areas in Afghanistan and found that Afghans view television and radio positively and perceive media as a worthwhile avenue to promote education and inform people of government actions.[36] The results point to effective ways to reach Afghan populations and address what they perceive as their immediate needs.

Health Behavior Efforts

Many social marketing efforts involve campaigns to address health-related behaviors, including the promotion of healthy eating, physical activity, children's health, safe sex, and HIV awareness and prevention. As such, social marketing is widely used for public health campaigns.[37]

Egyptian Television Minidramas: The Need for Well-Informed Efforts

In Egypt, televised health campaigns have been used for years; television watching is a popular national pastime, it is accessible to those who are illiterate, and it has been shown to be an effective route for information communication. In the 1980s and early 1990s, televised health messages in Egypt shifted from one-minute messages to complex, multiepisode programs. However, this shift was not well informed by theories of persuasive communication, formative research, or summative research. For example, the multiepisode health campaigns often attempted to communicate several different health messages, which may have confused audiences and reduced the potential impact of the campaigns.

In a study of Egyptian television minidramas, researcher Sandra Lane recommended that program producers tailor their messages by identifying the needs and preferences of their target audiences.[38] Most research on persuasive communication has been conducted in Western countries, and the results of these studies may not be applicable to a different local context.

---

[35] Author interview with Emmanuel de Dinechin, May 16, 2013.

[36] Altai Consulting, 2010.

[37] Grier and Bryant, 2005.

[38] Sandra D. Lane, "Television Minidramas: Social Marketing and Evaluation in Egypt," *Medical Anthropology Quarterly*, Vol. 11, No. 2, June 1997.

Tú No Me Conoces: Tailoring Efforts to the Local Context

Tú No Me Conoces (You Don't Know Me) was an eight-week health campaign implemented to address health behaviors among Latinos living along the California-Mexico border.[39] The campaign's goals included raising HIV/AIDS risk awareness and promoting HIV testing. Previous research suggested that the campaign's target audiences listened to the radio more often than they read the newspaper or watched television, so planners developed one-minute radio ads that aired for eight weeks.

Local organizations developed several potential advertisements, which were tested with focus groups from the audience of interest before being implemented. The messages included a toll-free telephone number, the URL for the campaign's website, and the locations of local health clinics.

To assess the efficacy of the campaign, data were collected on call history, website visits, and testing activity at the clinics referenced in the radio ads. Most of those who called the toll-free number were able to recall the Tú No Me Conoces campaign, and most of those who visited the website accessed it directly rather than via another website, suggesting that they learned of the site from the radio advertisements. Half of the clinics referenced in the campaign saw an increase in the number of HIV-test requests during the campaign. Further, of those who agreed to participate in a media survey at the local clinics, 30 percent could recall the campaign's message. By studying the target audience, developing messages with local organizations, pretesting the advertisements, and monitoring changes in behavior, the campaign was able to demonstrate its effectiveness.

Jeito Campaign: Failure to Take into Account the Local Context

In Mozambique, many women rely on sex work to provide income for themselves and their families.[40] Unprotected sex is common and has contributed to the spread of HIV in the country; in some areas, nearly 20 percent of the population is HIV-positive.[41] Many international organizations and governments, including the United States, have used social marketing approaches to promote behaviors to help reduce the spread of HIV in Mozambique. One long-term effort was known as Jeito, a campaign implemented by Population Services International, a U.S. NGO. One component of the Jeito campaign involved distributing at reduced cost and promoting the use of an eponymously named condom brand. Indicators, such as increased sales of Jeito-brand condoms, seemed to suggest that the campaign had been a success. However, some have questioned the campaign's impact on communities and the resulting perceptions

---

[39]   Alisa M. Olshefsky, Michelle M. Zive, Rosana Scolari, and María Zuñiga, "Promoting HIV Risk Awareness and Testing in Latinos Living on the U.S.-Mexico Border: The Tú No Me Conoces Social Marketing Campaign," *AIDS Education and Prevention*, Vol. 19, No. 5, October 2007.

[40]   James Pfeiffer, "Condom Social Marketing, Pentecostalism, and Structural Adjustment in Mozambique: A Clash of AIDS Prevention Messages," *Medical Anthropology Quarterly*, Vol. 18, No. 1, March 2004.

[41]   Adam Graham-Silverman, "Fighting AIDS in Mozambique," *Slate*, May 31, 2005.

of widespread behavior change.[42] For example, those with limited economic resources may not have purchased the condoms at all, whereas those already using condoms may have purchased more of them. In addition, respondents to surveys regarding sexual behavior may have felt disinclined to provide honest responses. An overreliance on sales figures and survey responses as measures of effectiveness may contribute to a misleading picture of campaign impact.

The Jeito campaign was implemented when Mozambique was restructuring its federal programs to reduce government spending. This contributed to a reduction in services targeting the poor, including decreases in the availability of public-sector health services. The economic effect of government cutbacks may have contributed to changes in behavior, such as an increased reliance on sex work for income among the poor, and the reduction in services may have increased the chances for HIV to spread unchecked. The Jeito campaign failed to take into account the potential impact of these structural changes and did not include local communities in its message development. Furthermore, religious groups sought to address the spread of HIV by promoting fidelity and family sanctity and, countering a primary message of the Jeito campaign, by discouraging the use of condoms, which they associated with promiscuity and immorality. At the same time, the Jeito campaign was using sexually suggestive slogans and images to encourage condom use. These conflicting messages angered religious leaders.

To address the limitations of this campaign, planners should have reached out to a wide range of organizations and encouraged community participation during the campaign's development. The social context and structure should also inform efforts, and it should not be automatically assumed that prepackaged approaches can be implemented with minimal changes in new contexts. Finally, a diversity of measures to assess effectiveness and a rigorous comparison across different measures may provide clearer information about a program's impact.

Other Social Marketing Effort Examples

Many additional examples can inform social marketing–related efforts. For example, the Institute for Public Relations has created a series of reports on public relations techniques used internationally.[43] Similarly, the University of Southern California's Lear Center has conducted extensive research on prosocial media effects, which suggests that entertainment media can be used to influence attitudes and behaviors. This research is

---

[42] Pfeiffer, 2004.

[43] Judy Turk VanSlyke and Linda H. Scanlan, *Evolution of Public Relations: Case Studies of Nations in Transition*, Gainesville, Fla.: Institute for Public Relations, 1999; Judy Turk VanSlyke and Linda H. Scanlan, *Evolution of Public Relations: Case Studies of Nations in Transition*, 2nd ed., Gainesville, Fla.: Institute for Public Relations, 2004; Judy Turk VanSlyke and Linda H. Scanlan, *Evolution of Public Relations: Case Studies of Nations in Transition*, 3rd ed., Gainesville, Fla.: Institute for Public Relations, 2008.

available online and examines a variety of media campaigns.[44] Other sources of information on social marketing efforts include scholarly articles, such as Sonya Grier and Carol Bryant's review of social marketing as applied to public health efforts.[45]

Assessment in Public Diplomacy

Public diplomacy involves communicating with foreign audiences in an attempt to persuade them on matters of international concern. More specifically, it has been described as "the process by which international actors seek to accomplish the goals of their foreign policy by engaging with foreign publics."[46] In this section, we review some recommendations for the development and assessment of public diplomacy efforts.

Public Diplomacy Frameworks: Conceptualizing Evaluation

NATO's Joint Analysis and Lessons Learned Centre (JALLC) has developed an extensive framework for how to develop, plan, evaluate, and communicate the results of public diplomacy efforts.[47] This guidance offers examples of efforts in two categories: engagement with individuals or groups (e.g., conferences for delegates) and mass communication (e.g., traditional media activities). Engagement with individuals or groups may help build relationships with key influencers, whereas mass media efforts are implemented to influence a larger audience.

These two categories entail distinct impact, outcome, and output objectives, and although there is some overlap, different research methods and tools are associated with the evaluation of each type of effort. For example, an IIP planner may seek to persuade political votes regarding funding for a particular organization. A group engagement activity to address this impact objective may be a conference for influential delegates. Evaluation of this effort may include a formative evaluation in the form of face-to-face interviews with delegates, an output evaluation in the form of conference exit polls, and an impact evaluation in the form of media content analysis. Another IIP planner may seek to influence mass public opinion. To address this mass media objective, the planner may turn to a traditional media channel, such as radio advertisements. Evaluation of this effort may include a formative evaluation in the form of broadcast media monitoring and analysis, an output evaluation in the form of omnibus surveys, and an impact evaluation in the form of observation. Of course, if possible, efforts should

---

[44] Mandy Shaivitz, *How Pro-Social Messages Make Their Way Into Entertainment Programming*, Los Angeles, Calif.: Council for Excellence in Government and USC Annenberg Norman Lear Center, 2003.

[45] Grier and Bryant, 2005.

[46] Nicholas J. Cull, "Public Diplomacy: Taxonomies and Histories," *Annals of the American Academy of Political and Social Science*, Vol. 616, No. 1, March 2008.

[47] NATO, Joint Analysis and Lessons Learned Centre, 2013.

involve multiple data collection methods and tools for each kind of evaluation (formative, output, and impact).

Another framework for public diplomacy evaluation, developed by James Pamment at the University of Texas, provides guidance for assessment within a given social context.[48] Pamment argues that evaluation practices are influenced by the characteristics of an organization, place, and time. Resource constraints, government guidelines, and desired results affect choices made regarding methods. Pamment identifies four (nonexclusive) approaches to public diplomacy evaluation: outputs, outcomes, perceptions, and networks. The first two approaches (outputs and outcomes) are rooted in the effects-based tradition of evaluation; the last two (perceptions and networks) are examples of contextualized approaches to evaluation.[49] Table C.2 summarizes these models.

Output-based models of evaluation focus on the activities of press officers and the extent to which they have disseminated the message. These evaluations may include counts of the number of press clippings on a particular topic or head counts at events. Advertising value equivalents may supplement output evaluations. Rather than examining whether these efforts have an effect, output evaluations focus on the extent of campaign efforts, or the level of production. As such, output evaluations may be used in an organization context that emphasizes the need for proof of labor and production, or evidence of effort. Outcome models of evaluation build on logic models and link campaign objectives to the campaign's impact on the public. The focus of outcome evaluations is on collecting data on whether an organization's objectives were met by

Table C.2
Pamment's Evaluation Models

Articulation	Methods	Theory of Influence	Anticipated Results
Output models	Press clippings, advertising value equivalents	Public diplomacy as output	Proof of labor/reach/volume
Outcome models	Logic models, impact assessments	Public diplomacy leads to effects	Proof that organization is effective/efficient
Perception models	Surveys, polls	Reputation management	Proof of influence over ideas and values
Network models	Hubs, alliance formation	Relationship management	Proof of attention to relationships and perspectives

SOURCE: Adapted from Pamment, 2014.

---

[48] James Pamment, "Articulating Influence: Toward a Research Agenda for Interpreting the Evaluation of Soft Power, Public Diplomacy, and Nation Brands," *Public Relations Review*, Vol. 40, No. 1, March 2014.

[49] Author interview with James Pamment, May 24, 2013; James Pamment, "Towards a Contextualized Interpretation of Public Diplomacy Evaluation," paper presented at the International Studies Association annual convention, San Francisco, Calif., April 3–6, 2013.

an effort. Focusing on outcomes in public diplomacy efforts can produce concrete effects that are easily assessed through traditional forms of outcome measures.

Perception-based models of evaluation focus on understanding and influencing an audience of interest. As such, assessment involves collecting information on the values, attitudes, and opinions of this audience and considering how an audience interpreted or was persuaded by a particular message. Perception-based models of evaluation involve tailoring an effort to a particular audience, rather than assuming that a prepackaged approach will produce equivalent outcomes across groups. Finally, network models of evaluation focus on identifying key influencers and the extent to which these individuals redistribute key messages. These key influencers can also provide valuable information that can be used to adjust messages and policies to better address the positions of a target audience. This approach to evaluation places a heavy emphasis on relationship management, and it overlaps with JALLC's category of evaluation involving engagement with individuals or groups.

Both JALLC's framework and Pamment's framework propose that evaluation methods should be adjusted based on both an organization's theory of influence and the results that are of greatest interest to the organization. Furthermore, the elements of an effort that are of greatest interest will likely affect what measures an organization uses to evaluate its public diplomacy efforts.

Broadcasting Board of Governors and Voice of America: Designing and Implementing Research

Another example of public diplomacy evaluation comes from the Broadcasting Board of Governors (BBG). BBG is a federal agency that provides oversight for U.S. civilian (i.e., nonmilitary) international broadcasting. As such, it oversees many public diplomacy efforts. Voice of America (VOA) is the largest broadcaster in the BBG network. It uses radio, television, and the web to disseminate news and cultural programs to between 134 million and 164 million people around the world, including populations in underserved and developing countries.[50]

In an effort to better understand the perceptions and interests of international audiences, the BBG has designed and implemented international survey collection efforts.[51] To design the surveys, research directors at BBG collaborate to determine topics for survey items that address their research interests and needs. Contracted research companies then develop or provide guidance regarding survey items that address these topics of interest. BBG then employs research contractors in the countries of interest to administer the surveys in person. By hiring local firms, the BBG can ensure that individuals who are familiar with a particular language and culture assist with survey development and dissemination.

---

[50] Voice of America, "The Largest U.S. International Broadcaster," factsheet, Washington, D.C., March 2013.

[51] Author interview with Kim Andrew Elliot, February 25, 2013.

In implementing these surveys, BBG and its contracted research firms must overcome several challenges. For example, obtaining a nationally representative sample may be difficult, so, in some cases, surveys are collected in urban areas only. In addition, surveys must be kept at a reasonable length to avoid survey fatigue among both interviewees and interviewers.

Trust Pays: Examining Cultural Relations Efforts

The British Council seeks to build "cultural relations" by educating international audiences on the culture and assets of the United Kingdom and by improving trust between the United Kingdom and other countries.[52] The British Council has sponsored research to evaluate its efforts, including a recently published report called *Trust Pays*.[53]

As part of the *Trust Pays* effort, the British Council sought to understand the perceptions of "future influencers" in a range of different countries, so data collection efforts focused on individuals between 18 and 34 years of age. The research was conducted by YouGov, Ipsos MORI, and their partner organizations, and involved the use of online panels.

The researchers collected baseline data on the extent to which participants trusted people from different countries and were willing to do business with them. Data collection also focused on the relationship between trust in people from the United Kingdom and participants' involvement in cultural relations activities (e.g., those sponsored by the British Council). Results showed that participants tended to have greater trust in people from the United Kingdom than in people from other countries, such as Germany and the United States. Those involved in British Council cultural relations activities had more trust in people from the United Kingdom. Participants who trusted people from the United Kingdom more were more interested in doing business with the United Kingdom.

Other Examples from Public Diplomacy

Robert Banks, a researcher at the University of Southern California's Center on Public Diplomacy, developed a comprehensive overview of various public diplomacy efforts, resources, and processes.[54] The guide covers metrics and measures for cultural programming, information campaigns and media agenda setting, new media, challenges and opportunities in polling, and the audience for public diplomacy evaluation. Banks's work is a potentially useful resource for IIP planners seeking additional examples from the public diplomacy sector.

---

[52] British Council, *Annual Report: 2012–13*, London, March 31, 2013.

[53] British Council, 2012.

[54] Banks, 2011.

Assessment in Politics

Persuasion is considered a fundamental element of politics.[55] Political communication efforts often seek to motivate and persuade voters to support or oppose a particular candidate or policy. Recent efforts to influence and examine political attitudes provide insights into evaluation options.

Tro Tros in Ghana: Examining Exposure to Partisan Radio Stations

Around the world, partisan media stations present audiences with information that favors a political party or political viewpoint. Although such stations are blamed for polarizing audiences, limited research has examined the influence of exposure to partisan media on audience attitudes or behaviors. To address this gap in the literature, Jeffrey Conroy-Krutz of Michigan State University and Devra Moehler of the University of Pennsylvania examined the influence of partisan media exposure on polarization (e.g., more extreme support of political party after listening to a like-minded radio station) and moderation (e.g., greater tolerance for other opinions). Their efforts may be adapted to assess IIP efforts in other contexts.

To examine the effects of partisan media exposure, field experiments were conducted in Ghana using commuter minibuses, called *tro tros*. Passengers on *tro tros* usually listen to the radio station of the driver's choice. However, passengers on selected *tro tros* were randomly assigned to listen to one of four radio stations with a particular partisan leaning during their commute: pro-government, pro-opposition, neutral political conversation, or no radio. After completing their ride, passengers were interviewed and different behavioral measures were collected from different passengers. Measures included the following: (1) giving passengers money for interview participation and asking them to donate a portion of that money to a cause associated with one side or the other of the partisan split; (2) giving passengers a choice of key chains, each associated with a different party or the government; (3) and asking passengers to join a petition about transportation policy by texting a number, which would measure political efficacy and engagement. These behavioral measures assist in addressing research concerns regarding biases in self-reports of attitudes.

Although the data are still being analyzed, initial results suggest that listening to like-minded radio did not have polarizing effects. Listening to a radio station that challenged partisan preferences had a moderating effect, contributing to greater acceptance and support for another political party.

---

[55] Morgen S. Johansen and Mark R. Joslyn, "Political Persuasion During Times of Crisis: The Effects of Education and News Media on Citizens' Factual Information About Iraq," *Journalism and Mass Communication Quarterly*, Vol. 85, No. 3, September 2008.

Big Data and Campaign Analytics: Synthesizing Data Collection Efforts

One of the campaign elements that assisted President Barack Obama during the 2012 presidential campaign was the collection and use of voter data.[56] During the 2008 campaign, Obama's team collected massive amounts of data. However, a severe limitation of these data collection efforts was that there were multiple disconnected databases for the different efforts. To address this limitation, one system that merged information from multiple sources was created. This single-system approach permitted more-sophisticated analyses and, thus, better identification of which individuals may be influenced by certain campaign messages. In other words, it permitted the micro-targeting of individual voters, rather than targeting voters by broad geographical locations. In addition to identifying whom to target and how to target voters, the system permitted metric-driven fundraising.

However, large amounts of information regarding voters do not assist with determining which messages are most effective for, for example, raising funds. To assist in examining the effects of messages and certain efforts, the Obama campaign used randomized control trials and other experimental designs. For example, the campaign planners altered the amount of money requested in fundraising emails and then tracked the amount of money raised from these different email efforts. They also randomly assigned voters to "treatment" and control groups, with those assigned to the treatment group receiving phone calls from campaign staffers. Later, they polled a sample of voters to determine the impact of the phone calls.[57] These efforts demonstrate how the use of big data and different research methods can inform persuasion efforts, thereby providing guidance regarding how to modify a particular campaign to best meet a given set of objectives.

---

[56] Michael Scherer, "Inside the Secret World of the Data Crunchers Who Helped Obama Win," *Time*, November 7, 2012.

[57] John Sides and Lynn Vavreck, "Obama's Not-So-Big Data," *Pacific Standard*, January 21, 2014.

Major Theories of Influence or Persuasion

This appendix briefly reviews several major theories of persuasion or influence that have been developed and used across disciplines and can inform the design, implementation, and assessment of an IIP effort. We begin with a description of communication factors associated with persuasion and factors that may be influenced by persuasion. We then describe the theory behind how participants understand persuasive messages, suggesting avenues to promote longer-lasting attitude changes.

Attitudes, attitude changes, and social factors play a role in persuasion. In this appendix, IIP planners may (1) recognize their implicit theory of change among those listed, (2) select from the described theories to inform their own efforts, or (3) build from these theories to create a new theory of change. Programs that are based on theory and research are more defensible than those based on intuition and assumption.

After exploring the theories, we present three approaches to organizing the elements of IIP efforts, concluding with examples from different disciplines, including business and marketing, public communication/social marketing, public diplomacy, and politics. IIP planners are encouraged to draw from the methods and theories presented here for their own purposes, either in developing IIP efforts or in planning assessments of those efforts.

Understanding and Using Existing Theories of Change

Understanding and incorporating the core elements of existing theories of behavior change and previous research into an effort's design can assist in creating and implementing an effort that will have the greatest chance of succeeding (i.e., the best chance of having the desired effects). When developing an IIP effort or program, planners may be inclined to build solely from their own intuition, untested patterns of effects that are perceived to exist in previous efforts, and readily available anecdotal evidence. However, research suggests that there are benefits to drawing on thoroughly tested theories of change during the design of an effort or when planning an interrelated set of efforts. Such efforts are more likely to be effective than those that lack a strong theoretical basis

and are based on commonly held and untested assumptions.[1] Thus, it is worthwhile, if not essential, to base a particular IIP effort on existing persuasion-relevant theories of change and research.[2]

In explaining the importance of the use of existing theories of change, Thomas Valente noted:

> To achieve and measure impact, researchers and programmers must understand the population and have articulated a theory of behavior change. Theory helps define behavior and specify the mechanism thought to influence it, which informs *program design*, *variable measurement*, *goal and objective setting*, and the ability to distinguish between *theory and program failure*. For example, if a theory validated by research posits that adolescents are most influenced by their peers, the program should be implemented by those within the audience's peer group.[3]

Among other things, empirically tested theories of change can inform assumptions about which effects are likely to result from certain actions or efforts, who is most likely to be affected by those actions, and when those effects are most likely to be seen. As such, awareness and utilization of previously assessed theories of change contribute to more-competent communication, enhancing the ability to achieve objectives in ways that are most appropriate for a particular context.[4]

This report focused primarily on one theory of change in developing an IIP effort. However, to communicate competently and build an effective program, IIP planners should not assume that one theory can or should be used for all messages, across all audiences, and at all times. Some theories may be more appropriate in certain contexts than others. Further, one should not assume that only one theory can be used to develop the logic model for a particular IIP effort. It is worthwhile to assume that multiple theories can address highly similar concepts, and different empirical assessments may address separate components contained within a single theory. Multiple theories should be considered when developing an IIP program.[5]

Program logic models can offer a structured approach for developing, managing, evaluating, and improving IIP efforts. Theories address broad principles of behavior change, whereas logic models embody a theory or set of theories in a particular con-

---

[1]   Karen Glanz and Donald B. Bishop, "The Role of Behavioral Science Theory in Development and Implementation of Public Health Interventions," *Annual Review of Public Health*, Vol. 31, 2010.

[2]   Magne Haug, "The Use of Formative Research and Persuasion Theory in Public Communication Campaigns: An Anti-Smoking Campaign Study," paper presented at the Nordic Mass Communication Research Conference, Reykjavik, Iceland, August 10–14, 2001.

[3]   Author interview with Thomas Valente, June 18, 2013; emphasis added.

[4]   Robert H. Gass and John S. Seiter, *Persuasion: Social Influence and Compliance Gaining*, 5th ed., New York: Pearson, 2014.

[5]   Author interview with Thomas Valente, June 18, 2013.

text.[6] To ensure that the logic model is useful and not too complex for program development and implementation, it is important to identify and summarize the central tenets of relevant theories of change. This step will assist in creating a clear and concise IIP effort.[7]

In this appendix, we present several IIP-relevant theories that have been developed across the social sciences, with a strong emphasis on theories that draw from social psychology. These theories of change have been empirically tested and are well known in the fields of communication and, more specifically, persuasive communication. Further, the different theories of change presented here highlight the distinct ways of conceptualizing persuasive messages and their differential effects on specific audiences in specific contexts. We also present several examples of how previous efforts in social marketing, public diplomacy, and other areas have incorporated theories of change into their program-specific efforts. These examples offer guidance to IIP program planners for applying components from broad theories to a specific context. This information can serve as a starting point for designing an IIP effort, helping planners identify a useful theory of change to inform their effort or build from the theories described here to explicate their own theory of change. Table D.1 summarizes the theories discussed in this appendix.

IIP Theory and Research Across the Social Sciences

Different theories can assist in the development of IIP efforts. In this section, we describe several theories of change that may be informative for IIP program planners. Of note, there are a multitude of potentially useful theories that can be used to connect an IIP effort's planned activities or messages with its intended effects and to articulate a logic model for the assessment of those effects. We present only a subset of the available theories.

Inputs and Outputs Involved in IIP Efforts

Multiple factors can influence how an IIP effort or campaign influences a particular audience. There are also many variables, or constructs, within an audience that can be influenced by a campaign. Theory and research in IIP-relevant areas differentiate among audience knowledge, attitudes, and behaviors. *Knowledge* includes the information that an audience has about a particular topic, object, person, or entity. This information may or may not be factual and accurate.[8] Thus, an audience can have knowl-

---

[6] Author interview with Thomas Valente, June 18, 2013.

[7] Heath and Heath, 2007.

[8] Richard E. Petty and John T. Cacioppo, *Attitudes and Persuasion: Classic and Contemporary Approaches*, Boulder, Colo.: Westview Press, 1996.

Table D.1
Summary of Major Theories of Influence or Persuasion

Approach	Theory	Brief Description
Social Science Theory and Research		
Inputs and outputs in persuasion	The input-output communication matrix	Influential communication factors, including source, message, channel, receiver, and destination characteristics, can affect an audience's responses. There are several responses to communication factors, which may roughly occur in a sequence.
Information processing	Elaboration likelihood model	There are two routes to persuasion: a central route and a peripheral route. The central route involves careful and effortful processing of information, whereas the peripheral route involves the use of superficial cues. The central route is associated with longer-lasting attitude change.
Functions of change	Social learning	People learn social behavior by observing role models or similar others and imitating their behavior. People are especially likely to imitate behavior if they see the role model rewarded for his or her actions.
	Opinion change	There are three categories of opinion (or attitude) change: compliance, identification, and internalization. Compliance involves performing an action without approval. Identification involves accepting a message without extensive thought about the message. Internalization involves fully accepting and supporting a message, uninfluenced by coercion or a need for affiliation.
Processes of change	Theory of planned behavior	Specific attitudes toward a behavior, beliefs about the relevance of a behavior, and the perceived ease of performing a behavior influence one's inclination to perform a behavior. These inclinations then influence the behavior.
	Cognitive dissonance	A person's behaviors can change their attitudes. When an individual behaves in a way that goes against his or her attitudes, he or she may feel discomfort or dissonance. To reduce this discomfort, people will change their previously held attitudes to align with their behaviors.
	Knowledge, attitudes, and practices	Knowledge, attitudes, and practices or behaviors can vary in terms of processing sequence. These three constructs can be ordered six different ways, each with implications for an IIP effort. For example, knowledge could lead to attitudes, which lead to practices, or attitudes could lead to knowledge, which leads to practices.

Approach	Theory	Brief Description
Social factors	Principles of social influence	Six principles of influence address social factors that can contribute to an individual meeting a request. These are liking, reciprocity, social proof, commitment, authority, and scarcity.
	Diffusion of innovation theory	Innovations are the new programs, practices, policies, and ideas that are communicated to an audience. Diffusion is the process through which these new innovations are spread or communicated in a social system. This communication is influenced by multiple factors and can contribute to change in the social system.
	Systems approach	An individual has interpersonal relationships, and these relationships occur within a community that operates within a society. Each of these social elements represents a different level of social influence.
Fear	Various theories	The relationship between fear-based appeals and attitude or behavioral change is complex. There is a strong potential for fear-based appeals to have the opposite effects of those intended. An audience may respond defensively to a fear-based message.
Culture	Various theories	It is important to consider the influence of culture on the efficacy of an IIP effort. Practitioners should be aware that the results of previous studies may not apply to different groups.

Organizing IIP Theory

Approach	Theory	Brief Description
Automatic processes	MINDSPACE	MINDSPACE is a mnemonic developed to address contextual influences of behaviors. It focuses on nine of the contextual factors that have robust effects: messenger, incentives, norms, defaults, salience, priming, affect, commitment, and ego.
Influencers and policies	Behavior change wheel	The behavior change wheel is a method for identifying the characteristics of certain efforts or interventions and connecting them to the behaviors that the interventions seek to change. It provides a pictorial representation of the six components that influence behavior, nine intervention functions, and seven types of policies.
Information environment	The Initiatives Group Information Environment Assessment Model	This framework provides guidance regarding what elements to change and identifying observable changes that will occur if certain effects have occurred.

edge that is based on incorrect and untrue information. *Attitudes* involve the positive or negative judgments that an audience has regarding a topic, object, person, or entity. In other words, attitudes involve forming opinions, not simply having basic knowledge.[9] Finally, *behaviors* involve the actions of an audience or individual. Communication is multidimensional and dynamic, and IIP program planners must keep these variables in mind when determining how to communicate a message.[10]

One of the most instrumental efforts in persuasion began with Carl Hovland at Yale University in the 1950s. During World War II, Hovland had worked with the Army's Information and Education Division on mass communication research.[11] At Yale, Hovland and his colleagues systematically examined the variables perceived to influence persuasive efficacy. Their focus was on understanding and influencing attitudes and attitude change. They considered source credibility, individual differences, and message order effects.

Building from this attitude research, William McGuire further developed several of the theoretical concepts reflected in his input-output communication matrix (see Figure D.1). The matrix identifies variables that can be manipulated or changed (i.e., independent variables) by program planners. In the input-output communication matrix, these are called *input communication variables*. The matrix also identifies variables that are likely to be influenced by these input variables: *output persuasion steps*. McGuire conceptualized these outputs as steps along the way to enduring behavior change.[12] The following sections discuss each concept in turn.

Input Communication Variables

Input communication variables can affect how an individual or audience perceives a particular object or entity. They include characteristics of the source, message, channel, receiver, and destination (see Figure D.1). These factors address the following classic question: "Who says what to whom, when, and how?"[13]

Source characteristics are the features of the individual, group, or organization communicating a message. A source may be an individual or group, appealing or unappealing, credible or not. For example, logic would suggest, and previous research has

---

[9]   Susan T. Fiske, *Social Beings: A Core Motives Approach to Social Psychology*, Hoboken, N.J.: John Wiley and Sons, 2004.

[10]   Nova Corcoran, "Theories and Models in Communicating Health Messages," in Nova Corcoran, ed., *Communicating Health: Strategies for Health Promotion*, Thousand Oaks, Calif.: Sage Publications, 2007.

[11]   Richard E. Petty and Duane T. Wegener, "Attitude Change: Multiple Roles for Persuasion Variables," in Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds., *The Handbook of Social Psychology*, 4th ed., New York: McGraw-Hill, 1998.

[12]   McGuire, 2012.

[13]   Richard E. Petty, Pablo Briñol, and Joseph R. Priester, "Mass Media Attitude Change: Implications of the Elaboration Likelihood Model of Persuasion," in Jennings Bryant and Mary Beth Oliver, eds., *Media Effects: Advances in Theory and Research*, 3rd ed., New York: Lawrence Erlbaum Associates, 2009.

Figure D.1
McGuire's Input-Output Communication Matrix

Input Communication Factors	Output Persuasion Steps
• Source characteristics (e.g., credibility, attractiveness) • Message characteristics (e.g., repetition, appeal) • Channel (e.g., modality) • Receiver (e.g., personality, ability) • Destination (e.g., immediate/delay, resistance)	• Exposure (tuning in) • Attention (attending to the message) • Liking (maintaining interest) • Comprehending (learning what) • Acquiring skills (learning how) • Agreeing with message (attitude change) • Remembering message (storing in memory) • Recall (retrieving new position from memory) • Intention (decision to act based on recall) • Action (performing the behavior) • Integrating (post-action cognitive integration) • Proselytizing (encouraging others)

SOURCE: Adapted from McGuire, 2012.

RAND *RR809/1-D.1*

demonstrated, that people are more likely to be persuaded by a highly credible source, someone perceived as trustworthy or having expertise on the topic being discussed.[14] Likewise, attractive or appealing sources (sources that are pleasant, familiar, and similar to the audience) can be more persuasive than those that are less familiar or similar to the audience.[15] This suggests that the source presenting a particular message should be carefully considered, and target-audience perceptions of this source should be determined before use in an IIP effort.

Message characteristics are the features of the persuasive communication that an audience receives. In other words, the message is the information that is being provided by the source. A message can be emotional or logical, specific or general, repetitious or not. When an audience is not highly motivated to attend to or recall a message, moderate repetition within the message or repeated exposure to the message may facilitate recall and persuasion.[16] Another message characteristic can be the inclusion of detailed factual information or broad evaluative information. For many audiences, messages based on broad evaluative information are more persuasive than those based

---

[14] Chanthika Pornpitakpan, "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence," *Journal of Applied Social Psychology*, Vol. 34, No. 2, February 2004.

[15] McGuire, 2012.

[16] Cornelia Pechmann and David W. Stewart, "Advertising Repetition: A Critical Review of Wearin and Wearout," *Current Issues and Research in Advertising*, Vol. 11, 1988.

on a detailed litany of facts.[17] The characteristics of the message must be developed in a manner that is most appropriate for the audience.

The message channel is how the message is transmitted to the audience. This can include television, radio, Internet, billboards, flyers, or letters. At a minimum, to ensure that the target audience is exposed to a particular message, the channel that is used should correspond to a format likely to be used or seen by the audience.[18] Clearly, receiver or audience characteristics can play a role in persuasion, and it follows that certain messages may be effective for certain audiences more than others.

Another category of input factors is destination-based. This may include whether the persuasion is intended to be immediate or delayed and whether the intent of the communication is to promote the acceptance of a message or resistance to another message. For example, research suggests that if participants are forewarned of the persuasive content of a message, they will be more resistant to being persuaded by that message.[19] Overall, these input factors suggest that both the goal of a communication campaign *and* its audience should be well understood, and the message should be tailored appropriately.

Output Persuasion Steps

In addition to different input factors, McGuire's input-output communication matrix outlines a general theoretical sequence of output steps corresponding to a hierarchy of output effects (behaviors).[20] These outputs can be used to evaluate the extent to which a particular message or campaign has been effective. In other words, they are different constructs that can be assessed when evaluating a campaign (i.e., dependent variables). According to McGuire, for a message to be persuasive, it must first reach the target audience, and, for example, it must be sufficiently nonthreatening that it minimizes the likelihood that the audience will tune out upon exposure. After exposure, an audience must give attention to the message. Even if a person sees an advertisement on television, he or she may not know what the advertisement is communicating. Then, that viewer must like the message, understand what the message is communicating, and know how to behave in accordance with the message. Once each of these steps is achieved, an audience must agree with the message, or evaluate it favorably, which suggests that an initial attitude has changed. Afterward, this message must be stored in the audience's memory and later recalled. This message recall must then lead to the intention to behave in a way that is supportive of the message, which contributes to actually

---

[17] Meera P. Venkatraman, Deborah Marlino, Frank R. Kardes, and Kimberly B. Sklar, "The Interactive Effects of Message Appeal and Individual Differences on Information Processing and Persuasion," *Psychology and Marketing*, Vol. 7, No. 2, Summer 1990.

[18] McGuire, 2012.

[19] Richard E. Petty and John T. Cacioppo, "Forewarning, Cognitive Responding, and Resistance to Persuasion," *Journal of Personality and Social Psychology*, Vol. 35, No. 9, September 1977.

[20] McGuire, 1989.

behaving in a way that is based on the intention. After behaving in accordance with the message, an individual must integrate this behavior into his or her thought and behavior patterns. Finally, an audience must begin to encourage similar behavior from others. Each of the outputs—exposure, attention, liking, comprehension, skill, agreement, memory, recall, intention, action, integration, and proselytization—is an effect of a communication campaign that can be assessed in addressing persuasive efficacy.

Note that there are limitations to this exact sequence of output persuasion steps. All of these steps may not be needed for behavioral change to occur, and fulfilling certain steps does not guarantee that behavioral change will occur. For example, the two-decade $23 million a year Got Milk? advertising campaign, which featured celebrities endorsing milk, was popular enough to inspire parodies and references in television shows and movies. Despite the campaign's longevity and reported 90-percent awareness among the U.S. public,[21] milk sales declined nationally over the period of the campaign, losing considerable ground to soft drinks, energy drinks, and nondairy milk alternatives.[22] The campaign successfully met each of the output persuasion steps in McGuire's matrix, yet it failed to influence behavior by getting people to drink (or, more importantly, purchase) more milk.

Information Processing Approach: Two Paths to Persuasion

Information processing encompasses the effortful thought involved in understanding a particular message. Some persuasive messages are associated with slow and analytic thought on the part of the audience, thus encouraging more-effortful information processing. By contrast, some messages can be understood through use of rapid and superficial processing, requiring minimal cognitive effort. Building from these two kinds of audience approaches to information processing, less effortful processing or more-effortful processing, Richard Petty and John Cacioppo developed the elaboration likelihood model.[23] They hypothesized that there are two routes to persuasion: a central route and a peripheral route (see Figure D.2). The central route involves careful and effortful processing of information, whereas the peripheral route involves the use of superficial cues, such as audience mood and the likeability of the message source.[24]

When members of an audience are willing (or motivated) and able to engage in effortful processing of a persuasive message, they are more likely to use the cen-

---

[21] "Got Milk? Is Here to Stay," PRNewswire, March 3, 2014.

[22] Gene Del Veccio, "Got Milk? Got Fired: 5 Valuable Lessons That All Executives Must Heed," *Huffington Post*, March 12, 2014. See also Renee J. Bator and Robert B. Cialdini, "The Application of Persuasion Theory to the Development of Effective Proenvironmental Public Service Announcements," *Journal of Social Issues*, Vol. 56, No. 3, Fall 2000.

[23] Petty and Wegener, 1998.

[24] Another very similar approach to processing, called the heuristic-systematic model, also proposes that there are two routes to processing, one involving simple decision rules and another involving more-systematic processing.

Figure D.2
Elaboration Likelihood Model: Two Routes to Persuasion
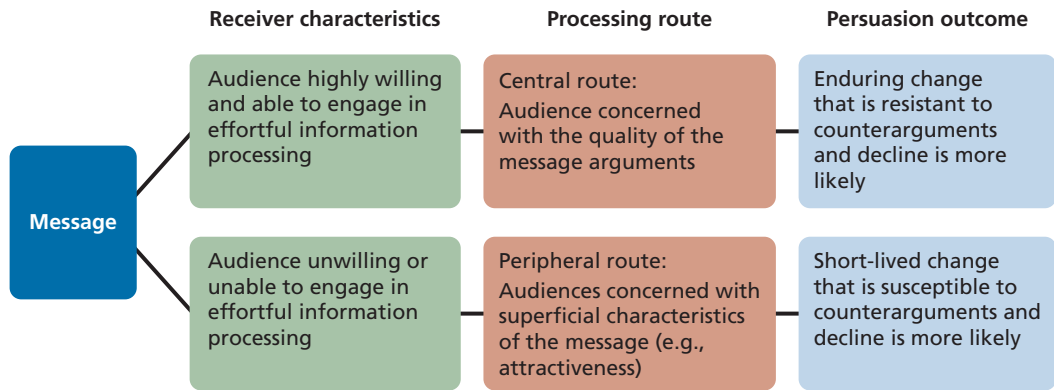
tral route to information processing and to carefully consider the message arguments (i.e., to elaborate on the message). However, when they are unwilling or unable to engage in effortful processing, they are more likely to use the peripheral route to processing and consider surface characteristics of the message, such as attractiveness. What may motivate an audience to give more attention to a message and thus engage in the central route to processing? When audiences find a message personally relevant and enjoy thinking things through, they are more likely to engage in "effortful" processing.[25] Fatigue, distractions, and an inability to understand a complex message are associated with an audience's inability to engage in the central route to processing. If a message is strong and has high-quality arguments, IIP planners will likely want to demonstrate the relevance of the message to the audience and obtain the audience's full and undistracted attention. If the message is somewhat weak, it may be worthwhile to embellish peripheral features, such as the attractiveness of the message.

The route by which an audience processes a message is important for ensuring a lasting impact on audience attitudes. The central, effortful route to information processing is associated with long-lasting attitude change, whereas the peripheral route is associated with temporary attitude change. In one study, people were asked to either analyze the logic of a message's arguments or assess peripheral aspects of the message. Both groups showed attitude change. However, ten days later, those who had analyzed the logic of the message's arguments demonstrated more lasting attitude change than those who assessed the message's peripheral cues.[26]

---

[25] Petty and Wegener, 1998.

[26] Shelly Chaiken, "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," *Journal of Personality and Social Psychology*, Vol. 39, No. 5, November 1980.

Richard Petty, one of the developers of the elaboration likelihood model, and his colleagues have argued that the media can play a role in promoting the idea that certain topics are highly relevant for an audience.[27] If the audience perceives a media-promoted topic as highly relevant, then it is more likely to engage in effortful, central processing that can lead to longer-lasting attitude change. For example, extensive media coverage of an issue, such as a political scandal, violence, or drug abuse, over an extended period may contribute to increases in the perceived personal relevance of the issue. Thus, the media can set the agenda of what is important to think about or evaluate and indirectly influence audience attitude change.

Functional Approaches: Considering Needs and Wants

Although some theory and research has focused on the paths to differential cognitive processing involved in persuasion, other research has focused on the reasons for people's attitudes and behaviors. According to these functional approaches, to be persuasive, messages should address these reasons.[28]

Social Learning

People may develop and hold attitudes to interpret the world and better interact with others. Consequently, audiences may determine how to behave and what to think based on what they see or hear from others who are similar or who are in positions of status and power. According to Albert Bandura's social learning theory, people learn social behavior by observing role models or similar others and imitating their behavior. People are especially likely to imitate behavior if they see the role model rewarded for his or her actions, suggesting that people perform actions that they believe are most likely to achieve desired results. Bandura's classic research showed that children who observed an adult showing aggressive behavior toward a doll were more likely to exhibit aggressive behavior than children who observed an adult who did not behave aggressively toward the doll. Further, children were especially likely to spontaneously imitate behavior when seeing adults get rewarded for their actions.[29] Social learning theory is often used when discussing the influence of violence in popular media on aggressive behavior among children.[30] It has also been used to discuss factors that influence par-

---

[27] Petty, Briñol, and Priester, 2009.

[28] Richard J. Lutz, "A Functional Approach to Consumer Attitude Research," in Kent Hunt, ed., *Advances in Consumer Research, North America Conference*, Vol. 5, Ann Arbor, Mich.: Association for Consumer Research, 1978.

[29] Albert Bandura, Dorothea Ross, and Sheila A. Ross, "Transmission of Aggression Through Imitation of Aggressive Models," *Journal of Abnormal and Social Psychology*, Vol. 63, No. 3, November 1961.

[30] Elliot Aronson, Timothy D. Wilson, and Robin M. Akert, *Social Psychology*, 5th ed., Upper Saddle River, N.J.: Prentice Hall, 2005.

ticipation in terrorism.[31] IIP planners may apply this theory when attempting to influence audience behavior.

Process of Opinion Change

Another theory that utilizes a functional approach to attitude and behavior change is Herbert Kelman's process of opinion change.[32] According to Kelman, there are three categories of opinion (or attitude) change: compliance, identification, and internalization. Compliance involves performing an action or suggested agreement with a request *without* actual approval of or agreement with the action or request. In other words, compliance involves audience actions performed without internalizing a message. People comply when they feel that they have no other choice or when they are not motivated to challenge the person requesting the action. For an audience, compliance functions to reduce the chances of negative repercussion and increase the chance of reward. Once the chances of repercussion or reward are eliminated, compliance would be expected to diminish.

Identification involves identifying with the message source, message role model or idea originator, and, subsequently, accepting and believing the message being communicated without giving extensive thought to the message. An audience will adopt the message or behavior of another in an effort to develop a relationship or affiliation with this other person or group. Thus, identification is a function of the desire to build or support a relationship or affiliation and involves supporting a message or performing a behavior for the purpose of relationship development. When the relationship or affiliation is no longer desired or no longer critical to an audience's self-perception, Kelman's theory suggests that agreement with the ideas or performance of the actions will cease.

Finally, internalization involves full acceptance and support of a message, uninfluenced by coercion or a need for affiliation. Internalization occurs when a message is congruent with a person's beliefs and values. The person adopts the message's arguments and engages in certain behaviors because they are intrinsically rewarding. Internalization is a function of the desire to agree with and support arguments and actions that are perceived to be rational and congruent with one's personal belief system.

Kelman's categories of opinion change provide a way to conceptualize categories of behavior and attitude change according to the functions they address. Compliance may best be achieved by using threats or rewards. This aligns with several of the concepts in Bandura's social learning theory. Identification may be best achieved by enlisting an appealing source or group to communicate a message, aligning with a peripheral route to information processing in the elaboration likelihood model. Internalization

---

[31] Ronald L. Akers and A. L. Silverman, "Toward a Social Learning Model of Violence and Terrorism," in Margaret A. Zahn, Henry H. Brownstein, and Shelly L. Jackson, eds., *Violence: From Theory to Research*, Cincinnati, Ohio: Matthew Bender and Co., 2004.

[32] Herbert C. Kelman, "Processes of Opinion Change," *Public Opinion Quarterly*, Vol. 25, No. 1, 1961.

may be best achieved by using a credible source that presents a clear and thorough message, aligning with the central route to information processing.

Theories of Influence Processes: When Attitudes Matter

As noted in the context of McGuire's input-output communication matrix, attitudes may be considered a step on the way to influencing behaviors. However, attitudes do not always predict behaviors, and previous research has consistently demonstrated that, even after expressing a particular attitude, people can behave in apparently contradictory ways.[33] Aware of this apparent contradiction, theorists and researchers have developed different theoretical models to determine when attitudes do predict behaviors and when other patterns of influence are more likely. One of the most well-known and often-used models addressing this is known as the theory of planned behavior, which builds from a previous theory of reasoned action.[34]

Theory of Planned Behavior

The theory of planned behavior begins with the basic notion that actions or behaviors are most strongly predicted by intentions to engage in the specified actions, termed *behavioral intentions* (see Figure D.3). Thus, an audience intends to engage in the behavior before actually doing so. Behavioral intentions can be influenced by a number of different factors. Attitudes are only one of the several factors that affect behavioral intentions.

Figure D.3
Basic Theory of Planned Behavior
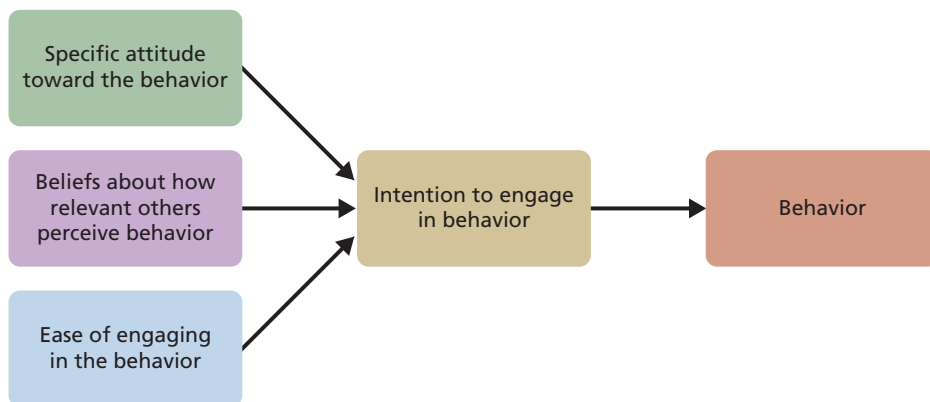
RAND *RR809/1-D.3*

---

[33] Richard T. LaPiere, "Attitudes vs. Actions," *Social Forces*, Vol. 13, No. 2, December 1934.

[34] Icek Ajzen, "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, December 1991.

Further, only certain attitudes predict behavioral intentions. The theory of planned behavior proposes that the specificity of the attitudes addressed and measured has consequences for the ability of the attitudes to predict behavioral intentions. In other words, the attitudes must be specific to the behavior of interest. General attitudes toward a topic are less likely to predict specific behavioral inclinations. For example, one study assessed women's attitudes toward birth control (general attitudes) and attitudes toward using birth control pills for the next two years (specific attitudes about a behavior).[35] Two years later, these same women were asked about their use of birth control. Results showed that general attitudes toward birth control did not predict birth control use, but specific attitudes about using birth control for the next two years did predict the women's behavior in terms of birth control use.

In addition to specific attitudes toward a behavior, another factor proposed to influence behavioral intentions is subjective norms. Subjective norms are an individual's perceptions about what other people will think about the behavior. For example, perceived support or lack of support for a behavior among family and friends will influence one's inclination to perform a certain action.

A third factor that influences behavioral intention is perceived behavioral control. In other words, a person's intention to perform a behavior is influenced by the perceived ease of the action. If a behavior is perceived as difficult, the theory of planned behavior proposes that people will be less inclined to engage in it.

This theory can help guide IIP planners in designing and assessing the efficacy of a campaign or persuasive message. Rather than assessing general attitudes toward a topic, planners should measure specific attitudes toward specific behaviors, as well as perceptions of how others perceive the behavior and perceptions regarding the ease of performing the behavior. Such a campaign will have the strongest impact on behavioral inclinations and behaviors.

Cognitive Dissonance

One of the best-known counterarguments to the basic process supported by the theory of planned behavior—namely, the notion that attitudes precede behaviors—comes from the theory of cognitive dissonance. This theory postulates that, under certain circumstances, a person's behaviors can change his or her attitudes. Although seemingly counterintuitive, there is a great deal of research to support this potential sequence of causality.[36]

Cognitive dissonance involves the feeling of tension, discomfort, or unease that a person feels when there are multiple incongruities in his or her thoughts and actions. Cognitive dissonance theory proposes that when individuals behave in a way that

---

[35] Andrew R. Davidson and James Jaccard, "Variables That Moderate the Attitude-Behavior Relation: Results of a Longitudinal Study," *Journal of Personality and Social Psychology*, Vol. 37, No. 8, August 1979.

[36] Petty and Wegener, 1998.

goes against their attitudes, they feel dissonance. This dissonance is uncomfortable and unpleasant, so people will attempt to find ways to reduce this negative feeling. To reduce the incongruence between thoughts and behaviors, people may avoid the thoughts that are inconsistent with their behaviors, focus more on thoughts that are consistent with their behaviors, or reduce their focus on thoughts that are inconsistent with their behaviors.[37] For example, after engaging in a behavior that goes against his or her attitudes, a person may subsequently change those attitudes to match the behavior (i.e., avoid thoughts that are inconsistent with the behaviors), thereby reducing dissonance. In other words, people try to find ways to justify their actions.

In one of the classic studies on cognitive dissonance, students were asked to spend an hour performing extremely boring, repetitive, and mundane tasks. After doing so, and after the experiment seemed to be over, the researcher offered students either $1 or $20 to tell another student that the study was actually fun and interesting. Students who were offered $1 to lie felt a great deal of dissonance. They had told another student something that did not align with their own attitudes, and they had not been paid enough to warrant lying. By contrast, those offered $20 felt very little dissonance; they lied because they had been paid enough to warrant doing so. To deal with their dissonance, students paid $1 needed to find a way to reconcile why they had lied to another student and their discomfort with doing so. Those paid $20 had no need to do so. As a result, those who were paid $1 changed their attitudes to be more positive toward the task, congruent with the lie they had told to the other students. Later assessments showed that students paid $1 recalled the tasks more favorably than the students paid $20. Thus, the behaviors of those paid $1, who lied to another student with insufficient reason, influenced their attitudes.[38] This suggests that when behavior change is desired, one option may be to promote attitude-changing dissonance.

More recently, the principles of cognitive dissonance have been applied as part of efforts to gain greater understanding of terrorism. For example, killing or harming people in one's own national or religious groups may cause dissonance, such that the behavior (e.g., bombing, shooting, kidnapping) goes against positive or even neutral attitudes toward certain groups. Therefore, terrorists may have a need to address this dissonance. This may include avoiding thoughts that are inconsistent with behaviors by blocking out opposite perspectives regarding the violent actions and engaging more in ways of thinking that are consistent with behaviors by framing the actions as good acts that are part of a fight against evil.[39] By understanding actions in terms of theory, the principles of the theory may be incorporated into programs that counter these actions.

---

[37] Andrea Kohn Maikovich, "A New Understanding of Terrorism Using Cognitive Dissonance Principles," *Journal for the Theory of Social Behaviour*, Vol. 35, No. 4, December 2005.

[38] Leon Festinger and James M. Carlsmith, "Cognitive Consequences of Forced Compliance," *Journal of Abnormal and Social Psychology*, Vol. 47, No. 2, March 1959.

[39] Maikovich, 2005.

Knowledge, Attitudes, and Practices

More recently developed models suggest that knowledge, attitudes, and behaviors can vary in terms of processing sequence, such that behaviors do not always result from attitudes. Individuals may change their behaviors before changing their attitudes. Thomas Valente and colleagues developed and tested different hypotheses regarding the relative order of knowledge (K), attitudes (A), and practices (P) or behaviors.[40] They proposed that there are six different orderings of these three constructs. For example, the traditional "knowledge leads to attitudes leads to practices" conception (K-A-P) is associated with learning and cognitive advancement from stage to stage. However, initial attitudes toward an idea or object may prompt a person to learn more, and this may then change behaviors, suggesting an attitudes-knowledge-practices sequence (A-K-P). This theory and research demonstrate that many possible causal sequences are conceivable and should be considered when designing a campaign.[41]

Social Norms and Social Context: Influential Social Factors

Building from theory about the influence of social norms in influencing intentions and behaviors, researchers have considered when norms will influence actions and which norms will influence these actions.[42] This theory and research focuses on the social factors that bolster or reduce the impact of persuasive messages.

Principles of Social Influence

Robert Cialdini has developed and researched many of the recent theoretical principles of social influence, focusing on the role that six principles play in social influence: reciprocity, social proof, commitment and consistency, liking, authority, and scarcity.[43] These six principles of influence address social factors that can contribute to an individual meeting a request.

The principle of reciprocity proposes that people repay others for what they have been given. There is a strong and pervasive social norm across human cultures to repay others for gifts or services received from them. This norm of reciprocity has been credited as the reason underlying the success of a technique called "door in the face."[44] This technique involves first making a large, extreme request that will likely be rejected and, after rejection, following with a smaller request. The technique is believed to be effec-

---

[40] Thomas W. Valente, Patricia Paredes, and Patricia R. Pope, "Matching the Message to the Process: The Relative Ordering of Knowledge, Attitudes, and Practices in Behavior Change Research," *Human Communication Research*, Vol. 24, No. 3, March 1998.

[41] Also see B. J. Fogg, "A Behavior Model for Persuasive Design," in *Proceedings of the 4th International Conference on Persuasive Technology*, New York: ACM, 2009.

[42] Robert B. Cialdini, Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter, "Managing Social Norms for Persuasive Impact," *Social Influence*, Vol. 1, No. 1, 2006.

[43] Robert B. Cialdini, *Influence: Science and Practice*, 4th ed., Boston: Allyn and Bacon, 2001a.

[44] Cialdini and Goldstein, 2004.

tive because the recipient of the request has been primed to reciprocate the requester's concession of moving from a large to a small request by making his or her own concession by moving from a rejection to an acceptance of the request. This suggests that giving to others can improve the chances that they will give in kind back to you.

A second principle of social influence is social proof. This principle proposes that people conform to the behaviors of similar others, suggesting that persuasion can be more effective when a message is presented by peers. For example, research has shown that people are more likely to donate to a charity when they learn that others like them have donated.[45] Thus, proof that someone's family, friends, or neighbors have performed an action or endorsed a cause can be a tool of influence.

Commitment or consistency is another principle of influence that can be used in the social arena. As noted, according to the theory of cognitive dissonance, people prefer to behave and think in ways that align. Once people have made clear commitments, they tend to align with (behave and think consistently with) these commitments. The success of the foot-in-the-door technique is credited to this principle of consistency.[46] This technique, the reverse of the "door in the face" technique described earlier, involves making a small request followed later by a larger request. Research has shown that, in an effort to maintain consistency, people will comply with a larger request after first complying with the smaller. For example, in one study, researchers asked individuals to place a small sign in their yards, and they later asked to enter the individuals' homes and catalog all of their household goods. Those who had agreed to place a small sign in their yard were more likely to later agree to have their household cataloged.[47]

The principle of commitment and consistency is also credited with contributing to the success of the low-ball technique, which is often used in car sales. In this case, an individual makes an active decision to purchase a product based on a certain characteristic, such as an extremely good price. Once the individual has actively decided to make the purchase, it is easier for the seller to negate the advantages that led the individual to make the decision in the first place, such as by pushing additional features or extra fees.[48] Influence using consistency is more likely to be effective when people make public, written, and voluntary commitments. Thus, making commitments public can reinforce the need to maintain a commitment and comply with subsequent requests.

---

[45] Peter H. Reingen, "Test of a List Procedure for Inducing Compliance with a Request to Donate Money," *Journal of Applied Psychology*, Vol. 67, No. 1, February 1982.

[46] Cialdini and Goldstein, 2004.

[47] Jonathan L. Freedman and Scott C. Fraser, "Compliance Without Pressure: The Foot-in-the-Door Technique," *Journal of Personality and Social Psychology*, Vol. 2, No. 2, August 1966.

[48] Robert B. Cialdini, John T. Cacioppo, Rodney Bassett, and John A. Miller, "Low-Ball Procedure for Producing Compliance: Commitment Then Cost," *Journal of Personality and Social Psychology*, Vol. 36, No. 5, May 1978.

According to the principle of liking, people may be willing to listen to and comply with requests from those they like and whom they perceive to like them.[49] Constructs that increase liking include the perceived existence of similarities between the requester, or influencer, and the audience and compliments or praise from the influencer.[50] Thus, influencers may consider engaging in informal conversations that address similar beliefs, habits, hobbies, and so on and by offering positive, unstinting remarks to those they wish to influence.

According to the fifth principle of social influence, people tend to defer to experts or others who are in positions of authority. This aligns with the notion that highly credible sources are more persuasive. In practice, those who want to influence others should establish their knowledge and expertise, rather than assume that an audience is already aware of their credentials. This may include promoting one's degrees or awards or referencing relevant previous experiences.[51]

Finally, the sixth principle is that of scarcity: People want more of something when there is less of it available. In other words, when goods or opportunities are perceived as being less available, people perceive them as more valuable. As a result, advertisers often claim that goods or opportunities are available for a limited time only. To make use of this principle, influencers may consider informing receivers that they are being given exclusive information or that an opportunity is available only for a limited time. These techniques will work only if they are genuine; if they are not, an audience will lose enthusiasm and trust.[52]

Each of these principles demonstrates the power of social norms and beliefs. Further, the principles provide guidance regarding how to structure an IIP effort. A previously developed theory, tested with research, can serve as a powerful tool in designing and implementing a specific persuasive campaign.

Diffusion of Innovation Theory

Other theories and research on the role of social factors have considered how an idea spreads through a population or social group over time. One of the oldest is the diffusion-of-innovation theory.[53] According to this theory, innovations, such as new programs, practices, policies, and ideas, are communicated to an audience, and diffusion is the process by which these innovations are spread or communicated among those in a social system. Individuals share information among their social systems to

---

[49] Robert B. Cialdini, "Harnessing the Science of Persuasion," *Harvard Business Review*, October 2001b.

[50] Robert B. Cialdini and Noah J. Goldstein, "Social Influence: Compliance and Conformity," *Annual Review of Psychology*, Vol. 55, 2004.

[51] Cialdini, 2001b.

[52] Cialdini, 2001b.

[53] James W. Dearing, "Applying Diffusion of Innovation Theory to Intervention Development," *Research on Social Work Practice*, Vol. 19, No. 5, September 2009.

achieve a sense of shared understanding, which can contribute to change within the social system.[54]

Numerous factors determine the extent to which an innovation diffuses throughout a social system. According to the theory, innovations proceed through five phases before an individual can spread the idea through a social system. This is known as the innovation-decision process. The first phase is knowledge, which occurs when a person learns about an innovation and its functions. In this phase, the individual learns about the cause-and-effect relationships associated with the innovation. The next phase is persuasion, in which the individual develops an attitude toward the innovation. He or she subsequently makes a decision about whether to endorse the innovation by participating in activities that test it. If a decision is made to adopt it, the individual then implements, or utilizes, the innovation. Finally, he or she seeks to confirm or reinforce his or her decision by sharing his or her knowledge with others. If this process is successful, a person can become an agent of change and assist in spreading the innovation. If it is unsuccessful, the person may attempt to hinder the diffusion process.

In terms of adopting the innovation, people are theorized to fall into one of five categories: innovators, early adopters, early majority, late majority, and laggards. Innovators are those within a social system who are the first to adopt the new idea, program, or policy, and these individuals tend to be more adventurous, better educated, and better able to handle uncertainty than their peers. By contrast, laggards tend to be the least educated and least adventurous. The community or social system in which an individual operates, and the prevalent characteristics and personalities in these social systems, can influence the adoption of an innovation, or how many people fall into certain categories.[55]

The diffusion-of-innovation theory proposes that those promoting an innovation should attempt to encourage its diffusion from innovators to laggards. This involves addressing three individual-level factors: adopter characteristics, personalities, and communication behavior. Adopter characteristics include formal education and socioeconomic status. Personality traits include the ability to handle uncertainty. Communication behavior includes how a person communicates information about an innovation.

How an innovation spreads among those in a community has implications for the development and evaluation of a campaign.[56] For example, before program implementation, it may be worthwhile to identify innovators, opinion leaders, and influencers who can serve as effective agents of change in a social system. In assessing the efficacy

---

[54] Muhiuddin Haider and Gary L. Kreps, "Forty Years of Diffusion of Innovations: Utility and Value in Public Health," *Journal of Health Communication*, Vol. 9, Suppl. 1, 2004.

[55] Everett M. Rogers, *Diffusion of Innovations*, 4th ed., New York: Free Press, 2010.

[56] Author interview with Ronald Rice, May 9, 2013.

of a campaign, it is helpful to identify the groups to which an innovation has spread and who still needs to be targeted.

Systems Approach

In recognizing that a specific persuasive effort occurs within and is influenced by social systems, researchers have developed different theories and approaches to conceptualize and tailor efforts depending on the system of interest. For example, according to the ecological perspective, an individual's behaviors are heavily influenced by his or her environment and social context. This perspective proposes that an individual has interpersonal relationships, and these relationships occur within a community that operates within a society. Each of these social elements—interpersonal relationships, community, and society—is a different level of social influence. Because there are multiple levels of social influence, IIP interventions and evaluations of those interventions may need to occur at these different levels—from the individual to society.[57]

Other theories have addressed how to model IIP efforts as part of a larger system. These models recognize that common structures in a particular social system often go unrecognized in campaign development and evaluation.[58] First, there are related and interacting elements within certain boundaries in a social system. Second, there are shared goals within the system. Finally, there are certain environmental factors—specifically, inputs and constraints—within the system. Rather than targeting an audience at the individual level and attempting to expose individuals to a persuasive message that has one goal (e.g., to sell a product), the systems approach proposes that multiple components be considered during the development and implementation of an IIP effort, including the political climate, community characteristics, available media forums, and audience characteristics. This approach suggests that efforts that involve linear communication from the influencer to the receivers should be phased out or avoided, and those that involve dialogue and consideration of how receivers perceive the message should be emphasized.[59]

Applying this systems approach to a specific context, a rural Afghan community may be considered a system of interacting members. This community exists in a certain environment, which may be characterized by limited security, limited formal education, limited economic resources, and prevalent propaganda from groups like the Taliban. This community then receives inputs, including persuasive messages from an IIP effort, and the audience processes these inputs. The shared goals and perceived constraints of members within a system influence how the persuasive messages are processed or interpreted. After the audience processes the IIP effort's messages, planners may observe certain outputs in the system, such as improved security or reduced sup-

---

[57] Valente and Kwan, 2012.

[58] Rice and Foote, 2013.

[59] Author interview with Steve Corman, March 2013.

port for the Taliban. However, these outputs will be either promoted or discouraged by other factors, including support or competition from other systems.[60]

According to the systems approach, measurement should also be adjusted to address the system in which an IIP effort is taking place. Specifically, the dimensions of the system should be assessed before, during, and after an IIP effort. There are four initial stages in this systems-based approach: Identify broad goals, assumptions, and related efforts (stage 1); describe and specify the social system (stage 2); determine the initial states and system phases (stage 3); and identify inputs and potential constraints (stage 4). Each stage involves identifying the parameters of a system in which an IIP effort will be implemented. These four initial stages specifying the effort's parameters are followed by four more: Establish the short- and long-term goals of the IIP effort (stage 5); outline individual-level processes (stage 6); select the approaches that are most appropriate for meeting the effort's goals within the specific social system (stage 7); and determine the design implications (stage 8).[61] These stages can be complemented by input from specialists who are familiar with the social context and related prior or ongoing efforts.

A systems-based approach can be useful in complex environments like Afghanistan. Many factors can influence the efficacy of a particular effort, including preexisting beliefs and countereffort. As such, planners should temper their expectations and carefully consider the available information about the system when selecting an approach.

Causing Fear: Often Ineffective in IIP Efforts

Several campaigns have attempted to use fear to persuade audiences to change their behaviors. This tactic is especially prevalent in health-based campaigns, and it is often unsuccessful.[62] The general structure of fear appeals involves presenting audiences with a risk or threat (e.g., lung cancer), clarifying their vulnerability or susceptibility to this risk or threat (e.g., smoking causes lung cancer), and informing them that the threat is severe (e.g., lung cancer kills).[63] Audiences are then provided with options to protect themselves from this threat (e.g., stop smoking).

Various theories have attempted to explain the reasons for the lack of success in many fear appeals.[64] One of the oldest and best-known theories proposed that there is a curvilinear relationship between fear and persuasion. This suggests that low to

[60] Author interview with Ronald Rice, May 9, 2013.

[61] Author interview with Ronald Rice, May 9, 2013.

[62] R. F. Soames Job, "Effective and Ineffective Use of Fear in Health Promotion Campaigns," *American Journal of Public Health*, Vol. 78, No. 2, February 1988.

[63] Robert A. C. Ruiter, Charles Abraham, and Gerjo Kok, "Scary Warnings and Rational Precautions: A Review of the Psychology of Fear Appeals," *Psychology and Health*, Vol. 16, No. 6, 2001.

[64] Ruiter, Abraham, and Kok, 2001.

moderate levels of fear may increase the efficacy of a persuasive message. However, if the recipient of a message feels too much fear in response to the message, his or her defensive mechanisms will activate (e.g., dismissal and denial of the message), and the persuasive impact of the message will diminish. It is worth noting that there is very little empirical support for this proposed curvilinear relationship.

The parallel response model proposes that fear appeals may trigger two different kinds of coping mechanisms. The first, fear control, involves coping with one's own negative emotion of fear by denying the existence of the threat causing the fear. The second, danger control, involves coping with the threat, not one's emotions, by taking appropriate action against the threat. These two mechanisms may operate independently or one may overshadow the other. Building from the notion of danger coping and emotion coping, a later theory proposed that behavioral intentions are influenced by the seriousness of the threat and the receiver's perceived susceptibility. This aligns with the danger coping mechanism. Further, according to this theory, behavioral intentions are also influenced by receivers' expectations that they can respond to the threat and that these responses will be effective. Research assessing these associations has shown mixed support.[65]

Generally, the relationship between fear-based appeals and attitude change or behavioral change is complex. There is a strong potential for fear-based appeals to have the opposite effects of those intended. Similar critiques have also been directed at shame-based appeals.[66] Thus, IIP planners should carefully consider whether or not these appeals are appropriate to use in a particular context, and they should pretest them to assess their potential effects on a target audience.

Approaches to Organizing and Understanding IIP Efforts

Much of the theory discussed thus far has been developed and researched over several decades. More-recent approaches have focused on synthesizing the available information about behavioral influence. These approaches build from many of the concepts from previous theory and research and highlight potential ways to improve IIP planning.

MINDSPACE: Considering Automatic Processes
MINDSPACE is a mnemonic developed to address contextual influences of behaviors.[67] Rather than focus on factors that contribute to deliberate changes in individual

---

[65] Ruiter, Abraham, and Kok, 2001.

[66] Lindsay Abrams, "Obesity Campaigns: The Fine Line Between Educating and Shaming," *Atlantic Online*, September 16, 2012.

[67] P. Dolan, M. Hallsworth, D. Halpern, D. King, R. Metcalf, and I. Vlaev, "Influencing Behaviour: The MINDSPACE Way," *Journal of Economic Psychology*, Vol. 33, No. 1, February 2012.

cognitions and behavioral intentions, MINDSPACE instead summarizes environmental influences that can change behavior in more-automatic and less cognitively controlled ways. There are multiple contextual factors that can influence actions, and MINDSPACE focuses on nine that have some of the most robust effects: messenger, incentives, norms, defaults, salience, priming, affect, commitment, and ego.

Reactions to the *messenger* include the automatic compliance that people grant to individuals who are in positions of authority (e.g., older peers, researchers, doctors) and the impulsive responses that people have toward those about whom they feel negatively or positively. Aspects of *incentives* that can elicit subconscious reactions include the reference point against which a particular incentive is evaluated; a stronger aversion to loss than desire for gain; overemphasizing the small, unlikely probabilities of something occurring; placing money into distinct mental budgets (e.g., salary, expenses); and preferences for immediate payoffs over later ones. Further, *norms* involve the standard or customary behaviors in a society to which society members conform. For example, people tend to modify their behavior to be more similar to those in their community. *Defaults* involve the options available to a person who does not make a considered decision or take a deliberate action. Often, people accept the default option that they are provided. These four factors are captured by the *MIND* portion of *MINDSPACE*.

Salience involves the factors that catch attention, including those that are novel (an unusual or unexpected message), accessible (immediately observable), and simple (easily understood, like a slogan). Further, automatic processes can be stimulated through *priming*, or exposure to cues that can elicit certain reactions. Experienced emotions, or *affect*, can influence behavior change as well. For example, disgust can lead people to increase soap use.[68] Further, after showing *commitment* to a cause or action, people prefer to continue to behave in ways that are congruent with their commitment. Finally, *ego* involves tendencies to behave in ways that maintain positive and consistent perceptions of one's self. For example, people tend to automatically attribute positive experiences to their own actions and negative experiences to the actions of others. Salience, priming, affect, commitment, and ego are captured by the *SPACE* portion of *MINDSPACE*.

Although many tools (such as laws) force audience compliance, MINDSPACE summarizes subtle actions that policymakers may consider when attempting to influence behaviors.[69] These concepts have also been described in popular media as contributing to the "stickiness" of an idea, or the extent to which a message remains in an audience's mind. Specifically, the concepts that tend to influence people are simple,

---

[68] Dolan et al., 2012.

[69] Paul Dolan, Michael Hallsworth, David Halpern, Dominic King, and Ivo Vlaev, *Mindspace: Influencing Behaviour Through Public Policy*, London: Institute for Government, 2010.

unexpected (i.e., novel), concrete (i.e., easily understood), credible, emotional, and salient—concepts that overlap with the MINDSPACE approach.[70]

Behavior Change Wheel: Influencers, Interventions, and Policies

The behavior change wheel is a method for identifying the characteristics of certain efforts, or interventions and connecting them to the behaviors that the interventions seek to change (see Figure D.4).[71] Many people who design interventions may not use existing theory or research because they think these preexisting frameworks do not adequately address their particular needs. The behavior change wheel was designed to assist developers in identifying a broad approach that can be tailored to a specific effort, thereby assisting in the application of broad concepts.

The model of the behavior change wheel begins with the idea that, to generate a behavior, there must be capability, opportunity, and motivation. In other words, a person must have the mental and physical ability to engage in a behavior and the

Figure D.4
The Behavior Change Wheel
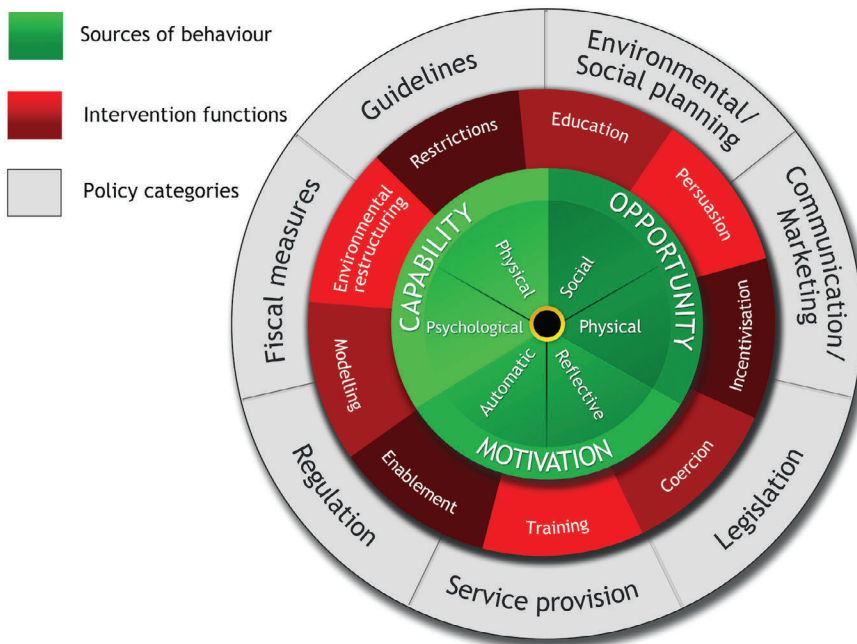
SOURCE: Mitchie, van Stralen, and West, 2011, Figure 2. Used under Creative Commons licensing guidelines (CC BY 4.0).
RAND *RR809/1-D.4*

---

[70] Heath and Heath, 2007.

[71] Susan Michie, Maartje M. van Stralen, and Robert West, "The Behaviour Change Wheel: A New Method for Characterizing and Designing Behaviour Change Interventions," *Implementation Science*, Vol. 6, 2011.

requisite drive (e.g., emotions, habits, decisions, goals). Further, external factors must permit or prompt the behavior. Any particular intervention may change one or more of these components. In addition, these three major components can be divided into subcomponents. Capability can be physical or mental. An opportunity can be physical or social. Motivation can involve intentional, reflective processes (i.e., planning) or automatic processes (i.e., emotions or habits).

There are also different categories of intervention. These can include training that involves developing skills, enablement through behavioral support and capacity building, modeling through the promotion of imitation, environmental restructuring through a change of context, and restricting the ability to perform a negative behavior. Further, intervention may include promoting knowledge and understanding, encouraging attitude change, creating reward-based incentives, and coercion through the threat of punishment in response to certain behaviors.

Policies are designed to support interventions that, in turn, target behaviors. Policies can include service delivery or provision, the regulation of behaviors or practices, measures to address financial costs (e.g., taxes), or guidelines that mandate sets of practices. In addition, policies may involve shaping or establishing jurisdiction over an environment, the use of different media forums, and the creation or modification of laws.

The behavior change wheel provides a pictorial representation of six components that influence behavior, nine intervention functions, and seven types of policies. The wheel can assist in quickly synthesizing information about an intervention. It is important to note, however, that the wheel has been criticized for not clearly articulating how one might move from identifying the underlying causes of a behavioral problem to determining the most effective intervention.[72]

The Initiatives Group Information Environment Assessment Model
The Initiatives Group has developed an assessment-based model of IIP that provides guidance on multiple assessment-relevant topics and a framework that applies theory to assessment.[73] The model begins by classifying the purpose of assessments—specifically, measuring processes, effects, and programs—and then provides informational support for effects assessments. Process assessments examine whether an organization's procedures are timely and efficient. Effects assessments address whether an organization's outputs are meeting goals and achieving desired results. Finally, programmatic assessments are related to effects assessments and involve examining ROI, or the financial value of impacts. Effects assessments are of particular interest in this context because they contribute to a better understanding of the extent of a campaign's success.

---

[72] Lee A. Rowland and Gaby van den Berg, *In Pursuit of a Contextual Diagnostic Approach to Behavior Change Interventions*, London: Behavioural Dynamics Institute, September 2012.

[73] The Initiatives Group, 2013.

To permit effects assessments, planners should establish clearly defined, measurable goals that are possible to accomplish. Further, planners should assess the will and capability of adversaries. The Initiatives Group's IE conditions framework, or theory of change, provides guidance for conceptualizing will and capability under given conditions (see Figure D.5). In this framework, *will* is defined as "the aggregate of variables that describe motivation and commitment to carry out an objective or execute a decision," and *capability* is defined as "the aggregate of instruments required to execute decisions."[74] Conditions are environmental variables that can be measured and influenced, and activities are the actions taken to influence conditions and cause effects.

The framework provides guidance about what elements to change and how to identify changes that will occur if certain effects have occurred. Building from social science theory, the framework outlines different components of will and how these components can be addressed and assessed.

Applying Theory to Practice Across Disciplines

Thus far, this appendix has focused on describing the principles and processes of different social science–focused theories and approaches. Diverse disciplines have consid-

Figure D.5
Information Environment Assessment Conditions Framework
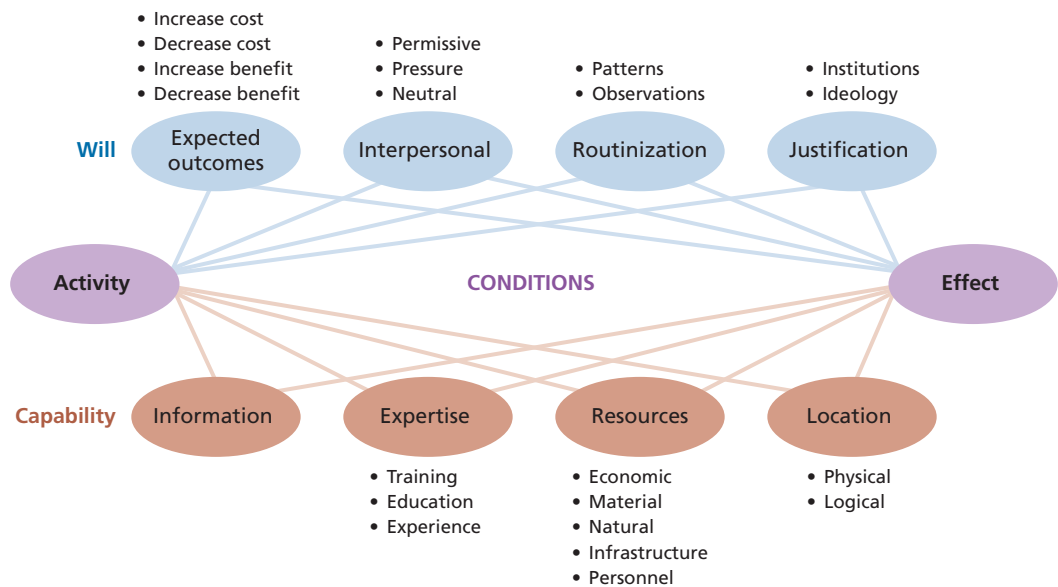
RAND *RR809/1-D.5*

---

[74] The Initiatives Group, 2013.

ered the application of theory to practice. These disciplines build from several of the previously discussed theoretical concepts and either implicitly or explicitly utilize their own theories of change. Although applied within a particular discipline, they may be informative for IIP efforts.

Business and Marketing

Persuasion theory and tactics are often used in business and marketing and sometimes utilize rigorous measurements to determine the effects of marketing efforts. The path to influencing behavior in business and marketing (the implicit theory of change) begins with the development of goals or identifying the clear and measurable objectives that apply to different stages of a persuasive process. Goals tailored to each stage of the persuasive process should incorporate the following concepts:

- Reach: How many or which customers were exposed to the message?
- Awareness: What new information should customers have?
- Comprehension: What should customers understand?
- Attitude: What should customers feel?
- Behavior: What should customers do?[75]

These tailored goals align with a broad objective of marketing activities, which is to "'funnel' customers from awareness to ultimately becoming loyal customers."[76]

Use of Social Media

When applying the previously outlined implicit theory of change to social media, businesses must establish their desired reach and raise awareness by disseminating relevant material. To assess these efforts, metrics could include unique visitors, page views, time spent on the site, and the number of comments associated with the message.[77] However, it is possible for businesses promoting a service or product to increase their reach and awareness without serving their ultimate goal (e.g., increased sales).

In business and marketing, social media is a communication tool that should be used for a purpose.[78] Thus, it is necessary to consider how customers interpret and respond to a message. Comprehension and attitudes can be measured through the amount of user-generated content and the valence of comments.[79] Finally, to assess

---

[75] Ketchum Global Research and Analytics, undated; interview with David Michaelson, April 1, 2013; interview with David Rockland, March 2013.

[76] Jeffery, 2010.

[77] Tia Fisher, "ROI in Social Media: A Look at the Arguments," *Journal of Database Marketing and Customer Strategy Management*, Vol. 16, No. 3, September 2009.

[78] Blanchard, 2011.

[79] Donna L. Hoffman and Marek Fodor, "Can You Measure the ROI of Your Social Marketing?" *MIT Sloan Management Review*, Vol. 52, No. 1, Fall 2010.

whether an effort has led to increased sales or changes in purchasing behavior, companies survey customers to inquire about how they learned of a product or track customers' purchases.[80]

Limitations

Although business and marketing provide some useful applications and tools, this discipline is not analogous to other domains, including DoD efforts to eliminate terrorism and support for terrorists. Product sales and purchase behavior are often the primary, if not the only, variable of interest in marketing.[81] Thus, trust and relationship building may be of less interest in these fields.[82] Further, the focus of advertisements is often product promotion rather than countering the message of an adversarial group or changing beliefs.[83] As such, research and practice in the discipline of social marketing, discussed next, may be of greater use for DoD-related projects.[84]

Public Communication and Social Marketing

Public communication, or social marketing, builds from techniques used in business and marketing but applies these techniques to efforts that seek to benefit individuals and communities.[85] Rather than focusing on product sales and purchasing behaviors, social marketing efforts seek to change individual and community behaviors for the purpose of promoting social good. For example, health-promotion efforts often utilize social marketing.[86] To improve effectiveness, social marketing efforts often involve identifying a target audience and attempting to convince this audience of the rewards of altering a given behavior.[87] Thus, the audience is informed of the issue and then targeted with a call to action to address the issue.[88]

South Dakota 24/7 Sobriety Project

South Dakota developed the 24/7 Sobriety Project to reduce alcohol consumption among chronic drunk drivers. The project seeks to increase awareness of the conse-

---

[80] Fisher, 2009.

[81] Author interview with Victoria Romero, June 24, 2013.

[82] Author interview with Craig Hayden, June 21, 2013.

[83] Author interview with Victoria Romero, June 24, 2013.

[84] William A. Smith, "Social Marketing: An Overview of Approach and Effects," *Injury Prevention*, Vol. 12, Suppl. 1, June 2006.

[85] Charles K. Atkin and Ronald E. Rice, "Theory and Principles of Public Communication Campaigns," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2012.

[86] Schneider and Cheslock, 2003.

[87] Schneider and Cheslock, 2003.

[88] Author interview on a not-for-attribution basis, April 18, 2013.

quences of alcohol consumption among these individuals. South Dakota residents who have served jail time for DUIs must submit to continuous monitoring, involving either two breathalyzer tests per day or wearing an alcohol-monitoring bracelet.[89] Those who consume alcohol are immediately served with sanctions. The monitoring and sanctions are designed to increase awareness of the consequences of violating the program's terms and offenders' certainty of punishment, which are theorized to reduce problem drinking. As evidence of its effectiveness, the program points to decreases in repeated DUI arrests, domestic violence arrests, and traffic accidents.

Cure Violence

Another social marketing campaign that applies a theory of change to practice is Cure Violence, originally known as CeaseFire. This program targets specific audiences identified as high risk. One program staff member described the approach as follows:

> They focus on targeting a narrow subset of individuals that are identified as high risk. High-risk individuals are those who meet at least four of the following criteria: between 16 and 25 years of age, have a history of arrests or offenses, have served time in prison, have been the victim of a shooting, and are involved in illegal street activity. They then utilize specific groups of individuals to target established causal factors that contribute to violence among this high-risk group.[90]

The program includes six categories of actions to reduce the risk of violent behaviors, increase the cost of engaging in such behaviors, change norms surrounding violence, and provide alternatives to violence: street intervention, client outreach, clergy involvement, community mobilization, educational campaign work, and police and prosecution. Multiple assessments of the program in different U.S. neighborhoods suggest that the program has had some success in achieving its ultimate goal of violence reduction. Comparisons between neighborhoods that have implemented the program and similar neighborhoods that have not show reductions in shootings and gang homicides in neighborhoods with the program.[91]

Limitations

A potential limitation in the application of social marketing theory and practice to DoD efforts is that social marketing efforts often involve targeting a narrow audience, such as DUI offenders in South Dakota or high-risk individuals in specific neighborhoods. However, in a defense context, more limitation might be a strength. Narrower

---

[89] Beau Kilmer, Nancy Nicosia, Paul Heaton, and Greg Midgette, "Efficacy of Frequent Monitoring with Swift, Certain, and Modest Sanctions for Violations: Insights from South Dakota's 24/7 Sobriety Project," *American Journal of Public Health*, Vol. 103, No. 1, January 2013; interview with Beau Kilmer, March 14, 2013.

[90] Author interview with Joshua Gryniewicz, August 23, 2013.

[91] Welsey G. Skogan, Susan M. Hartnett, Natalie Bump, and Jill Dubois, *Evaluation of CeaseFire-Chicago*, Evanston, Ill.: Institute for Policy Research, Northwestern University, June 2009.

and more precisely defined target audiences for defense IIP objectives would certainly make assessment more straightforward, and it might make achieving those objectives more manageable as well.

Critics also note that some attempts to apply traditional marketing techniques to social marketing campaigns may be misguided, as the characteristics of social marketing campaigns and interactions between implementers and their target audiences are different from those of sellers and buyers.[92]

Public Diplomacy

Public diplomacy may be defined as "an international actor's attempt to manage the international environment through engagement with a foreign public."[93] Public diplomacy has been likened to persuasion, and it may be most successful when it aligns with principles and theories of persuasion.[94] To assess the success of public diplomacy efforts, clear goals and measurements are needed. Measuring public diplomacy efforts can be difficult, in part due to the length of time required before success can be seen and the multiple events and programs that make up a public diplomacy effort.[95] However, incorporating measurement is worthwhile, as it can assist with allocating resources, identifying best practices, and demonstrating effectiveness to policymakers.

To evaluate a public diplomacy effort, a country must identify its goals for engaging with its target audience. Goals may include improving perceptions of the country (nation branding), building support for country objectives, or improving mutual understanding with other countries.[96] Steps in addressing these goals could include: (1) determining baseline perceptions among a target audience, (2) disseminating targeted persuasive communication, (3) disseminating targeted messages to foreign media and leaders, (4) changing attitudes among the target audience, and (5) changing foreign support.[97]

Rebranding Switzerland

One public diplomacy effort that has been considered successful was undertaken by Presence Switzerland to improve international perceptions of and investments in Switzerland.[98] In the late 1990s, Switzerland had a poor international image. Mil-

---

[92] Author interview with Steve Booth-Butterfield, January 7, 2013.

[93] Nicholas J. Cull, *Public Diplomacy: Lessons from the Past*, Los Angeles, Calif.: Figueroa Press, 2009.

[94] Mohan J. Dutta-Bergman, "U.S. Diplomacy in the Middle East: A Critical Culture Approach," *Journal of Communication Inquiry*, Vol. 30, No. 2, April 2006.

[95] Banks, 2011.

[96] Banks, 2011.

[97] Michael Egner, *Between Slogans and Solutions: A Frame-Based Assessment Methodology for Public Diplomacy*, dissertation, Santa Monica, Calif.: Pardee RAND Graduate School, RGSD-255, December 2009.

[98] Cull, 2009.

lions of dollars had been deposited in Swiss banks during World War II, a large portion of which may have been taken from victims of the Holocaust. The phrase "Swiss bank account," ubiquitous in popular media, was shorthand for a place where criminals could hide wealth from tax collectors and law enforcement agencies. Previous efforts to address the country's reputation had been unsuccessful. Presence Switzerland attempted to address the problem by involving members from diverse sectors, including the Swiss media, banking and tourism industries, and foreign ministries. As its first task, Presence Switzerland identified target countries in which to initiate public diplomacy activities, including Germany, Italy, France, the United States, and the United Kingdom. Then, it administered image surveys in these countries, including polling and media analysis. Among other things, the surveys showed that international audiences had little knowledge of Switzerland's humanitarian commitment or modernity. Subsequently, Swiss efforts focused on increasing exposure to information about the country's positive attributes among targeted audiences. Presence Switzerland promotes its message in conjunction with international events, sponsoring a pavilion at the World Expo and the House of Switzerland at the Olympics, and operates a website in eight languages. Planners continue to collect international surveys in target countries, allowing them to identify changes in knowledge and attitudes regarding Switzerland, which they can then address as part of the campaign. In 2012, to attract younger audiences, Presence Switzerland released a series of free games for mobile devices featuring Swiss legends and literary characters, and its website features videos on a range of topics. The campaign's data suggest that international attitudes toward Switzerland improved significantly after its implementation.[99]

Politics

Another area in which persuasion efforts play a key role in influencing behavior is politics. In the words of researchers Michael Cobb and James Kuklinski, "Persuasion, changing another's beliefs and attitudes, is about influence; and influence is the essence of politics."[100] The three major stages of "winning a vote" are registering a voter, demonstrating to the voter that a particular candidate should be his or her preferred choice, and mobilizing the voter to go to the polls.[101] Measurements can assist in determining which areas to target for political persuasion efforts, how to target these areas, and the success of efforts implemented in these areas.

---

[99] For more on the campaign's process for continuously monitoring its effects, see Federal Department of Foreign Affairs, Switzerland, "Monitoring and Analysis," web page, last updated November 26, 2014.

[100] Michael D. Cobb and James H. Kuklinski, "Changing Minds: Political Arguments and Political Persuasion," *American Journal of Political Science*, Vol. 41, No. 1, January 1997.

[101] Sasha Issenberg, *The Victory Lab*, New York: Crown Publishers, 2012.

Obama's Presidential Campaigns

The success of President Barack Obama's 2008 and 2012 campaigns has been credited, in part, to the tactical use of theory and data. Campaign planners calculated the probability that potential voters would cast a ballot and endorse Obama,[102] and they continually updated their statistical models with newly collected data. This ensured that the Obama campaign had up-to-date information regarding voters' behavioral intentions. The collected data were then used to target appropriate audiences through phone calls, door-to-door campaigning, and paid political advertisements, ensuring that potential Obama voters were provided with information that they perceived as interesting and relevant.

Implications for DoD IIP Efforts

In business and marketing, planners have developed implicit theories of change that consider awareness, comprehension, attitudes, and behaviors. However, the applicability of business efforts to DoD IIP efforts may be limited by the strong focus on a business goal, such as increased sales. As a result, social marketing efforts may be more applicable to DoD efforts. Social marketing seeks to produce behavioral change within a community, and business-related efforts have developed and utilized innovative measures and theories of change that may have utility for IIP planners. Similarly, persuasion efforts in public diplomacy and politics have sought to affect the attitudes and behaviors of broad audiences by collecting and using data in effective ways.

Practitioners should carefully consider the applicability of research from different disciplines when designing an IIP effort. The research should inform, but not dictate, planning and assessment. That said, there are some common themes across disciplines and approaches. For example, characteristics of the audience and context should guide message content and delivery mode to increase the chances of prompting a desired behavior change. To determine these characteristics, baseline information regarding the context of interest should be collected, analyzed, and applied during the design of an IIP effort. Further, data collection during and after the effort can assist in refining processes and, ideally, ensuring success.

---

[102] Sasha Issenberg, "How President Obama's Campaign Used Big Data to Rally Voters," *MIT Technology Review Magazine*, January–February 2013.

References

Abrams, Lindsay, "Obesity Campaigns: The Fine Line Between Educating and Shaming," *Atlantic Online*, September 16, 2012. As of October 20, 2014:
http://www.theatlantic.com/health/archive/2012/09/
obesity-campaigns-the-fine-line-between-educating-and-shaming/262401

Affholter, Dennis, "Outcome Monitoring," in Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds., *Handbook of Practical Program Evaluation*, San Francisco: Jossey-Bass, 1994, pp. 96–118.

Ajzen, Icek, "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, December 1991, pp. 179–211.

Ajzen, Icek, and Martin Fishbein, "A Theory of Reasoned Action," in *Understanding Attitudes and Predicting Social Behavior*, Upper Saddle River, N.J.: Pearson, 1980.

Akers, Ronald L., and A. L. Silverman, "Toward a Social Learning Model of Violence and Terrorism," in Margaret A. Zahn, Henry H. Brownstein, and Shelly L. Jackson, eds., *Violence: From Theory to Research*, Cincinnati, Ohio: Matthew Bender and Co., 2004.

Altai Consulting, "Afghan Media in 2010," prepared for the U.S. Agency for International Development, 2010. As of August 22, 2014:
http://www.altaiconsulting.com/docs/media

Andreason, Alan R., "Social Marketing: Its Definition and Domain," *Journal of Public Policy and Marketing*, Vol. 13, No. 1, Spring 1994, pp. 108–114.

Aronson, Elliot, Timothy D. Wilson, and Robin M. Akert, *Social Psychology*, 5th ed., Upper Saddle River, N.J.: Prentice Hall, 2005.

Arsenault, Amelia, Sheldon Himelfarb, and Susan Abbott, *Evaluating Media Interventions in Conflict Countries*, Washington, D.C.: United States Institute of Peace, 2011. As of August 22, 2014:
http://www.usip.org/sites/default/files/resources/PW77.pdf

Atkin, Charles K., and Ronald E. Rice, "Theory and Principles of Public Communication Campaigns," in Ronald E. Rice and Charles K. Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2012, pp. 3–20.

———, "Advances in Public Communication Campaigns," in Erica Scharrer, ed., *The International Encyclopedia of Media Studies*, Vol. 5, London: Wiley-Blackwell, 2013, pp. 526–551.

Atkinson, Carol, "Does Soft Power Matter? A Comparative Analysis of Student Exchange Programs 1980–2006," *Foreign Policy Analysis*, Vol. 6, No. 1, January 2010, pp. 1–22.

Averch, Harvey A., "Using Expert Judgment," in Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds., *Handbook of Practical Program Evaluation*, San Francisco: Jossey-Bass, 2004, pp. 292–309.

Babbie, Earl, *Survey Research Methods*, 2nd ed., Belmont, Calif.: Wadsworth Publishing Company, 1990.

Bandura, Albert, Dorothea Ross, and Sheila A. Ross, "Transmission of Aggression Through Imitation of Aggressive Models," *Journal of Abnormal and Social Psychology*, Vol. 63, No. 3, November 1961, pp. 575–582.

Banks, Robert, *A Resource Guide to Public Diplomacy Evaluation*, Los Angeles, Calif.: Figueroa Press, November 2011.

"Barcelona Declaration of Measurement Principles," 2nd European Summit on Measurement, International Association for Measurement and Evaluation of Communication, July 19, 2010. As of August 3, 2014:
http://amecorg.com/2012/06/barcelona-declaration-of-measurement-principles

Bator, Renee J., and Robert B. Cialdini, "The Application of Persuasion Theory to the Development of Effective Proenvironmental Public Service Announcements," *Journal of Social Issues*, Vol. 56, No. 3, Fall 2000, pp. 527–541.

Becker, David C., and Robert Grossman-Vermaas, "Metrics for the Haiti Stabilization Initiative," *Prism*, Vol. 2, No. 2, March 2011, pp. 145–158.

Biersmith, Eric, "Logic Model as a Tool to Evaluate Prevention," paper presented at Evaluation 2013, the annual conference of the American Evaluation Association, Washington, D.C., October 14–19, 2013.

Blakley, Johanna, "Movies for a Change," presentation at TEDxPhoenix, February 12, 2012. As of October 20, 2014:
https://www.youtube.com/watch?v=Pb0FZPzzWuk

Blanchard, Olivier, *Social Media ROI: Managing and Measuring Social Media Efforts in Your Organization*, Indianapolis, Ind.: Que, 2011.

Blanken, Leo J., and Jason J. Lepore, *Performance Measurement in Military Operations: Information Versus Incentives*, Monterey and San Luis Obispo, Calif.: Naval Postgraduate School and California Polytechnic State University, November 12, 2012.

Blumenthal, Mark, "Gallup Presidential Poll: How Did Brand-Name Firm Blow Election?" *Huffington Post*, March 8, 2013. As of October 20, 2014:
http://www.huffingtonpost.com/2013/03/08/gallup-presidential-poll_n_2806361.html

Bohan-Baker, Marielle, "Pitching Policy Change," *Evaluation Exchange*, Vol. 7, No. 1, Winter 2001, pp. 3–4.

Booth-Butterfield, Steve, "Standard Model," *Persuasion Blog*, undated. As of October 20, 2014:
http://healthyinfluence.com/wordpress/steves-primer-of-practical-persuasion-3-0/standard-model/

Brick, J. M., and G. Kalton. "Handling Missing Data in Survey Research," *Statistical Methods in Medical Research*, Vol. 5, No. 3, September 1996, pp. 215–238.

British Council, *Trust Pays: How International Cultural Relationships Build Trust in the UK and Underpin the Success of the UK Economy*, Edinburgh, UK, 2012. As of October 17, 2014:
http://www.britishcouncil.org/sites/britishcouncil.uk2/files/trust-pays-report.pdf

———, *Annual Report: 2012–13*, London, March 31, 2013. As of October 20, 2014:
http://www.britishcouncil.org/sites/britishcouncil.uk2/files/annual-report-2012-13.pdf

Bruce, Sarah, and Mary Tiger, *A Review of Research Relevant to Evaluating Social Marketing Mass Media Campaigns*, Durham, N.C.: Clean Water Education Partnership, undated.

Bulmer, Martin, "Introduction: The Problem of Exporting Social Survey Research," *American Behavioral Scientist*, Vol. 42, No. 2, October 1998, pp. 153–167.

Bumiller, Elisabeth, "We Have Met the Enemy and He Is PowerPoint," *New York Times*, April 26, 2010.

Campbell, Jason, Michael O'Hanlon, and Jeremy Shapiro, *Assessing Counterinsurgency and Stabilization Missions*, Washington, D.C.: Brookings Institution, Policy Paper No. 14, May 2009.

Carroll, Wallace, *Persuade or Perish*, New York: Houghton Mifflin, 1948.

Chaiken, Shelly, "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," *Journal of Personality and Social Psychology*, Vol. 39, No. 5, November 1980, pp. 752–766.

Chatterjee, Joyee S., Anurudra Bhanot, Lauren B. Frank, Sheila T. Murphy, and Gerry Power, "The Importance of Interpersonal Discussion and Self-Efficacy in Knowledge, Attitude, and Practice Models," *International Journal of Communication*, Vol. 3, 2009, pp. 607–634.

Childs, Geoffrey, "Military Information Support to Contingency Operations in Libya," *Special Warfare*, Vol. 26, No. 1, January–March 2013.

Chouinard, Jill Anne, J. Bradley Cousins, and Swee C. Goh, "Case 7: United Way of Greater Toronto (UWGT)," in J. Bradley Cousins and Isabelle Bourgeois, eds., *Organizational Capacity to Do and Use Evaluation*, No. 141, Spring 2014.

Cialdini, Robert B., *Influence: Science and Practice*, 4th ed., Boston: Allyn and Bacon, 2001a.

———, "Harnessing the Science of Persuasion," *Harvard Business Review*, October 2001b, pp. 72–81.

Cialdini, Robert B., John T. Cacioppo, Rodney Bassett, and John A. Miller, "Low-Ball Procedure for Producing Compliance: Commitment Then Cost," *Journal of Personality and Social Psychology*, Vol. 36, No. 5, May 1978, pp. 463–476.

Cialdini, Robert B., Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter, "Managing Social Norms for Persuasive Impact," *Social Influence*, Vol. 1, No. 1, 2006, pp. 3–15.

Cialdini, Robert B., and Noah J. Goldstein, "Social Influence: Compliance and Conformity," *Annual Review of Psychology*, Vol. 55, 2004, pp. 591–621.

Cioppa, Thomas M., "Operation Iraqi Freedom Strategic Communication Analysis and Assessment," *Media, War, and Conflict*, Vol. 2, No. 1, April 2009, pp. 25–45.

Clary, Tim A., *USAID/Haiti: Social Marketing Assessment, 2008*, Washington, D.C.: Global Health Technical Assistance Project, 2008. As of October 20, 2014:
http://pdf.usaid.gov/pdf_docs/Pnadn850.pdf

Coalition for Innovative Media Measurement, *CIMM Lexicon 1.0, Terms and Definitions: A Common Language for Set-Top Box Media Measurement*, New York, May 2010. As of October 20, 2014:
http://cimmusorg.startlogic.com/wp-content/uploads/2012/08/RPD_LEX.10.pdf

Cobb, Michael D., and James H. Kuklinski, "Changing Minds: Political Arguments and Political Persuasion," *American Journal of Political Science*, Vol. 41, No. 1, January 1997, pp. 88–121.

Coffman, Julia, *Public Communication Campaign Evaluation*, Washington, D.C.: Communications Consortium Media Center, May 2002.

Coffman, Julia, and Ehren Reed, *Unique Methods in Advocacy Evaluation*, Washington, D.C.: Innovation Network, 2009. As of October 20, 2014:
http://www.innonet.org/resources/files/Unique_Methods_Brief.pdf

Connable, Ben, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, Santa Monica, Calif.: RAND Corporation, MG-1086-DOD, 2012. As of August 22, 2014:
http://www.rand.org/pubs/monographs/MG1086.html

Conroy-Krutz, Jeffrey, and Devra Coren Moehler, "Moderation from Bias: A Field Experiment on Partisan Media in a New Democracy," draft manuscript, May 20, 2014.

Corcoran, Nova, "Theories and Models in Communicating Health Messages," in Nova Corcoran, ed., *Communicating Health: Strategies for Health Promotion*, Thousand Oaks, Calif.: Sage Publications, 2007, pp. 5–31.

Corman, Steven, "Understanding Extremists' Use of Narrative to Influence Contest Populations," paper prepared for the Workshop on Mapping Ideas: Discovering and Information Landscape, San Diego State University, San Diego, Calif., June 29–30, 2011. As of October 20, 2014:
http://mappingideas.sdsu.edu/old_Mappingideas/SummerWorkshop/2011/Papers/Corman_Position.pdf

Cosgrave, John, Ben Ramalingam, and Tony Beck, *Real-Time Evaluations of Humanitarian Action: An ALNAP Guide*, pilot version, London: Active Learning Network for Accountability and Performance in Humanitarian Action, 2009.

Crano, William D., *Lessons Learned from Media-Based Campaigns, or, It Takes More Than Money and Good Intentions*, Claremont, Calif.: Claremont Graduate University, 2002. As of October 20, 2014:
http://www.unodc.org/documents/drug-prevention-and-treatment/PreventionStandards/meeting2012_01/-_crano_media.pdf

———, "Theory-Driven Evaluation and Construct Validity," in Stewart I. Donaldson and Michael Scriven, eds., *Evaluating Social Programs and Problems: Visions for the New Millennium*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2003.

Crano, William D., and Marilynn B. Brewer, *Principles and Methods of Social Research*, 2nd ed., Mahwah, N.J.: Lawrence Erlbaum Associates, 2002.

Cull, Nicholas J., "Public Diplomacy: Taxonomies and Histories," *Annals of the American Academy of Political and Social Science*, Vol. 616, No. 1, March 2008, pp. 31–54.

———, *Public Diplomacy: Lessons from the Past*, Los Angeles, Calif.: Figueroa Press, 2009.

Cull, Nicholas J., and Ali Fisher, *The Playbook: Case Studies of Engagement*, online database, undated. As of February 18, 2015:
https://playbookbeta.wordpress.com

Daugherty, William E., and Morris Janowitz, *A Psychological Warfare Casebook*, Baltimore, Md.: Johns Hopkins University Press, 1958.

Davidson, Andrew R., and James Jaccard, "Variables That Moderate the Attitude-Behavior Relation: Results of a Longitudinal Study," *Journal of Personality and Social Psychology*, Vol. 37, No. 8, August 1979, pp. 1364–1376.

Dearing, James W., "Applying Diffusion of Innovation Theory to Intervention Development," *Research on Social Work Practice*, Vol. 19, No. 5, September 2009, pp. 503–518.

Debeljak, Klara, *Youth in Iran: A Story Half Told: Values, Priorities and Perspectives of Iranian Youth*, Young Publics Research Paper Series No. 1, Washington, D.C.: InterMedia, May 2013. As of October 20, 2014:
http://www.intermedia.org/wp-content/uploads/Young-Publics-Research-Paper-Series-Iran.pdf

Debeljak, Klara, and Joe Bonnell, *Citizen Access to Information in Papua New Guinea*, Washington, D.C.: InterMedia, June 2012. As of October 20, 2014:
http://www.intermedia.org/wp-content/uploads/PNG-Report.pdf

Del Veccio, Gene, "Got Milk? Got Fired: 5 Valuable Lessons That All Executives Must Heed," *Huffington Post*, March 12, 2014. As of November 20, 2014:
http://www.huffingtonpost.com/gene-del-vecchio/got-milk-got-fired-5-valu_b_4938176.html

De McLaurin, Ronald, Carl F. Rosenthal, and Sarah A. Skillings, eds., *The Art and Science of Psychological Operations: Case Studies of Military Application*, Vol. 1, Washington, D.C.: American Institutes for Research, April 1976a.

———, *The Art and Science of Psychological Operations: Case Studies of Military Application*, Vol. 2, Washington, D.C.: American Institutes for Research, April 1976b.

DeVellis, Robert F., *Scale Development: Theory and Applications*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2012.

Development Media International, "Proving Impact," web page, undated. As of October 20, 2014:
http://www.developmentmedia.net/proving-impact

Dewey, Caitlin, "Measuring Happiness Through Twitter," *Washington Post*, May 6, 2013.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian, *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd ed., Hoboken, N.J.: John Wiley and Sons, 2009.

Dolan, P., M. Hallsworth, D. Halpern, D. King, R. Metcalf, and I. Vlaev, "Influencing Behaviour: The MINDSPACE Way," *Journal of Economic Psychology*, Vol. 33, No. 1, February 2012, pp. 264–277.

Dolan, Paul, Michael Hallsworth, David Halpern, Dominic King, and Ivo Vlaev, *Mindspace: Influencing Behaviour Through Public Policy*, London: Institute for Government, 2010.

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

———, *Program Theory–Driven Evaluation: Strategies and Applications*, New York: Lawrence Erlbaum Associates, 2007.

Downes-Martin, Stephen, "Operations Assessment in Afghanistan Is Broken: What Is to Be Done?" *Naval War College Review*, Vol. 64, No. 4, Fall 2011, pp. 103–125.

Dutta-Bergman, Mohan J., "U.S. Public Diplomacy in the Middle East: A Critical Culture Approach," *Journal of Communication Inquiry*, Vol. 30, No. 2, April 2006, pp. 102–124.

Egner, Michael, *Between Slogans and Solutions: A Frame-Based Assessment Methodology for Public Diplomacy*, dissertation, Santa Monica, Calif.: Pardee RAND Graduate School, RGSD-255, December 2009. As of October 20, 2014:
http://www.rand.org/pubs/rgs_dissertations/RGSD255.html

Eles, P. T., E. Vincent, B. Vasiliev, and K. M. Banko, *Opinion Polling in Support of the Canadian Mission in Kandahar: A Final Report for the Kandahar Province Opinion Polling Program, Including Program Overview, Lessons, and Recommendations*, Ottawa, Ont.: Defence R&D Canada, Centre for Operational Research and Analysis, DRDC CORA TR 2012-160U, September 2012.

Entman, Robert M., "Framing: Toward Clarification of a Fractured Paradigm," *Journal of Communication*, Vol. 43, No. 4, December 1993, pp. 51–58.

Eppler, Martin J., and Peter Muenzenmayer, "Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology," *Proceedings of the 7th International Conference on Information Quality*, Cambridge, Mass.: MIT Sloan School of Management, 2002, pp. 187–196.

Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang, "Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses," *Behavior Research Methods*, Vol. 41, No. 4, November 2009, pp. 1149–1160.

Federal Department of Foreign Affairs, Switzerland, "Monitoring and Analysis," web page, last updated November 26, 2014. As of December 1, 2014:
https://www.dfae.admin.ch/eda/en/fdfa/fdfa/communication-abroad/monitoring-analysis.html

Fehlenberg, Kate, "Critical Juncture: Applying Assessment Tools and Approaches to Scaling-Up: A New Focus for External Validity," paper presented at Evaluation 2013, the annual conference of the American Evaluation Association, Washington, D.C., October 14–19, 2013.

Festinger, Leon, and James M. Carlsmith, "Cognitive Consequences of Forced Compliance," *Journal of Abnormal and Social Psychology*, Vol. 47, No. 2, March 1959, pp. 203–210.

Fishbein, Martin, and Robert Hornik, "Measuring Media Exposure: An Introduction to the Special Issue," *Communication Methods and Measures*, Vol. 2, Nos. 1–2, 2008, pp. 1–5.

Fishbein, Martin, Harry C. Triandis, Frederick H. Kanfer, Marshall Becker, Susan E. Middlestadt, and Anita Eichler, "Factors Influencing Behavior and Behavior Change," in Andrew Baum, Tracey A. Revenson, and Jerome E. Singer, eds., *Handbook of Health Psychology*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2001, pp. 3–17.

Fisher, Tia, "ROI in Social Media: A Look at the Arguments," *Journal of Database Marketing and Customer Strategy Management*, Vol. 16, No. 3, September 2009, pp. 189–195.

Fiske, Susan T., *Social Beings: A Core Motives Approach to Social Psychology*, Hoboken, N.J.: John Wiley and Sons, 2004.

Fogg, B. J., "A Behavior Model for Persuasive Design," in *Proceedings of the 4th International Conference on Persuasive Technology*, New York: ACM, 2009.

Fowler, Floyd J., *Improving Survey Questions: Design and Evaluation*, Thousand Oaks, Calif.: Sage Publications, 1995.

Freedman, Jonathan L., and Scott C. Fraser, "Compliance Without Pressure: The Foot-in-the-Door Technique," *Journal of Personality and Social Psychology*, Vol. 2, No. 2, August 1966, pp. 195–202.

Gaarder, Marie, and Jeannie Annan, *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*, New York: World Bank Independent Evaluation Group, June 2013.

Gagliardone, Iginio, and Nicole A. Stremlau, "Public Opinion Research in a Conflict Zone: Grassroots Diplomacy in Darfur," *International Journal of Communication*, Vol. 2, 2008, pp. 1085–1113.

Garland, Ian, "Return Path Data: A 21st Century Business Tool," undated. As of October 20, 2014:
http://www.casbaa.com/privacy-policy/1297-return-path-data-a-21st-century-business-tool

Gass, Robert H., and John S. Seiter, *Persuasion: Social Influence and Compliance Gaining*, 5th ed., New York: Pearson, 2014.

Gates, Robert, Secretary of Defense, "Strategic Communication and Information Operations in the DoD," memorandum, Washington, D.C., January 25, 2011.

Gilligan, Michael J., Eric N. Mvukiyehe, and Cyrus Samii, *Reintegrating Rebels into Civilian Life: Quasi-Experimental Evidence from Burundi*, Washington, D.C.: United States Institute of Peace, 2010.

Glanz, Karen, and Donald B. Bishop, "The Role of Behavioral Science Theory in Development and Implementation of Public Health Interventions," *Annual Review of Public Health*, Vol. 31, 2010, pp. 399–418.

"Got Milk? Is Here to Stay," PRNewswire, March 3, 2014. As of November 20, 2014: http://www.prnewswire.com/news-releases/got-milk-is-here-to-stay-248187531.html

Graham-Silverman, Adam, "Fighting AIDS in Mozambique," *Slate*, May 31, 2005. As of October 20, 2014: http://www.slate.com/articles/news_and_politics/dispatches/features/2005/fighting_aids_in_mozambique/the_abcsof_aids_prevention.html

Grier, Sonya, and Carol A. Bryant, "Social Marketing in Public Health," *Annual Review of Public Health*, Vol. 26, 2005, pp. 319–339.

Haider, Muhiuddin, and Gary L. Kreps, "Forty Years of Diffusion of Innovations: Utility and Value in Public Health," *Journal of Health Communication*, Vol. 9, Suppl. 1, 2004, pp. 3–11.

Haims, Marla C., Melinda Moore, Harold D. Green, and Cynthia Clapp-Wincek, *Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects*, Santa Monica, Calif.: RAND Corporation, TR-784-OSD, 2011. As of August 22, 2014: http://www.rand.org/pubs/technical_reports/TR784.html

Hanssen, Carl E., Frances Lawrenz, and Diane O. Dunet, "Concurrent Meta-Evaluation: A Critique," *American Journal of Evaluation*, Vol. 29, No. 4, December 2008, pp. 572–582.

Harrell, Margaret C., and Melissa A. Bradley, *Data Collection Methods: Semi-Structured Interviews and Focus Groups*, Santa Monica, Calif.: RAND Corporation, TR-718-USG, 2009. As of October 20, 2014: http://www.rand.org/pubs/technical_reports/TR718.html

Haug, Magne, "The Use of Formative Research and Persuasion Theory in Public Communication Campaigns: An Anti-Smoking Campaign Study," paper presented at the Nordic Mass Communication Research Conference, Reykjavik, Iceland, August 10–14, 2001. As of February 10, 2014: http://people.oregonstate.edu/~flayb/TTI%20citations/Substance%20use/Haug%2001%20Anti-smoking%20campaigns.pdf

Hayden, Craig, *The Rhetoric of Soft Power: Public Diplomacy in Global Contexts*, Landham, Md.: Lexington Books, 2012.

———, "Another Perspective on IIP Social Media Strategy," *Intermap*, July 23, 2013. As of August 22, 2014: http://intermap.org/2013/07/23/another-perspective-on-iip-social-media-strategy

Headquarters, U.S. Department of the Army, *Psychological Operations Leaders Planning Guide*, Graphic Training Aid 33-01-001, Washington, D.C., November 2005.

———, *Inform and Influence Activities*, Field Manual 3-13, Washington, D.C., January 2013a.

———, *Military Information Support Operations*, Field Manual 3-53, Washington, D.C., January 2013b.

Headquarters, U.S. Department of the Army, and Headquarters, U.S. Marine Corps, *Psychological Operations, Tactics, Techniques, and Procedures*, Field Manual 3-05.301/Marine Corps Reference Publication 3-40.6A, Washington, D.C., December 2003.

Healey, Joseph F., *Statistics: A Tool for Social Research*, 9th ed., Belmont, Calif.: Wadsworth/Cengage Learning, 2012.

Heath, Chip, and Dan Heath, *Made to Stick: Why Some Ideas Survive and Others Die*, New York: Random House, 2007.

Hedonometer, homepage, undated. As of October 20, 2014:
http://hedonometer.org

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan, "Most People Are Not Weird," *Nature*, Vol. 466, No. 7302, July 1, 2010, p. 29.

Henry, Gary T., and Craig S. Gordon, "Tracking Issue Attention: Specifying the Dynamics of the Public Agenda," *Public Opinion Quarterly*, Vol. 65, No. 2, 2001, pp. 157–177.

Heppner, P. Paul, Dennis M. Kivlighan, and Bruce E. Wampold, *Research Design in Counseling*, 3rd ed., Belmont, Calif.: Thomas Higher Education, 2008.

Herz, Martin F., "Some Psychological Lessons from Leaflet Propaganda in World War II," *Public Opinion Quarterly*, Vol. 13, No. 3, Fall 1949, pp. 471–486.

Hoffman, Donna L., and Marek Fodor. "Can You Measure the ROI of Your Social Marketing?" *MIT Sloan Management Review*, Vol. 52, No. 1, Fall 2010, pp. 41–49.

Howell, David C., *Statistical Methods for Psychology*, 5th ed., Pacific Grove, Calif.: Duxbury, 2002.

Hubbard, Douglas W., *How to Measure Anything: Finding the Value of "Intangibles" in Business*, 2nd ed., Hoboken, N.J.: John Wiley and Sons, 2010.

Hughes, Gail, "The Impact of Incorrect Responses to Reverse-Coded Survey Items," *Research in the Schools*, Vol. 16, No. 2, Fall 2009, pp. 76–88.

The Initiatives Group, *Information Environment Assessment Handbook*, version 2.0, Washington, D.C.: Office of the Under Secretary of Defense for Intelligence, 2013.

Issenberg, Sasha, *The Victory Lab*, New York: Crown Publishers, 2012.

———, "How President Obama's Campaign Used Big Data to Rally Voters," *MIT Technology Review Magazine*, January–February 2013.

Jeffery, Mark, *Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know*, Hoboken, N.J.: John Wiley and Sons, 2010.

Job, R. F. Soames, "Effective and Ineffective Use of Fear in Health Promotion Campaigns," *American Journal of Public Health*, Vol. 78, No. 2, February 1988, pp. 163–167.

Johansen, Morgen S., and Mark R. Joslyn, "Political Persuasion During Times of Crisis: The Effects of Education and News Media on Citizens' Factual Information About Iraq," *Journalism and Mass Communication Quarterly*, Vol. 85, No. 3, September 2008, pp. 591–608.

Joint Information Operations Warfare Center, *Joint Information Operations Assessment Framework*, October 1, 2012.

Joint United Nations Programme on HIV/AIDS, *Condom Social Marketing: Selected Case Studies*, Geneva, Switzerland, 2000. As of October 20, 2014:
http://data.unaids.org/publications/IRC-pub02/jc1195-condsocmark_en.pdf

Jonas, Jeff, "Fantasy Analytics," blog post, *Jeff Jonas*, November 9, 2012. As of October 20, 2014:
http://jeffjonas.typepad.com/jeff_jonas/2012/11/fantasy-analytics.html

Keeney, Ralph L., "Developing Objectives and Attributes," in Ward Edwards, Ralph F. Miles, Jr., and Detlof von Winterfeldt, eds., *Advances in Decision Analysis: From Foundations to Applications*, Cambridge, UK: Cambridge University Press, 2007, pp. 104–128.

Kelman, Herbert C., "Processes of Opinion Change," *Public Opinion Quarterly*, Vol. 25, No. 1, 1961, pp. 57–78.

Ketchum Global Research and Analytics, *The Principles of PR Measurement*, undated. As of August 3, 2014:
http://www.ketchum.com/sites/default/files/principles_of_pr_measurement_0.pdf

Khoo, Michael, Joe Pagano, Anne L. Washington, Mimi Recker, Bart Palmer, and Robert A. Donahue, "Using Web Metrics to Analyze Digital Libraries," *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, New York: ACM, 2008, pp. 375–384.

Kilmer, Beau, Nancy Nicosia, Paul Heaton, and Greg Midgette, "Efficacy of Frequent Monitoring with Swift, Certain, and Modest Sanctions for Violations: Insights from South Dakota's 24/7 Sobriety Project," *American Journal of Public Health*, Vol. 103, No. 1, January 2013, pp. 37–43.

Kreuger, Richard A., and Mary Ann Casey, *Focus Groups: A Practical Guide for Applied Research*, Thousand Oaks, Calif.: Sage Publications, 1994.

Lane, Sandra D., "Television Minidramas: Social Marketing and Evaluation in Egypt," *Medical Anthropology Quarterly*, Vol. 11, No. 2, June 1997, pp. 164–182.

LaPiere, Richard T., "Attitudes vs. Actions," *Social Forces*, Vol. 13, No. 2, December 1934, pp. 230–237.

LaRivee, Dave, *Best Practices Guide for Conducting Assessments in Counterinsurgencies*, Washington, D.C.: U.S. Air Force Academy, December 2011.

Leviton, Laura C., "Some Underexamined Aspects of Evaluation Capacity Building," *American Journal of Evaluation*, Vol. 35, No. 1, March 2014, pp. 90–94.

Lindzey, Gardner, and Elliot Aronson, eds., *The Handbook of Social Psychology: Research Methods*, Vol. 2, 2nd ed., Reading, Mass.: Addison-Wesley, 1968.

Lunt, Peter, and Sonia Livingstone, "Rethinking the Focus Group in Media and Communications Research," *Journal of Communication*, Vol. 46, No. 2, June 1996, pp. 79–98.

Lutz, Richard J., "A Functional Approach to Consumer Attitude Research," in Kent Hunt, ed., *Advances in Consumer Research, North America Conference*, Vol. 5, Ann Arbor, Mich.: Association for Consumer Research, 1978, pp. 360–369.

Mackay, Andrew, Steve Tatham, and Lee Rowland, "The Effectiveness of US Military Information Operations in Afghanistan 2001–2010: Why RAND Missed the Point," *IO Sphere*, December 3, 2012.

Maikovich, Andrea Kohn, "A New Understanding of Terrorism Using Cognitive Dissonance Principles," *Journal for the Theory of Social Behaviour*, Vol. 35, No. 4, December 2005, pp. 373–397.

Mares, Marie-Louise, and Zhongdang Pan, "Effects of Sesame Street: A Meta-Analysis of Children's Learning in 15 Countries," *Journal of Applied Developmental Psychology*, Vol. 34, No. 3, May–June 2013, pp. 140–151.

Matel, John, "Hidden Prosperity and the Banana Index in Iraq," blog post, *DipNote*, April 8, 2008. As of October 20, 2014:
http://blogs.state.gov/stories/2008/04/08/hidden-prosperity-and-banana-index

May, Henry, "Making Statistics More Meaningful for Policy Research and Program Evaluation," *American Journal of Evaluation*, Vol. 25, No. 4, 2004, pp. 525–540.

McGuire, William J., "Theoretical Foundations of Campaigns," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, 2nd ed., Newbury Park, Calif.: Sage Publications, 1989, pp. 43–65.

———, "McGuire's Classic Input-Output Framework for Constructing Persuasive Messages," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2012, pp. 133–146.

McManus, John, and Lori Dorfman, "Silent Revolution: How U.S. Newspapers Portray Child Care," *Issue* (Berkeley Media Studies Group), No. 11, January 2002.

Mertens, Donna M., and Amy T. Wilson, *Program Evaluation Theory and Practice: A Comprehensive Guide*, New York: Guilford Press, 2012.

Metzgar, Emily T., *Promoting Japan: One JET at a Time*, CPD Perspectives on Public Diplomacy No. 3, Los Angeles, Calif.: University of Southern California Center on Public Diplomacy, 2012.

Michaelson, David, and Sandra Macleod, "The Application of Best Practices in Public Relations Measurement and Evaluation Systems," *Public Relations Journal*, Vol. 1, No. 1, October 2007, pp. 1–14.

Michie, Susan, Maartje M. van Stralen, and Robert West, "The Behaviour Change Wheel: A New Method for Characterizing and Designing Behaviour Change Interventions," *Implementation Science*, Vol. 6, 2011, article 42.

Military Operations Research Society, *Assessments of Multinational Operations: From Analysis to Doctrine and Policy*, proceedings of the Military Operations Research Society Conference special meeting, MacDill Air Force Base, Tampa, Fla., November 5–8, 2012.

Moehler, Devra C., Jeffrey Conroy-Krutz, and Rosario Aguilar Pariente, "Parties on the Ballot: Visual Cues and Voting Behavior in Uganda," paper presented at the International Communication Association annual conference, Boston, Mass., May 26–30, 2011.

Muñoz, Arturo, *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations 2001–2010*, Santa Monica, Calif.: RAND Corporation, MG-1060-MCIA, 2012. As of August 22, 2014:
http://www.rand.org/pubs/monographs/MG1060.html

———, "Response to 'Why RAND Missed the Point,'" *IO Sphere*, January 15, 2013.

NATO—See North Atlantic Treaty Organization.

Nichols, Wes, "Advertising Analytics 2.0," *Harvard Business Review*, March 2013, pp. 60–68.

Nisbett, Richard E., and Timothy D. Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, Vol. 84, No. 3, March 1977, pp. 231–259.

Norman Lear Center, "Research Study Finds That a Film Can Have a Measurable Impact on Audience Behavior," press release, February 22, 2012. As of October 20, 2014:
http://www.learcenter.org/pdf/FoodInc.pdf

North Atlantic Treaty Organization, *NATO Operations Assessment Handbook*, interim version 1.0, January 29, 2011. Not available to the general public.

North Atlantic Treaty Organization, Joint Analysis and Lessons Learned Centre, *A Framework for the Strategic Planning and Evaluation of Public Diplomacy*, Lisbon, Portugal, 2013. As of October 20, 2014:
http://www.jallc.nato.int/newsmedia/docs/A%20Framework%20for%20the%20Strategic%20Planning%20and%20Evolution%20of%20Public%20Diplomacy.pdf

Office of the Inspector General, U.S. Department of State, *Inspection of the Bureau of International Information Programs*, May 2013. As of August 22, 2014:
http://oig.state.gov/system/files/211193.pdf

Office of the Under Secretary of Defense for Intelligence, *Information Environment Assessment Handbook*, version 2.0, Washington, D.C., 2013.

Olshefsky, Alisa M., Michelle M. Zive, Rosana Scolari, and María Zuñiga, "Promoting HIV Risk Awareness and Testing in Latinos Living on the U.S.-Mexico Border: The Tú No Me Conoces Social Marketing Campaign," *AIDS Education and Prevention*, Vol. 19, No. 5, October 2007, pp. 422–435.

Oral History Project Team, "The Oral History of Evaluation, Part 5: An Interview with Michael Quinn Patton," *American Journal of Evaluation*, Vol. 28, No. 1, March 2007, pp. 102–114.

Osburg, Jan, Christopher Paul, Lisa Saum-Manning, Dan Madden, and Leslie Adrienne Payne, *Assessing Locally Focused Stability Operations*, Santa Monica, Calif.: RAND Corporation, RR-387-A, 2014. As of October 20, 2014:
http://www.rand.org/pubs/research_reports/RR387.html

Paluck, Elizabeth Levy, "Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda," *Journal of Personality and Social Psychology*, Vol. 96, No. 3, March 2009, pp. 574–587.

———, "Is It Better Not to Talk? Group Polarization, Extended Contact, and Perspectives Taking in Eastern Republic of Congo," *Personality and Social Psychology Bulletin*, Vol. 36, No. 9, September 2010, pp. 1170–1185.

Pamment, James, "Towards a Contextualized Interpretation of Public Diplomacy Evaluation," paper presented at the International Studies Association annual convention, San Francisco, Calif., April 3–6, 2013.

———, "Articulating Influence: Toward a Research Agenda for Interpreting the Evaluation of Soft Power, Public Diplomacy, and Nation Brands," *Public Relations Review*, Vol. 40, No. 1, March 2014, pp. 50–59.

Papa, Michael, and Arvind Singhal, "How Entertainment-Education Programs Promote Dialogue in Support of Social Change," paper presented at the 58th annual International Communication Association Conference, Montreal, May 22, 2008.

Patton, Michael Quinn, *Qualitative Research and Evaluation Methods*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2002.

Paul, Christopher, *Strategic Communication: Origins, Concepts, and Current Debates*, Santa Barbara, Calif.: Praeger, 2011.

———, "Foundations for Assessment: The Hierarchy of Evaluation and the Importance of Articulating a Theory of Change," *Small Wars Journal*, Vol. 10, No. 3, 2014.

Paul, Christopher, Harry J. Thie, Elaine Reardon, Deanna Weber Prine, and Laurence Smallman, *Implementing and Evaluating an Innovative Approach to Simulation Training Acquisitions*, Santa Monica, Calif.: RAND Corporation, MG-442-OSD, 2006. As of August 22, 2014:
http://www.rand.org/pubs/monographs/MG442.html

Paul, Christopher, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*, Santa Monica, Calif.: RAND Corporation, RR-809/3-OSD, 2015. As of April 2015:
http://www.rand.org/pubs/research_reports/RR809z3.html

Paul, Christopher, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*, Santa Monica, Calif.: RAND Corporation, RR-809/2-OSD, 2015. As of April 2015: http://www.rand.org/pubs/research_reports/RR809z2.html

Pechmann, Cornelia, and David W. Stewart, "Advertising Repetition: A Critical Review of Wearin and Wearout," *Current Issues and Research in Advertising*, Vol. 11, 1988.

Perry, Robert L., "A Multi-Dimensional Model for PSYOP Measures of Effectiveness," *IO Sphere*, Spring 2008, pp. 9–13.

Petty, Richard E., Pablo Briñol, and Joseph R. Priester, "Mass Media Attitude Change: Implications of the Elaboration Likelihood Model of Persuasion," in Jennings Bryant and Mary Beth Oliver, eds., *Media Effects: Advances in Theory and Research*, 3rd ed., New York: Lawrence Erlbaum Associates, 2009, pp. 125–164.

Petty, Richard E., and John T. Cacioppo, "Forewarning, Cognitive Responding, and Resistance to Persuasion," *Journal of Personality and Social Psychology*, Vol. 35, No. 9, September 1977, pp. 645–655.

———, *Attitudes and Persuasion: Classic and Contemporary Approaches*, Boulder, Colo.: Westview Press, 1996.

Petty, Richard E., and Duane T. Wegener, "Attitude Change: Multiple Roles for Persuasion Variables," in Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds., *The Handbook of Social Psychology*, 4th ed., New York: McGraw-Hill, 1998, pp. 323–390.

Pfeiffer, James, "Condom Social Marketing, Pentecostalism, and Structural Adjustment in Mozambique: A Clash of AIDS Prevention Messages," *Medical Anthropology Quarterly*, Vol. 18, No. 1, March 2004, pp. 77–103.

Phippen, A., L. Sheppard, and S. Furnell, "A Practical Evaluation of Web Analytics," *Internet Research*, Vol. 14, No. 4, 2004, pp. 284–293.

Pornpitakpan, Chanthika, "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence," *Journal of Applied Social Psychology*, Vol. 34, No. 2, February 2004, pp. 243–281.

Power, Gerry, Samia Khatun, and Klara Debeljak, "'Citizen Access to Information': Capturing the Evidence Across Zambia," in Ingrid Volkmer, ed., *The Handbook of Global Media Research*, Chichester, West Sussex, UK: Wiley-Blackwell, 2012, pp. 245–275.

Ramirez, Juan R., and William D. Crano, "Deterrence and Incapacitation: An Interrupted Time Series Analysis of California's Three Strikes Law," *Journal of Applied Social Psychology*, Vol. 33, No. 1, January 2003, pp. 110–144.

Rate, Christopher R., and Dennis M. Murphy, *Can't Count It, Can't Change It: Assessing Influence Operations Effectiveness*, Carlisle Barracks, Pa.: U.S. Army War College, March 14, 2011.

Reingen, Peter H., "Test of a List Procedure for Inducing Compliance with a Request to Donate Money," *Journal of Applied Psychology*, Vol. 67, No. 1, February 1982, pp. 110–118.

Rice, Ronald E., and Dennis R. Foote, "A Systems-Based Evaluation Planning Model for Health Communication Campaigns in Developing Countries," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2013, pp. 69–82.

Ries, Eric, *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, New York: Random House Digital, 2011.

Robertson, Andrew, Eran Fraenkel, Emrys Schoemaker, and Sheldon Himelfarb, *Media in Fragile Environments: The USIP Intended-Outcomes Needs Assessment Methodology*, Washington, D.C.: United States Institute of Peace, April 2011. As of October 20, 2014: http://www.usip.org/publications/media-in-fragile-environments

Rogers, Everett M., *Diffusion of Innovations*, 4th ed., New York: Free Press, 2010.

Rosenthal, Robert, and Ralph L. Rosnow, *Essentials of Behavioral Research*, 3rd ed., New York: McGraw-Hill, 2008.

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

Rowland, Lee A., and Gaby van den Berg, *In Pursuit of a Contextual Diagnostic Approach to Behavior Change Interventions*, London: Behavioural Dynamics Institute, September 2012.

Ruiter, Robert A. C., Charles Abraham, and Gerjo Kok, "Scary Warnings and Rational Precautions: A Review of the Psychology of Fear Appeals," *Psychology and Health*, Vol. 16, No. 6, 2001, pp. 613–630.

Schaefer, Joseph L., and John W. Graham, "Missing Data: Our View of the State of the Art," *Psychological Methods*, Vol. 7, No. 2, June 2002, pp. 147–177.

Scherer, Michael, "Inside the Secret World of the Data Crunchers Who Helped Obama Win," *Time*, November 7, 2012. As of March 3, 2014: http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win

Schneider, Barbara, and Nicole Cheslock, *Measuring Results: Gaining Insight on Behavior Change Strategies and Evaluation Methods from Environmental Education, Museum, Health, and Social Marketing Programs*, San Francisco, Calif.: Coevolution Institute, April 2003.

Schroden, Jonathan, "Why Operations Assessments Fail: It's Not Just the Metrics," *Naval War College Review*, Vol. 64, No. 4, Fall 2011, pp. 89–102.

Schwartz, Shalom H., "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries," in Mark P. Zanna, ed., *Advances in Experimental Social Psychology*, Vol. 25, San Diego, Calif.: Academic Press, 1992, pp. 1–65.

———, "Beyond Individualism/Collectivism: New Dimensions of Values," in Uichol Kim, Harry C. Triandis, Cigdem Kagitcibasi, Sang-Chin Choi, and Gene Yoon, eds., *Individualism and Collectivism: Theory, Method, and Applications*, Thousand Oaks, Calif.: Sage Publications, 1994, pp. 85–119.

Schwarz, Norbert, "Attitude Measurement," in William D. Crano and Radmila Prislin, eds., *Attitudes and Attitude Change*, New York: Psychology Press, 2008, pp. 51–60.

Schwarz, Norbert, and Gerald L. Clore, "Mood, Misattribution, and Judgments of Well-Being: Information and Directive Functions of Affective States," *Journal of Personality and Social Psychology*, Vol. 45, No. 3, September 1983, pp. 513–523.

Sen, Arun, Peter A. Dacin, and Christos Pattichis, "Current Trends in Web Data Analysis," *Communications of the ACM*, Vol. 49, No. 11, November 2006, pp. 85–91.

SenseMaker, homepage, undated. As of October 20, 2014: http://www.sensemaker-suite.com

Shaivitz, Mandy, *How Pro-Social Messages Make Their Way into Entertainment Programming*, Los Angeles, Calif.: Council for Excellence in Government and USC Annenberg Norman Lear Center, 2003. As of October 20, 2014:
http://www.learcenter.org/pdf/MCDReport.pdf

Sides, John, and Lynn Vavreck, "Obama's Not-So-Big Data," *Pacific Standard*, January 21, 2014. As of October 20, 2014:
http://www.psmag.com/navigation/politics-and-law/
obamas-big-data-inconclusive-results-political-campaigns-72687

Skogan, Welsey G., Susan M. Hartnett, Natalie Bump, and Jill Dubois, *Evaluation of CeaseFire-Chicago*, Evanston, Ill.: Institute for Policy Research, Northwestern University, June 2009.

Small, Deborah A., George Loewenstein, and Paul Slovic, "Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims," *Organizational Behavior and Human Decision Processes*, Vol. 102, No. 2, March 2007, pp. 143–153.

Smith, William A., "Social Marketing: An Overview of Approach and Effects," *Injury Prevention*, Vol. 12, Suppl. 1, June 2006, pp. 38–43.

Stanaland, Andrea J. S., and Linda L. Golden, "Consumer Receptivity to Social Marketing Information: The Role of Self-Rated Knowledge and Knowledge Accuracy," *Academy of Marketing Studies Journal*, Vol. 13, No. 2, 2009, pp. 26–32.

Stevens, James Paul, *Applied Multivariate Statistics for the Social Sciences*, 5th ed., New York: Routledge, 2009.

Suedfeld, Peter, "Suedfeld's Integrative Complexity Research," web page, last updated June 7, 2004. As of October 20, 2014:
http://www2.psych.ubc.ca/~psuedfeld/index2.html

Suedfeld, Peter, and Susan Bluck, "Changes in Integrative Complexity Prior to Surprise Attacks," *Journal of Conflict Resolution*, Vol. 32, No. 4, December 1988, pp. 626–635.

Taylor, Maureen, "Methods of Evaluating Media Interventions in Conflict Countries," paper prepared for the workshop "Evaluating Media's Impact in Conflict Countries," Caux, Switzerland, December 13–17, 2010. As of August 22, 2014:
http://www.global.asc.upenn.edu/fileLibrary/PDFs/taylorcaux2.pdf

Thornton, Ronald J., "Likert Scales: An Assessment Application," *IO Sphere*, Summer 2013.

Tufte, Edward, "PowerPoint Is Evil," *Wired*, September 2003. As of August 22, 2014:
http://www.wired.com/wired/archive/11.09/ppt2.html

UK Ministry of Defence, *Assessment*, Joint Doctrine Note 2/12, Shriveham, UK, February 2012. As of September 12, 2012:
http://www.mod.uk/NR/rdonlyres/30D3468C-75B3-4FD9-944D-
975D40E3DCAA/0/20120221jdn212_assessment.pdf

UK Polling Report, *Support for the Iraq War*, online database, undated. As of August 22, 2014:
http://ukpollingreport.co.uk/iraq

Upshur, William P., Jonathan W. Roginski, and David J. Kilcullen, "Recognizing Systems in Afghanistan: Lessons Learned and New Approaches to Operational Assessments," *Prism*, Vol. 3, No. 3, 2012, pp. 87–104.

U.S. Agency for International Development, "Logical Framework Template: Basic," web page, undated. As of October 20, 2014:
http://usaidprojectstarter.org/content/logical-framework-template-basic

U.S. Department of Defense Directive 3600.01, *Information Operations (IO)*, Washington, D.C., May 2, 2013.

U.S. Joint Chiefs of Staff, *Joint Operation Planning*, Joint Publication 5-0, Washington, D.C., August 11, 2011a.

———, *Joint Operations*, Joint Publication 3-0, Washington, D.C., August 11, 2011b.

———, *Commander's Handbook for Assessment Planning and Execution*, version 1.0, Suffolk, Va.: Joint and Coalition Warfighting, September 9, 2011c.

———, *Joint Intelligence*, Joint Publication 2-0, Washington, D.C., October 22, 2013.

U.S. Marine Corps, *Assessment: MAGTF Staff Training Program (MSTP)*, MSTP Pamphlet 6-9, Quantico, Va.: Marine Air-Ground Task Force Staff Training Program, October 25, 2007.

———, *Marine Corps Operating Concept for Information Operations*, Washington, D.C., February 4, 2013.

U.S. Pacific Fleet, "Pacific Partnership 2012 to 2013: Assessment Transition Brief," briefing, undated.

Valente, Thomas W., *Network Models of the Diffusion of Innovations*, New York: Hampton Press, 1995.

———, *Evaluating Health Promotion Programs*, Oxford, UK: Oxford University Press, 2002.

———, *Social Networks and Health: Models, Methods, and Applications*, New York: Oxford University Press, 2010.

———, "Network Interventions," *Science*, Vol. 337, No. 6090, July 2012, pp. 49–53.

Valente, Thomas W., and Patchareeya P. Kwan, "Evaluating Communication Campaign," in Ronald E. Rice and Charles K. Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2012, pp. 83–98.

Valente, Thomas W., Patricia Paredes, and Patricia R. Pope, "Matching the Message to the Process: The Relative Ordering of Knowledge, Attitudes, and Practices in Behavior Change Research," *Human Communication Research*, Vol. 24, No. 3, March 1998, pp. 366–385.

Valente, Thomas W., and Walter P. Saba, "Reproductive Health Is in Your Hands: The National Media Campaign in Bolivia," *SIECUS Report*, Vol. 25, No. 2, December 1996–January 1997, pp. 10–13.

———, "Mass Media and Interpersonal Influence in a Reproductive Health Communication Campaign in Bolivia," *Communication Research*, Vol. 25, No. 1, February 1998, pp. 96–124.

———, "Campaign Recognition and Interpersonal Communication as Factors in Contraceptive Use in Bolivia," *Journal of Health Communication*, Vol. 6, No. 4, 2001, pp. 303–322.

VanSlyke, Judy Turk, and Linda H. Scanlan, *Evolution of Public Relations: Case Studies of Nations in Transition*, Gainesville, Fla.: Institute for Public Relations, 1999.

———, *Evolution of Public Relations: Case Studies of Nations in Transition*, 2nd ed., Gainesville, Fla.: Institute for Public Relations, 2004.

———, *Evolution of Public Relations: Case Studies of Nations in Transition*, 3rd ed., Gainesville, Fla.: Institute for Public Relations, 2008.

Venkatraman, Meera P., Deborah Marlino, Frank R. Kardes, and Kimberly B. Sklar, "The Interactive Effects of Message Appeal and Individual Differences on Information Processing and Persuasion," *Psychology and Marketing*, Vol. 7, No. 2, Summer 1990, pp. 85–96.

Voice of America, "The Largest U.S. International Broadcaster," factsheet, Washington D.C., March 2013. As of October 20, 2014:
http://www.bbg.gov/wp-content/media/2013/05/VOA-Fact-Sheet-March-2013.pdf

Wainer, Howard, "How to Display Data Badly," *American Statistician*, Vol. 38, No. 2, May 1984, pp. 137–147.

Wandersman, Abraham, "Getting To Outcomes: An Evaluation Capacity Building Example of Rationale, Science, and Practice," *American Journal of Evaluation*, Vol. 35, No. 1, March 2014a, pp. 100–106.

———, "Moving Forward with the Science and Practice of Evaluation Capacity Building (ECB): The Why, How, What, and Outcomes of ECB," *American Journal of Evaluation*, Vol. 35, No. 1, March 2014b, pp. 87–89.

Wechsler, William F., Deputy Assistant Secretary of Defense for Special Operations and Combating Terrorism, "Information Operations (IO) 2nd Quarter Reporting Requirement," memorandum, undated.

West, Stephen G., Jeremy C. Biesanz, and Steven C. Pitts, "Causal Inference and Generalization in Field Settings: Experimental and Quasi-Experimental Designs," in Harry T. Reis and Charles M. Judd, eds., *Handbook of Research Methods in Social and Personality Psychology*, New York: Cambridge University Press, 2000, pp. 40–84.

Wong, Carolyn, *How Will the e-Explosion Affect How We Do Research? Phase 1: The E-DEL+I Proof-of-Concept Exercise*, Santa Monica, Calif.: RAND Corporation, DB-399-RC, 2003. As of October 20, 2014:
http://www.rand.org/pubs/documented_briefings/DB399.html

World Bank Group, *Worldwide Governance Indicators*, online database, undated. As of October 20, 2014:
http://info.worldbank.org/governance/wgi/index.aspx

Yarbrough, Donald B., Lyn M. Shulha, Rodney K. Hopson, and Flora A. Caruthers, T*he Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2011.

Yeats, Jessica M., and Walter L. Perry, *Review of the Regional Center Enterprise Measures of Effectiveness Plan*, unpublished RAND research, 2011.

Yin, Robert K., *Case Study Research: Design and Methods*, 5th ed., Thousand Oaks, Calif.: Sage Publications, 2014.

Index

A

Abbott, Susan, 90n38, 99n57
Abraham, Charles, 337n, 338n65, n.63–64
Abrams, Lindsay, 338n66
A/B testing, 157, 177
academia
    informative results from, 8
    interviews with SMEs in, 7
academic evaluation research
    best practices from, 263
    design of, 142
    insights from successful research in, 2
    represented in study, 7
access, as barrier in survey research, 235
accidental sampling, 284, 285
accountability
    in assessment, 49–51
    clarity in demands for, 269
    congressional mandate for, 20–25, 269
    in data collection, 55–59
    DoD reporting to meet requirement for, 269
    environment of, 45
    as motive for assessment, 12, 13
    tied to measures, 124
    and time horizons, 51–55
achievable objectives, 74, 76–77
Acosta, Joie, 29, 52n39, 96n50, 198n24, 222n136
ACSOR/D3 systems, 59, 283, 289, 290
actionable metrics, 203
activities
    connecting objectives and, 32–33
    inform and influence, 4
    in logic models, 89, 95–96
    performing vs. assessing, 48
    pilot tests of, 134
Aday, Sean, 156
adversary, measuring attitudes toward, 207
advertising analytics, 303–304

advocacy evaluation (bellwether method), 159, 183
advocacy for assessment
    by broader DoD IIP community, 267–268
    lack of, 68
affect (in MINDSPACE), 339
Affholter, Dennis, 124
Afghan Assessments Group, 292
Afghanistan
    challenges of assessment in, 65–66
    documentation of activities/programs in, 195
    ISAF psychological operations in, 299–300
    media and public perceptions in, 306–307
    problems with polling programs in, 289
    quasi-experimental designs in, 156
    surveys in, 56–59, 246–247
    systems approach to communities in, 336–337
Africa
    conflict resolution media programs in, 152
    marketing trend in, 63
after-action review, 8–9, 97
aggregate data, 220–221, 258
aggregation
    of assessments, 224–225
    of IIP and campaign assessments, 25–26
agreement bias, 245
Air Land Sea Application Center manual, 3
Ajzen, Icek, 205n56, 329n34
Akers, Ronald L., 328n31
Akert, Robin M., 327n30
Albany Associates, 188
allocation, 304. *See also* resource allocation
Altai Consulting, 56, 59, 161, 244, 290, 306–307
American Educational Research Association, 260
American Psychological Association, 260
analysis of variance (ANOVA), 227

Y

Z

To achieve key national security objectives, the U.S. government and the U.S. Department of Defense (DoD) must communicate effectively and credibly with a broad range of foreign audiences. DoD spends more than $250 million per year on inform, influence, and persuade (IIP) efforts, but how effective (and cost-effective) are they? How well do they support military objectives? Could some of them be improved? If so, how? It can be difficult to measure changes in audience behavior and attitudes, and it can take a great deal of time for DoD IIP efforts to have an impact. DoD has struggled with assessing the progress and effectiveness of its IIP efforts and in presenting the results of these assessments to stakeholders and decisionmakers. To address these challenges, a RAND study compiled examples of strong assessment practices across sectors, including defense, marketing, public relations, and academia, distilling and synthesizing insights and advice for the assessment of DoD IIP efforts and programs. These insights and attendant best practices will be useful to personnel who plan and assess DoD IIP efforts and those who make decisions based on assessments, particularly those in DoD and Congress who are responsible for setting national defense priorities and allocating the necessary resources. In addition to identifying where and why efforts have been successful, assessment can help detect imminent program failure early on, saving precious time and resources. An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*, offers a quick-reference guide to the best practices presented here for personnel responsible for planning, executing, and assessing DoD IIP efforts.

RAND | NATIONAL DEFENSE RESEARCH INSTITUTE

www.rand.org

$30.00