Assessing Impact across Policy Domains
A Workshop at the Center for Strategic & International Studies

PRINCIPAL AUTHORS:
MAREN LEED
JENNIFER TAYLOR

CONTRIBUTING AUTHOR:
ZACK FELLMAN

DECEMBER 2010
About CSIS

In an era of ever-changing global opportunities and challenges, the Center for Strategic and International Studies (CSIS) provides strategic insights and practical policy solutions to decisionmakers. CSIS conducts research and analysis and develops policy initiatives that look into the future and anticipate change.

Founded by David M. Abshire and Admiral Arleigh Burke at the height of the Cold War, CSIS was dedicated to the simple but urgent goal of finding ways for America to survive as a nation and prosper as a people. Since 1962, CSIS has grown to become one of the world’s preeminent public policy institutions.

Today, CSIS is a bipartisan, nonprofit organization headquartered in Washington, DC. More than 220 full-time staff and a large network of affiliated scholars focus their expertise on defense and security; on the world’s regions and the unique challenges inherent to them; and on the issues that know no boundary in an increasingly connected world.

Former U.S. senator Sam Nunn became chairman of the CSIS Board of Trustees in 1999, and John J. Hamre has led CSIS as its president and chief executive officer since 2000.

CSIS does not take specific policy positions; accordingly, all views expressed herein should be understood to be solely those of the author(s).

© 2010 by the Center for Strategic and International Studies. All rights reserved.
On December 2, 2010, the Center for Strategic and International Studies hosted a workshop attended by approximately forty experts in impact assessment and performance measurement. By intention this group represented a diverse range of policy focus—from defense and security issues to international development, intelligence, education, and health care. It included representatives from governmental organizations, academia, and the private sector.

The day’s discussion had three primary objectives:

- To generate discussion between communities that address similar issues from different vantage points;
- To provide a forum for enhancing understanding shared measurement challenges and state-of-the-art approaches to addressing them; and
- To develop a network of expertise in support of future interactions.

The workshop was held on a non-attribution basis, so the discussion below summarizes themes without reference to specific participants. The day was organized around five questions to spark discussion, though the resulting conversation highlighted how thoroughly they are interrelated. The five questions were as follows:

- What are we trying to measure—effectiveness or efficiency?
- How are good indicators identified?
- Among relevant indicators, how do we identify the most important?
- How do we move from measuring outputs to outcomes?
- How can we best develop effective feedback loops?

Two dominant themes came out during the course of the day. One was the real yet often overlooked distinction between performance monitoring and program evaluation. Participants spent a fair amount of time discussing the distinctions between evaluating how well a given program or policy is being implemented, which requires one set of indicators, and assessing its impact, which requires a different set of information. The task of monitoring is principally concerned with efficiency, or the relationship of inputs to outputs, whereas evaluation focuses on effectiveness indicators to gauge outcomes. The need to be clear about the ultimate purpose and use of the resulting analysis—to determine whether a given program or project was actually having the intended effect versus understanding whether it was being well-run—was emphasized repeatedly as a crucial first step to ensuring a useful product.

The second theme was the challenge of increasing the impact of the impact assessments themselves. There was discussion throughout the day about how to bridge the gap between high-quality analysis and getting that analysis acted upon by decision-makers. One of the workshop’s main objectives of the day was to cross-pollinate communities that normally do not engage with one another as a means of identifying practices that might be easily shared or transferred. Somewhat surprisingly, the discussion revealed a relatively high degree of common practice, but
also common challenges and frustrations that continue to plague assessment practitioners. These include the difficulty of ensuring the integration of analysis into decisions, but also of finding the most effective way to manage disparate stakeholder interests, ensuring sufficient financial and organizational support for impact assessment in the first place, and conducting the most rigorous analysis in the face of extremely difficult or complex field conditions.

Summary of Discussion

The day began with a discussion of Minard’s famous graph depicting the size of the French Grande Armée during Napoleon’s march on Moscow in 1812. The discussant noticed that despite relatively accurate data about casualties, temperature, terrain, and campaign progress, the ultimate determination of whether the loss of life and treasure was ultimately “worth it” was highly dependent on the outcome. If Moscow had fallen, history likely would have judged the campaign differently. Further, even if Napoleon somehow had access to lots of data before and during his march, it still would not have answered the question of whether he should go to war. More broadly, the point was raised that data does not answer policy questions, though it can inform them.

As the day progressed, the discussion moved past the types of cost-benefit calculations like those relating to Napoleon’s campaign and focused more directly on impact assessment, which is aimed at identifying causal relationships. A few key features of impact assessment were raised repeatedly. First, all stakeholders—policy makers, relevant donors or funders, implementers, and assessors—have varying interests, ways of understanding the problem and potential solution sets, and frequently different objectives. Second, regardless of initial intentions, there is always the possibility that politics rather than analysis will determine a program’s fate. Third, in ideal circumstances, the strategic objectives of a project or policy should drive the indicators and methodology of determining impact or efficiency. However, these circumstances rarely exist. Far more frequently, desired effects are not identified up front or are changed ex post facto, measurements are begun in the middle of implementation, and adequate baselines are lacking. Project design is all too often done without evidence-based analysis as a foundation, rather than beginning with objectives and outcomes that could a significant effect on how desired aims are pursued.

There was some variability across the group in the degree to which their assessments focused on effectiveness versus efficiency. Evaluations in some fields (e.g., education) are almost exclusively focused on effectiveness, though it was noted that perhaps efficiency will play a more prominent role in the future. In other areas efficiency receives substantially greater attention. Overall, however, it was clear that evaluating effectiveness is the principal concern, and that determining whether the desired impact is being achieved in the most efficient manner is, if raised at all, typically a secondary issue. While participants recognized that efficiency may take on greater import, the larger concern expressed about fiscal pressure was that it might result in decreased support for impact assessment overall.
One participant described the clarity of a well-defined outcome as “liberating.” Increased gross domestic product is easier to assess, it was acknowledged, than is the contribution a given program might have to the absence of a terrorist act.

Frequent engagement with stakeholders was also a common theme. The most successful impact assessments work with them to go from broad goals and objectives to determine the appropriate set of indicators. What is the objective of the indicator? To tell a story—how effective is this project? Compare to other projects—are there established benchmarks? To provide a cost analysis—are results being provided to seniors? Stakeholder education is frequently in order—to assist with feasibility of measurements and determine explicit, vice implicit, objectives. These explicit objectives will determine success of the project—e.g., while building relationships with the host nation is frequently an important impact; it cannot be determined to be an important aspect of project success after the fact. From a more procedural standpoint, attendees also asserted that involving their targeted consumers in the designing and implementation of a study could be a strong way to ensure that consumers have a vested interest in seeing (and using) the results.

One participant put forth a set of criteria to help identify good assessment indicators. He asserted that the best indicators meet the following four criteria, and if conditions change so that they no longer do, a change may be in order:

1. They are objective, with transparent assumptions.
2. They are frequent, which increases an evaluator’s ability to find effects.
3. They are collected over a long period of time to allow for trend analysis. Measuring trends over time provide a great value. An insistence on the precise mapping should not lead to change of measures too frequently.
4. They are not burdensome to collect, which increases the likelihood the data will actually be available. This participant acknowledged that this is not always possible, but for some outcomes (e.g., violence, wages paid for fixed services) it is.

Another participant noted that these criteria related more directly to outputs than to outcomes, and were more appropriately suited to performance measurement purposes than for impact assessment. This led to a broader discussion of the importance of conducting randomly controlled trials (RCTs), which ultimately most participants agreed are the gold standard of impact assessment. They vastly decrease the likelihood of turning to output indicators, and they obviate the need for a counter-factual. That said, there are a wide range of instances in which an RCT may be extremely difficult to implement. There were repeated mentions of circumstances in which RCTs may not be a viable option — for example, in providing a program believed to be beneficial to one group but denying access to another. Participants discussed a number of alternative approaches to address such circumstances that included phasing in the program over different time periods, varying the design of the intervention, etc. Ultimately, the group agreed that getting as close to an RCT design as possible is desirable, but that there may be some instances in which that distance is substantial.
The group noted that finding good indicators is most difficult when measuring the intangible, for example democracy or openness in the private sector. Data limitations are also always an issue. Frequently data is impractical to collect, and projects may evolve or demonstrate bias towards positive impact. Ultimately, the group reiterated, it is important to recognize that there is no inherent hierarchy between quantitative and qualitative approaches. Instead, it is highly dependent on the circumstances, and good qualitative data will support a better impact assessment than bad quasi-experimental quantitative data. While quantitative studies are in theory replicable, testable, and scientifically rigorous, participants noted that high-quality qualitative studies ought to be able to meet these same standards.

The day concluded by returning to the subject of making assessments more meaningful. An informal poll revealed that only a handful had recently read an impact assessment report produced by their organizations. The implication was drawn that if few impact assessors are reading these reports, probably even fewer outside of the community are reading them. The challenges of getting to the proper audience are many, and include lack of time or lack of organizational understanding. Policymakers are pressed for time, and may lack the appropriate technical expertise to fully understand the implications of an assessment. (On the front end, they may also lack the requisite knowledge to inform good assessment design.)

The group also discussed the importance of understanding the “customers’” worldview in order to more effectively communicate results in that context. They also noted that the nature of the assessors’ relationship to the decision makers can have a significant effect on how to ensure that assessments are utilized. The obstacles that face an assessment organization that is within a larger institution are different than those of an external one that has been called in by one or more stakeholders. At least one organization has been successful in building a culture around the importance of evaluating impact, and project managers would rather shrink some aspects of programming than reduce evaluation funds. This success was attributed to the fact that the organization is relatively new, and its creation was in part motivated by a strong desire for greater accountability. Creating this culture where it does not already exist within an existing organization would be a much larger challenge, the group acknowledged. That said, others reported that they have found success in building an incentive structure to encourage assessments, principally by grounding their efforts in theories of organizational change.

Participants highlighted the importance of targeting the audience who can effect change, which in turn requires a good understanding of organizational structure and centers of influence. Organizational culture and attitudes about the importance of impact assessment more specifically, can also play a role in determining how best to conduct and/or market an assessment. Assessors must be cognizant of perverse incentives that might prevent action, such as evaluations that demonstrate that a project is not having its intended effect.

There was a brief discussion about the need to incentivize projects that produce vice those that underperform, but also of the practical difficulties if disaggregating poor implementation and poor design. A failure to produce the desired effects may be because the design is not being
correctly put into practice, but may be because the design is ill-conceived. Sometimes it takes putting programs into practice to identify both sets of issues, and it may take multiple iterations before it is clear which problem is actually being manifested. Funders thus need to find the right balance between allowing for implementation problems to be identified and resolved and ensuring that scarce resources are directed toward the most efficacious programs.

Ultimately, the workshop reinforced participants’ understanding of the complexities involved with conducting high quality impact assessment but also the range of tools available to enable better insight in support of policy decisions. Those tools may not provide perfect clarity, but can yield additional information. As one participant repeatedly noted, having some empirical basis for action—even if imperfect—is better than none at all. Thus assessors can benefit from continued sharing of approaches both to solving evaluation problems but also the challenge of making those evaluations a more meaningful component of policy choices. CSIS looks forward to continuing to support such efforts.