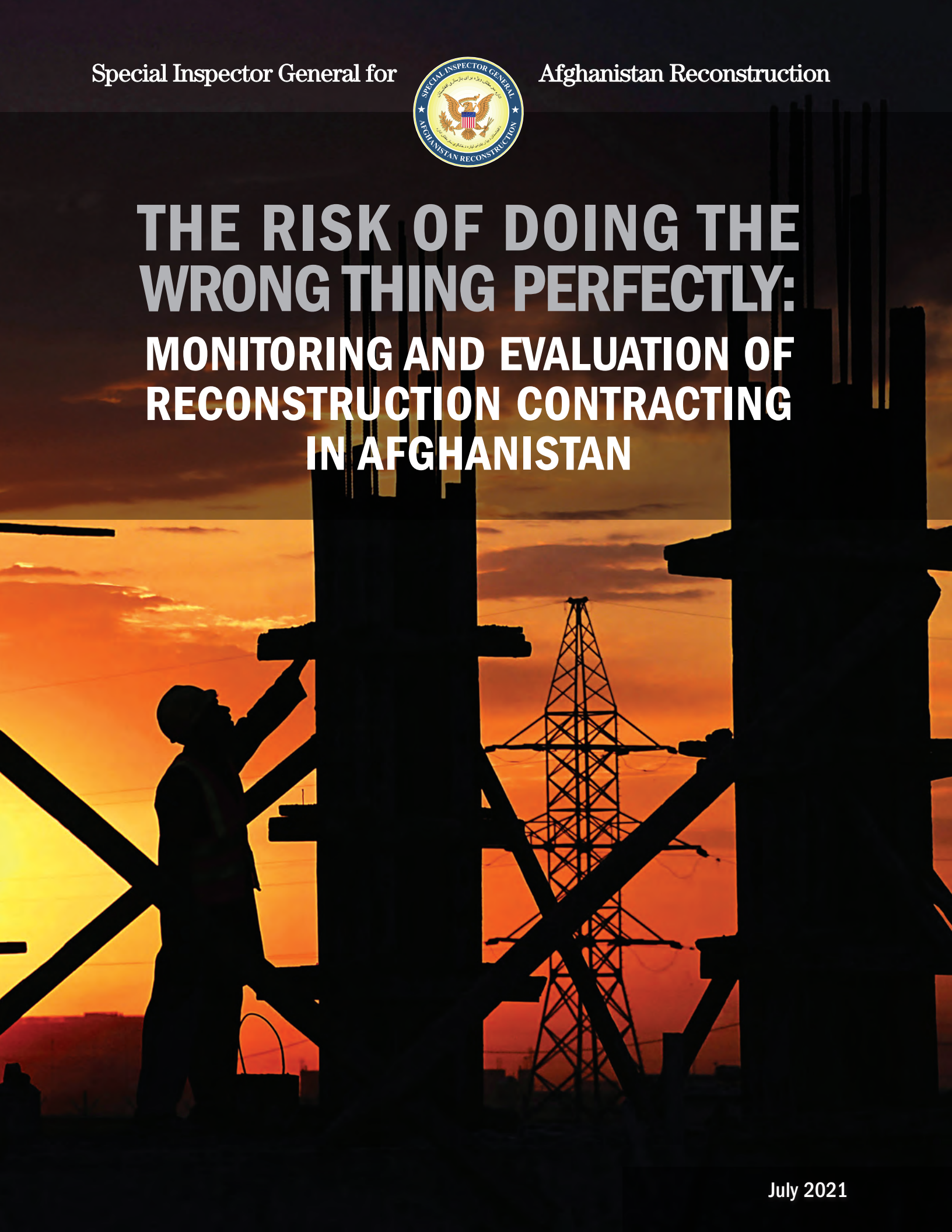




THE RISK OF DOING THE WRONG THING PERFECTLY: MONITORING AND EVALUATION OF RECONSTRUCTION CONTRACTING IN AFGHANISTAN





Cover photo credit:

Afghan laborers work at a construction site during sunset in Mazar-e Sharif on April 22, 2014. (AFP photo by Farshad Usyan)



Special Inspector General
for Afghanistan Reconstruction

The Risk of Doing the Wrong Thing Perfectly: Monitoring and Evaluation of Reconstruction Contracting in Afghanistan is the 10th lessons learned report to be issued by the Special Inspector General for Afghanistan Reconstruction. The report examines how reconstruction contracts have been monitored and evaluated in Afghanistan since 2001. In addition to exploring how agencies applied monitoring and evaluation (M&E) to individual contracts, the report also analyzes how multiple contracts—for example, portfolios of contracts in sectors like health or education—were assessed to determine their net effect and overall impact. For the purposes of the report, “contracting” includes other mechanisms such as grants and cooperative agreements, through which U.S. funding was—as through contracts—provided directly to third parties in support of Afghanistan reconstruction.

As U.S. efforts to reconstruct Afghanistan evolved, contractors became a prominent feature of the reconstruction landscape, substantially augmenting U.S. government capacity. However, heavy reliance on contracts also presented significant challenges. Widely documented by SIGAR and others, contingency contracting is notoriously vulnerable to waste, fraud, and abuse. In unpredictable and chaotic environments such as Afghanistan, poor oversight or improper implementation can threaten relationships with local communities, endanger the lives of U.S. and Afghan government personnel and civilians, and undermine strategic goals. Consequently, proper M&E—broadly defined as the processes used to determine the performance and effectiveness of programs and projects—is vital.

We identify findings and lessons from the last two decades of U.S. efforts to apply M&E to reconstruction contracting in Afghanistan. We found that agencies placed far more emphasis on tracking program activities and outputs than on assessing outcomes and impacts. Although M&E processes generally evolved and improved over time in response to widespread concerns about corruption and a lack of reconstruction progress, M&E systems that existed on paper were not always put into practice. In many cases, M&E became a “work-maker” that generated an excess of data, not all of which were relevant to the broader goals contracts were supposed to advance. Moreover, although agencies developed processes to change course on projects that were not working, these processes were not always fully used. This shortcoming undermined the basic purpose of M&E.

Most importantly, we found that M&E systems were not always used in ways that allowed agencies to periodically revalidate the theories and assumptions underlying programs. The absence of periodic reality checks created the risk of doing the wrong thing perfectly: A project that met contracted deliverables and performance-indicator targets would be considered “successful,” whether or not it had achieved or contributed

to broader, more important goals. On a more positive note, we found that agencies have relatively robust M&E—or M&E-like—policies in place. Consequently, key aspects of these M&E systems have the potential to improve both programmatic and strategic outcomes, if they are fully embraced and implemented.

Recent developments—including the decision to withdraw U.S. troops from Afghanistan—make this report quite timely. The decision to withdraw presents a unique opportunity to reset U.S. aspirations for reconstruction based on what can be reasonably accomplished, and to rethink how progress should be measured.

In addition to its timeliness, the report has further-reaching application in two important ways. First, it is almost axiomatic that the United States periodically becomes involved in large-scale reconstruction efforts. Should the United States find itself involved in another—even several years or decades from now—the findings, lessons, and recommendations presented here may prove useful. Second, even in the present, many of the report’s observations are likely generalizable to other contexts. The fundamental strategic logic that stability can be achieved by providing security sector assistance, promoting economic growth and accountable governance, and fostering societal inclusivity is hardly unique. Indeed, this core intervention logic remains remarkably constant across numerous other countries and regions characterized by instability. As long as the United States continues its effort to promote stability across the globe, the question of how to properly apply M&E to the innumerable contracts, grants, and cooperative agreements that are intended to advance that objective will remain highly relevant.

Lessons learned reports such as this one comply with SIGAR’s legislative mandate to provide recommendations to promote economy, efficiency, and effectiveness of U.S.-funded reconstruction programs and operations; to prevent and detect waste, fraud, and abuse; and to inform the Congress and the Secretaries of State and Defense about reconstruction-related problems and the need for corrective action.

Congress created SIGAR as an independent agency focused solely on the Afghanistan mission and devoted exclusively to reconstruction issues. Unlike most inspectors general, which have jurisdiction only over the programs and operations of their respective departments or agencies, SIGAR has jurisdiction over all programs and operations supported with U.S. reconstruction dollars, regardless of the agency involved. Because SIGAR has the authority to look across the entire reconstruction effort, it is uniquely positioned to identify and address whole-of-government lessons.

Our lessons learned reports synthesize not only the body of work and expertise of SIGAR, but also that of other oversight agencies, government entities, current and former officials with on-the-ground experience, academic institutions, and independent scholars. The reports document what the U.S. government sought to accomplish, assess what it achieved, and evaluate the degree to which these efforts helped the United States reach its reconstruction goals in Afghanistan. They also provide recommendations to address the challenges stakeholders face in ensuring effective and sustainable reconstruction efforts in Afghanistan as well as in future contingency operations.

SIGAR's Lessons Learned Program comprises subject matter experts with considerable experience working and living in Afghanistan, aided by a team of seasoned research analysts. I want to express my deepest appreciation to the team members who produced this report. I thank the report team: Daniel Fisher, project lead; Paul Fishstein, supervisory research analyst; Matt Bentrrott, senior research analyst; Anna Andriychuk, Alyssa Goodman, Lauren Helinski, and Patrick O'Malley, research analysts; and Joshua Herman, Stephanie Pillion, and Holly Ratcliffe, student trainees. I also thank Nikolai Condee-Padunov, program manager; Tracy Content, editor; Jason Davis and Vong Lim, visual information specialists; and Joseph Windrem, Lessons Learned Program director.

In producing its reports, the program also uses the significant skills and experience found in SIGAR's Audits, Investigations, and Research and Analysis directorates. I thank all of the individuals who provided their time and effort to contribute to this report.

In addition, I am grateful to the many U.S. government officials at the Department of Defense, Department of State, and the U.S. Agency for International Development, as well as contractor and implementing partner staff who provided valuable insights and feedback. This report is truly a collaborative effort intended not only to identify problems, but also to learn from them and propose reasonable solutions to improve future reconstruction efforts.

I believe lessons learned reports such as this will be a key legacy of SIGAR. Through these reports, we hope to reach a diverse audience in the legislative and executive branches, at the strategic and programmatic levels, both in Washington and in the field. Using our unique interagency mandate, we intend to do everything we can to make sure the lessons from the most ambitious reconstruction effort in U.S. history are both identified and fully applied.

A handwritten signature in black ink, appearing to read 'John F. Sopko', with a long, sweeping horizontal stroke extending to the right.

John F. Sopko,

Special Inspector General for Afghanistan Reconstruction

Arlington, Virginia

	<p>EXECUTIVE SUMMARY VIII</p> <p>CHAPTER 1 1 INTRODUCTION</p> <p>Scope 5</p> <p>A Brief History of Reconstruction Contracting: Over Time, U.S. Government Outsourcing Became More Prevalent 10</p> <p>Contracting and M&E in Afghanistan: All the Usual Challenges, Plus a Wartime Environment 16</p>
	<p>CHAPTER 2 27 USAID</p> <p>Introduction 27</p> <p>The Challenge: Selecting What to Measure, and How 30</p> <p>Putting M&E to Use 52</p> <p>Chapter Summary and Key Findings 74</p>
	<p>CHAPTER 3 77 STATE</p> <p>Introduction 77</p> <p>State Reconstruction Programming in Afghanistan 79</p> <p>Finding the Right Metrics to Assess Progress: Measuring the Measurable 83</p> <p>Learning and Acting on Information 95</p> <p>Chapter Summary and Key Findings 109</p>
	<p>CHAPTER 4 111 DOD</p> <p>DOD's Approach to M&E Has Evolved Over Time 111</p> <p>Reconstructing the ANDSF: Measuring Operational Effectiveness Proved Difficult 114</p> <p>Counterinsurgency through Economic Development: Was the Underlying Theory Valid? 124</p> <p>Operation Assessment: DOD's Overarching M&E Analogue 144</p> <p>Chapter Summary and Key Findings 154</p>



CHAPTER 5	157
INTERNATIONAL EXPERIENCE AND PRACTICES	
The United States Was Not Alone in Struggling with M&E	157
Many Key Themes from the International Experience Reinforce Those from the United States	167
Key Findings	169



CHAPTER 6	171
CONCLUSION	
Findings	174
Lessons	175
Recommendations	180



APPENDICES & ENDNOTES	193
Appendix A: Methodology	193
Appendix B: Legislation, Authorities, and Policies Governing Contracting and M&E	196
Appendix C: USAID Country Development Cooperation Strategy Evidence Grades	220
Appendix D: Acronyms	224
Acknowledgements	226
Endnotes	227

FIGURES

Figure 1: How M&E Practices Create the Risk of Doing the Wrong Thing Perfectly	X
Figure 2: Contracting Instruments	5
Figure 3: Contracting Personnel	12
Figure 4: How M&E Practices Create the Risk of Doing the Wrong Thing Perfectly	17
Figure 5: USAID Spending on Contracts, Grants, and Cooperative Agreements by Sector	28
Figure 6: Country Development Cooperation Strategy Evidence Grading Scale	40
Figure 7: Evolution of Money as a Weapon System–Afghanistan: Key Changes to CERP M&E Requirements after May 2009	130
Figure 8: Example of a Resolute Support District Stability Assessment	153
Figure 9: At USAID, the Country-level Results Framework Helps Organize M&E	201
Figure 10: Visualizing USAID’s Results Framework	202
Figure 11: The Performance Management Plan in Action: An Example from USAID’s 2011–2015 Performance Management Plan	203
Figure 12: Illustrating the Functions of an Activity Monitoring, Evaluation, and Learning Plan	205
Figure 13: M&E in USAID’s Program Cycle: Key Documents and Processes	207
Figure 14: State Program Design and Performance Management Process	210

TABLES

Table 1: A Pattern of Poor M&E: Examples from Prior SIGAR Work	3
Table 2: Country Development Cooperation Strategy Evidence Grades (Selected)	42
Table 3: Illustrative Contradictions or Logical Flaws in USAID Evaluations	62
Table 4: Major State Department Funds	79
Table 5: Critical Analysis of Performance Indicators for State’s Legal Aid through Legal Education Program, as Presented in the Program’s FY 2018 PMP	86
Table 6: Illustrative Contradictions or Logical Flaws in State Evaluations	102
Table 7: DOD Reconstruction Programming in Afghanistan	113
Table 8: Strengths and Weaknesses of DOD’s Impact Evaluation of CERP	136
Table 9: Select Examples of Legislation Impacting Contracting and M&E	197
Table 10: What M&E Means to Each Agency	199
Table 11: Standard Foreign Assistance Indicator Categories	212
Table 12: State’s Indicator Selection Criteria	213

CALLOUT AND HIGHLIGHT BOXES

Callout: M&E vs. Oversight	2
Highlight: What Is M&E?	7
Callout: The Federal Acquisition Regulation	11
Callout: What Is Corruption?	13
Highlight: The Concept of Inherently Governmental Functions Reflects the Risks and Benefits of Contracting	15
Callout: Standard Foreign Assistance Indicators	19
Highlight: Technological Advancements Created New Tools for M&E, But Potential Benefits Were Sometimes Stymied	22
Callout: Defining Key USAID Terms: Projects and Activities	30
Highlight: Public Opinion Polling Was a Key Source of Information, but Data Were Sometimes Questionable	32
Callout: Defining Key USAID Terms: Performance Management Plan	34
Highlight: To What Extent is USAID’s Strategy Grounded in Evidence?	40

MONITORING AND EVALUATION OF RECONSTRUCTION CONTRACTING

Callout:	SIGAR Has Audited Agency Efforts in the Health and Education Sectors	47
Highlight:	The Measuring Impacts of Stabilization Initiatives Program: A Unique Attempt at Assessing Impact	48
Callout:	Plausible Contribution May Be a Reasonable Substitute for Impact Assessment	50
Callout:	Levels of Multitiered Monitoring	53
Highlight:	Third-party Monitoring in the Health Services Sector Illustrates the Strengths and Limitations of Outsourcing Data Collection	54
Highlight:	Monitoring Whether Development Programming Benefitted the Taliban: USAID Sends Mixed Messages	57
Callout:	Defining Key USAID Terms: Results Framework	60
Highlight:	What Makes for a Good Evaluation in an Environment such as Afghanistan?	64
Highlight:	Are There Limitations to the Utility of USAID's Strategic Review Processes?	67
Callout:	Burn Rate: The Amount of Money Spent Became an Indicator of Progress	68
Callout:	Defining Key USAID Terms: Activity, Monitoring, Evaluation, and Learning Plan	69
Highlight:	USAID's Power Transmission Expansion and Connectivity Project: A Poster Child for Failed M&E	71
Callout:	On-budget and Off-budget Assistance	71
Callout:	Anecdotes Can Supplement Raw Indicators, but Have Their Own Limitations	85
Callout:	Data Collection Mechanisms, Data Interpretation and Use Remain Key Challenges	89
Callout:	Weighing the Relative Importance of U.S. Programs and Examining Macro Trends Can Help Determine Whether Political Aims Are Being Achieved	93
Highlight:	Lack of Performance Metrics Hindered M&E of Rule of Law Assistance	94
Highlight:	Evaluations of State's Reconstruction Programming Suffered from Analytical Flaws	98
Callout:	Shortcomings in CSSP's Performance Management Plan	101
Callout:	Differing Bureaucratic Timelines May Prevent State from Making Necessary Changes to Programs	107
Callout:	The Military's Perspectives on ANDSF Capability Have Sometimes Been Called into Question	120
Callout:	The Afghanistan Compact	122
Callout:	The Legacy and Afghanistan Source Operations Management Programs: Lack of Performance Metrics Made It Difficult to Determine Return on Investment	123
Highlight:	What Was CERP's Purpose?	127
Callout:	CERP Impact Assessments Had to Grapple with Data Quality Issues	134
Highlight:	The Task Force for Business and Stability Operations: Some Impact Assessments Do Not Actually Examine Impact	140
Callout:	The Afghan First Initiative: Sometimes Metrics Were Straightforward but Difficult to Reliably Track	145
Highlight:	District Control: A Case Study of the Utility of a Key Campaign-level Metric	151
Callout:	Defining Key World Bank Terms	158
Callout:	A Metareview Commissioned by the German Government was a Useful and Unique Approach to M&E	165
Callout:	Randomized Controlled Trial and Quasi-experimental Design	182
Callout:	Types of Evaluation	206
Callout:	Defining Key State Department Entities	208
Highlight:	Road-testing State's Performance Management Tool Kit: The Importance of Continually Reassessing Programmatic Assumptions	211

EXECUTIVE SUMMARY

The United States has appropriated more than \$144 billion for Afghanistan reconstruction since 2001. A large proportion of this spending funded contracts, grants, and cooperative agreements through which contractors and nongovernmental organizations implemented programs and projects that aimed to achieve reconstruction objectives. These private organizations provided an astounding array of services—everything from dam and road construction to developing the Afghan private sector and promoting economic growth, to advising Afghan government ministries and maintaining helicopters for the Afghan Air Force. Contracting for these services allowed the U.S. government to augment its capacity and to implement programs and projects deemed vital to the reconstruction effort. It also created serious challenges.

Particularly in a contingency environment like Afghanistan, contracting is vulnerable to waste, fraud, and abuse. It is also susceptible to inappropriate and ineffective programming that at times undermined—or at least did not meaningfully advance—the reconstruction effort. There is more than a decade’s worth of literature documenting various kinds of contracting excesses in Afghanistan, and feedback from those reports has led to increasing emphasis on contract monitoring and evaluation (M&E). But the quality of actionable information generated by M&E has not always been commensurate with either the complexity of agencies’ M&E systems, or with the laborious data collection they required. SIGAR’s body of work, which has touched every major reconstruction sector—health, education, rule of law, women’s rights, infrastructure, security assistance, and others—paints a picture of U.S. agencies struggling to effectively measure outcomes and impacts while sometimes relying on shaky data to make claims of success. This report builds on SIGAR’s prior work, pursues new lines of inquiry, and examines ways in which the United States can improve its approach to M&E.

Broadly speaking, M&E is the effort to objectively determine whether a contractor or implementing partner fulfilled the terms of a program’s contract, whether or not a program accomplished its intended goals, and what useful lessons can be drawn. Rigorous M&E is intended to enable U.S. agencies to measure performance and effectiveness, learn from success and failure, make course corrections in ongoing programs, inform future programming, and hold contractors or implementing partners accountable for their work. To be effective, M&E must be accurate, timely, informative, and actionable—and both mission leadership and contract oversight personnel must be willing and able to act on the information gathered. In a contingency environment, M&E practices must be flexible enough to adapt to rapidly shifting environments without sacrificing effectiveness or accountability.

The need for effective M&E is reinforced by the decades-long trend toward “downsizing,” “reinventing” or outsourcing governmental functions. This trend significantly increased the number and size of government contracts overall, and thus the need for contract

accountability. At the same time, staffing levels and the technical capacity in government agencies which might be expected to do that work—most notably, for purposes of this report, the Department of State, the Department of Defense (DOD) and the U.S. Agency for International Development (USAID)—did not keep pace with the ever-increasing number and value of contracts. Every major report on contracting has noted the chronic shortages of government employees to do basic contract oversight. Illustrating the extent of the problem, at one point in 2011, USAID would have had to send nearly its entire overseas workforce to work only in Afghanistan in order to meet the U.S. government’s average ratio of dollars to contracting officers.

Yet, despite its challenges, contracting is essential: There is no practical way that U.S. agencies can do what contractors do. As the U.S. government continues to rely extensively on contractors for reconstruction work, improving M&E and ensuring that officials act on its findings will be a critical need. This is especially true in Afghanistan and in other countries and regions where the U.S. seeks to counter instability. Such contexts, where assistance programs are often subject to considerable political scrutiny and frequently carried out under conditions un conducive to direct U.S. government oversight, tend to exacerbate all the normal frictions of program management.

In particular, they present unique obstacles to accurately assessing performance and effectiveness. A key finding of this report is that, as implemented in Afghanistan, M&E created a very high risk of doing the wrong thing perfectly: A project that met contracted deliverables and performance-indicator targets would generally be considered “successful,” whether or not it had achieved broader, more important goals. Figure 1 on the next page presents illustrative examples of this risk. These examples are discussed in detail later in this report.

The current situation in Afghanistan presents heightened challenges. Recent developments suggest that 2021 may actually be the year that fulfills perennial predictions that some critical inflection point is imminent. These developments include:

- a February 2020 agreement between the United States and the Taliban for a withdrawal of troops, in exchange for the Taliban’s commitment to prevent the use of Afghan soil by groups or individuals that threaten the United States and its allies;
- throughout 2020 and into 2021, the continuing, albeit very slow-moving, talks between the Afghan government and the Taliban on the possibility of intra-Afghan peace;
- increasing levels of violence.

In the meantime, President Joseph R. Biden has set a deadline for the withdrawal of all U.S. troops by September 11, 2021. The decision to withdraw creates a new and different set of uncertainties in the relationship between the United States, the Afghan government, and the Taliban, and raises the question of how U.S. funds may be spent in Afghanistan moving forward. Although Afghanistan’s social and economic needs remain significant—and have increased in complexity due to the devastation caused by the spread of COVID-19—the extent of future U.S. support and what form it will take

FIGURE 1

HOW M&E PRACTICES CREATE THE RISK OF DOING THE WRONG THING PERFECTLY

	Why the Risk is Created	Example	Why M&E Does Not Address the Flaw	
Permutations of Risk	Weak strategic evidence base	According to USAID's current Country Development Cooperation Strategy for Afghanistan, all USAID programs and projects "are designed to contribute to the stability, and ... increased self-reliance of Afghanistan because they address key drivers of conflict." However, the strategy does not present convincing evidence that this assertion is plausible.	M&E is concentrated at the individual program and project level and rarely reexamines how, why, or if a program or project might materially contribute to strategic goals. If the strategic evidence base is weak, M&E may conclude a program or project is successful regardless of whether it contributes to these broader goals.	Risk of Doing the Wrong Thing Perfectly
	Flawed performance indicators	USAID's Incentives Driving Economic Alternatives for the North, East and West project aimed to reduce opium-poppy cultivation. However, during implementation the program dropped indicators that were directly related to reducing opium-poppy cultivation.	M&E cannot track progress towards programmatic objectives if performance indicators are unrelated to the objectives.	
	Potentially shaky programmatic assumptions	The Commander's Emergency Response Program was implemented as part of an effort to defeat the Taliban insurgency and promote stability. However, the assumption that economic and social development could be deliberately used for these purposes has been contested. Meanwhile, CERP metrics were conflated with the act of implementation.	M&E generally relies on metrics that are input or output oriented. Such metrics are therefore of little use in stress-testing programmatic assumptions.	
	The problem a program is trying to solve is too entrenched	The Justice Training Transition Program sought to increase Afghans' confidence in the formal justice sector. However, the program's success was largely contingent on significant, chronic factors outside the program's control, such as addressing widespread corruption and catalyzing political will to enact meaningful reforms.	M&E errs on the side of positivity, searching for evidence of progress rather than stagnation or regression even in situations where obstacles to progress are substantial. Consequently, it may determine a program or project is more successful than it actually is.	
	Unreliable or anecdotal data	Agencies have made claims of success in Afghanistan's education sector using exaggerated school-enrollment data. Actual school-attendance rates are likely far lower.	When M&E relies on misleading data, it provides inaccurate information about program effectiveness.	

Source: USAID, "Afghanistan Country Development Cooperation Strategy FY 2019–2023," September, 2018, p. 26; USAID Office of Inspector General, "Audit of USAID/Afghanistan's Incentives Driving Economic Alternatives for the North, East, and West Program," Audit Report No. F-306-12-004-P June 29, 2012, pp. 2, 5; Wilton Park, "Report on Wilton Park Conference 1022: Winning 'Hearts and Minds' in Afghanistan: Assessing the Effectiveness of Development Aid in COIN Operations, Thursday 11–Sunday 14 March 2010," April 1, 2010, p. 1; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. ix–x, 92–93, 98–99, 100; Daniel Egel, Charles P. Ries, Ben Connable, et al., "Investing in the Fight: Assessing the Use of the Commander's Emergency Response Program in Afghanistan," RAND Corporation, 2016, p. 213; SIGAR, *Commander's Emergency Response Program: DOD Has Not Determined the Full Extent to Which Its Program and Projects, Totalling \$1.5 Billion in Obligations, Achieved Their Objectives and Goals in Afghanistan from Fiscal Years 2009 through 2013*, SIGAR-18-45-AR, April 2018, pp. i, 25; SIGAR, *Primary and Secondary Education in Afghanistan: Comprehensive Assessments Needed to Determine the Progress and Effectiveness of Over \$759 Million in DOD, State, and USAID Programs*, SIGAR-16-32-AR, April 26, 2016, p. ii; Conor Foley and Orsolya Szekeley, "External Final Evaluation of Afghanistan Justice Training Transition Program (JTTP)," March 2016, pp. 8, 13, 24–26, 40; Conor Foley and Katerina Stolyarenko, "External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014)," November 2014, p. 17; The Asia Foundation, "Afghanistan in 2015: A Survey of the Afghan People," 2015, p. 39; The Asia Foundation, "Afghanistan in 2014: A Survey of the Afghan People," 2014, p. 38; The Asia Foundation, "Afghanistan in 2013: A Survey of the Afghan People," 2013, pp. 34–35; The Asia Foundation, "Afghanistan in 2012: A Survey of the Afghan People," 2012, p. 42; SIGAR, *Quarterly Report to the United States Congress*, July 30, 2017, p. 182; SIGAR analysis.

are still far from clear. Notwithstanding these uncertainties, the U.S. troop withdrawal presents a unique opportunity to reset aspirations for reconstruction based on what can be reasonably accomplished, and to rethink how progress should be measured. For these reasons, this report is quite timely.

As U.S. agencies continue to draw down their presence in Afghanistan, the U.S. government's ability to oversee its reconstruction activities may become even more difficult than in recent years. In the short term, mitigating the risk of waste may require

more creativity in using remote monitoring and other techniques for activity-level monitoring, and more rigor in strategic-level monitoring. Conversely, in the long term, an intra-Afghan peace agreement could prove a significant boon to M&E efforts if it materially reduces violence. For example, improved security could increase third-party access to project sites—including, potentially, those presently contested or controlled by the Taliban. In either case, M&E will have to be responsive to shifting conditions on the ground.

While the road ahead may be very different from the one traveled so far, it is clear that U.S. involvement in Afghanistan's reconstruction is far from ended, and that contracting will continue to play a central role. Given this reality, understanding the successes, failures, and inherent limitations of M&E in those contracts, and how it can be improved, is a necessary starting point for U.S. agencies attempting to help rebuild a country devastated by 40 years of war.

This report aims to contribute to that understanding. It has six chapters accompanied by a technical appendix:

- **Chapter 1** explores the history and practice of reconstruction-related contracting in the post-9/11 warfighting environment, and the strengths and weaknesses of contracting itself. Chapter 1 further details the M&E processes that evolved in Afghanistan. It concludes by placing contracting and M&E within the broader context of recent developments, including peace negotiations and the COVID-19 pandemic.
- **Chapter 2** explores how USAID's detailed, highly elaborated M&E processes were put to use in Afghanistan in an effort to understand what projects were actually accomplishing. In doing so, it describes the challenges USAID faced in applying M&E in a complex environment where pressure to demonstrate progress was immense, and where confounding variables frequently made impact difficult to determine.
- **Chapter 3** discusses the use of M&E at State. As with the chapter on USAID, it is structured around two key aspects of M&E: finding the right metrics to assess progress, and acting on information provided by M&E. It is organized principally around core findings. The chapter shows how State's M&E efforts evolved over time, gradually becoming more systematized. In the end, even with increased sophistication, State's M&E systems suffered from many of the same shortcomings as USAID's. Overall, tracking program activities and measuring outputs was generally prioritized over outcome and impact assessment.
- **Chapter 4** explores how DOD implemented M&E in Afghanistan. DOD used a variety of assessment, M&E and oversight practices that were not always directly comparable to USAID's and State's processes. The chapter examines the challenges DOD faced in monitoring and evaluating its programming through case studies in the two key areas in which DOD spending was concentrated: reconstructing the Afghan National Defense and Security Forces, and pursuing counterinsurgency through economic development. It focuses on high-dollar-value programs within these areas

and discusses measurement systems used to assess overall progress in building the capacity of the security forces, which involved a wide variety of contracts. The last part of the chapter briefly explores DOD’s overarching M&E-like process: operation assessment.

- **Chapter 5** reviews international experiences and practices of M&E. The chapter examines several different entities, including the World Bank, the UN, the German Federal Ministry for Economic Cooperation and Development, and the UK Department for International Development. It shows that M&E challenges were not unique to the United States, and concludes with a list of suggested best practices derived from the international experience.
- **Chapter 6** presents findings, lessons and recommendations based on the previous five chapters. Some of these apply broadly to challenges faced by all three agencies; others are addressed to USAID, State or DOD individually. We provide a summary of the report’s findings, lessons, and recommendations below.
- **Appendix B** presents important technical context for contracting and M&E. The appendix provides a comprehensive baseline against which the actual implementation of M&E is assessed in chapters 2, 3, and 4. It describes the role of the Federal Acquisition Regulation and examines how key legislative actions have impacted M&E and contracting—both within and outside of the context of Afghanistan reconstruction. It then details current M&E policies at DOD, State, and USAID, showing how M&E is supposed to be integrated into the broader, cyclical processes through which each agency designs programming and uses it to advance strategy.

KEY FINDINGS

SIGAR identified 11 key findings:

1. The assumption was that work completed well would lead to good results. However, it is possible to do the wrong thing perfectly. As implemented, even if M&E systems were able to determine that work was completed well, those systems did not always determine whether good work was actually contributing to achievement of strategic U.S. goals.
2. Although there were some exceptions, DOD, State, and USAID generally placed more emphasis on tracking inputs and outputs than on assessing impact.
3. DOD, State and USAID now have relatively robust M&E—or M&E-like—systems in place. But in practice, M&E was often treated more like a compliance exercise than a genuine opportunity to learn and adapt programming and strategy.
4. DOD, State, and USAID began to place more emphasis on deliberate and methodical monitoring and evaluation during the 2009 to 2012 surge period. The trend during and shortly after this period was towards increasingly institutionalized and complex M&E, particularly at State and DOD.

5. Pressure to demonstrate that gains were being made discouraged candid assessment of progress toward outcomes and impacts, and often led to selective or overly positive reporting.
6. Although agencies developed processes to weed out programs and projects that were unlikely to succeed and to change course on those that were not working, these processes were not always fully used, undermining the fundamental purpose of M&E.
7. Frequently, program- and project-level metrics reflected what was easy to measure rather than what was most relevant. Discrete, quantitatively-oriented metrics had a tendency to oversimplify what constituted “success.”
8. Confusion about what data mattered, in conjunction with a compulsion to overmeasure, led to a tendency to collect data with little actual assessment value.
9. Reporting and administrative requirements can occupy valuable staff time that could otherwise be spent on programs or projects themselves. Contracting officer’s representatives were often overworked, in part because of M&E and oversight requirements, an issue exacerbated by personnel shortages and short rotations.
10. Some of the most useful, but also most challenging, aspects of M&E systems—such as policies requiring the development of robust, evidence-based models that connect programs and projects to higher-order strategic outcomes and that are periodically reassessed—exist on paper, but are not always implemented in meaningful ways.
11. Key aspects of existing M&E policies have the potential to be very useful for improving both programmatic and strategic efficacy in Afghanistan and elsewhere, if they are fully embraced and implemented.

LESSONS

In addition to these findings, SIGAR identified 10 lessons:

1. **Measuring outcomes and impacts is critical.**

Too often, DOD, State, and USAID failed to measure programs and projects against the ultimate outcomes and impacts those programs and projects sought to achieve. With M&E relegated to input and output measurement, it was often difficult to understand what was and was not working. With numerous confounding variables, complex environments like Afghanistan can present significant obstacles to this kind of assessment. Nevertheless, making a reasonable attempt to determine outcomes and impacts is crucial.

2. **M&E policies and practices are less likely to be effective if they are not meaningfully implemented.**

Each agency examined for this report has strong M&E systems in place. But the mere existence of policy and procedure on paper does not guarantee that these practices are implemented fully. Whether because of strategic or programmatic inertia, or because meaningfully meeting every M&E requirement is practically

infeasible, programs and projects can continue even if they are not necessarily achieving outcomes and impacts.

3. Continually stress-testing the theories and assumptions guiding programming is crucial.

In complex environments, causal processes of change are usually not well understood. Yet assumptions about those causal processes are often used to justify programming. In such environments, many projects are likely to be implemented because they are *believed*—rather than *proven*—to be effective. Especially in these environments, M&E must be used to continually stress-test such assumptions.

4. Performance metrics are only useful if they are relevant.

Determining what to measure is a function of what programs and projects aim to accomplish and how they intend to accomplish it. If metrics are unrelated to objectives, they are not useful for assessing effectiveness.

5. Pressure to demonstrate progress can undermine the utility of M&E.

External pressure, whether political or interagency, to demonstrate immediate and tangible results can shift the incentive structure surrounding M&E. If the perception is that there is a requirement (implicit or explicit) to demonstrate progress, M&E is both less likely to accept evidence of a lack thereof and more likely to be biased towards positive data. This can result in unsupported claims of success.

6. Measurability alone should not determine which metrics are prioritized and what is defined as success.

Measurability can determine key choices surrounding M&E. For example, it is relatively easy—or, if not easy, simply more tangible—to estimate or quantify the number of children enrolled in school, changes in life expectancy, or reductions in maternal mortality. In that regard, measurability and quantifiability help explain why improvements in health and education appear so frequently as evidence of progress in Afghanistan. However, these metrics have well-documented faults. Moreover, even allowing for substantial gains in these indicators, it is unclear whether they are accurate proxies of progress towards the political objectives motivating U.S. support to the health and education sectors. In short, precision can be a façade and quantifiability can obscure important nuance or qualification.

Because measurability is not always a good proxy for efficacy, measurability alone should not determine which metrics are prioritized and what is defined as a success.

7. Anecdotes and success stories can be useful, but only if they accurately represent the broader picture.

Anecdotes can serve the very useful purpose of adding nuance or humanity to issues that would otherwise be captured only in numbers. However, anecdotes that are not representative of the wider situation can also create misperceptions: Success stories that suffer from selective presentation of facts or hyperbole may not be successes at all.

8. In cases where programs or projects aim to achieve political outcomes, tracking performance against those outcomes is critical.

Effectiveness must be evaluated against relevant outcomes. Some projects may be implemented primarily because they carry significant political symbolism. For example, the ultimate intent of a project may be to signal U.S. commitment to the host government. Beyond political symbolism, many reconstruction programs and projects are intended to achieve political objectives that are more explicit—in particular, “stability.” In either case, the actual effectiveness of projects can only be measured against these political outcomes.

9. High levels of spending outpaced the number of contract oversight personnel.

An enduring feature of contingency environments is that when spending increases, oversight generally does not keep pace. Virtually every report written on reconstruction contracting has pointed out that contract oversight personnel are overworked, overburdened, or too few in number.

10. Poor M&E can result in waste.

Poor M&E can reduce program effectiveness, imperil the achievement of mission objectives, and result in waste. In its most recent calculation, SIGAR estimated that total reconstruction losses were approximately \$19 billion. At least two programmatic categories—stabilization and counternarcotics—appear to have failed entirely. M&E that candidly examines programmatic prospects of success can help mitigate the risk that taxpayer funds are spent to no purpose.

RECOMMENDATIONS

Based on the analysis in this report, SIGAR makes 14 recommendations:

Recommendations to DOD, State, and USAID

- 1. When reporting claimed successes to external stakeholders such as the Congress and the public, the Secretary of Defense, Secretary of State, and Administrator of USAID should report only those claims that can be supported by multiple data points, and acknowledge any important context, qualifications, and data limitations.**

Selective presentation of facts can misrepresent the situation on the ground. A large body of SIGAR work that includes audits of U.S. efforts to develop Afghanistan's health, education, rule of law, and power sectors, as well as an audit of Promote (USAID's largest single investment to advance women globally), point to unclear outcomes or impact, inadequate assessment of effects, or the use of incomplete or faulty data. Taken as a whole, this body of work—as well as the new research presented in this report—raises serious questions about many claims of success advanced by agencies. Moving forward, agencies should be candid and transparent in how they report such claims. Specifically, they should support them with multiple data points while openly acknowledging any important context, qualifications, and data limitations.

- 2. To maximize the effectiveness of future reconstruction programming, the Secretary of Defense, Secretary of State, and the Administrator of USAID should determine the 10 most successful and 10 least successful reconstruction programs or projects of their respective department or agency. The determination should be based on the extent to which the programs or projects contributed to the accomplishment of U.S. strategic goals, and should include a detailed explanation of how the programs and projects were evaluated and selected. Its findings should be incorporated into future planning, including planning for reconstruction-like programs or projects in other countries, if applicable.**

Although the present situation in Afghanistan is characterized by uncertainty, some aspects of the future mission are more predictable than others. In particular, resources for reconstruction are likely to continue to decline, signaling difficult decisions ahead. The critical question will be where to best invest available funds for the highest possible strategic return. Perhaps more than ever, it is important for agencies to articulate how they evaluate and perceive their past successes and failures.

For these reasons, with this report, SIGAR is renewing its unanswered March 2013 request of the Secretary of Defense, the Secretary of State, and the Administrator of USAID to determine the 10 most successful and 10 least successful reconstruction programs or projects of their respective department or agency.

- 3. The Secretary of Defense, the Secretary of State, and the Administrator of USAID should more regularly conduct impact evaluations to assess the effects of contracted reconstruction and other foreign assistance programs, including security sector assistance.**

The most credible way to test whether contracted programs and projects are effective is through impact evaluation. In an impact evaluation, a program or project's underlying theory of change is submitted to rigorous testing. Good impact evaluations leave little room for overly optimistic findings because they force an answer to the most fundamental M&E question: Is the program or project achieving desired effects? DOD, State, and USAID should conduct impact evaluations with greater regularity.

- 4. The Secretary of State and the Administrator of USAID should assess whether minimizing or modifying administrative requirements for compliance and M&E would result in more time available to assess program effectiveness.**

Overburdened oversight staff emerges as one of the most consistent themes from both the existing literature and SIGAR's work for this report. Overly burdensome administrative requirements related to both compliance and M&E may exacerbate the problem. These issues raise the important question of whether more time should be devoted to understanding effectiveness and periodically revalidating larger programmatic assumptions than on meeting requirements that may not answer more fundamental questions. The question applies not only to how the time and bandwidth of individuals might be better used, but also to how USAID and State determine organizational priorities and choose how to allocate limited resources.

One way to allocate more time to answering such questions would be to hold spending constant; another might be to increase staffing levels, or reduce the administrative burden for existing staff. Assuming it is not possible to increase staffing, especially in light of the COVID-19 pandemic and the ongoing drawdown of U.S. personnel in Afghanistan, State and USAID should comprehensively assess whether there are tradeoffs between meeting every administrative requirement and performing quality M&E. Such an assessment could help begin a conversation around whether certain requirements could be eliminated or modified to free up more time for analyzing program effectiveness.

DOD's AM&E policy for security cooperation is still relatively new and the full effects of M&E formalization are not yet clear. Additionally, DOD's most articulated M&E requirements prior to the issuance of the policy applied primarily to the Commander's Emergency Response Program, which is no longer in wide use in Afghanistan. We therefore make this recommendation only to State and USAID.

Recommendations for USAID

- 5. The Administrator of USAID should conduct a systematic review of the available evidence relating to core assumptions of the current Country Development Cooperation Strategy for Afghanistan.**

Chapter 201 of USAID’s Automated Directives System requires each mission to integrate evidence into strategic planning. Such evidence is supposed to provide rigorous justification for USAID’s selected development approach in a country. However, key assumptions presented in USAID’s Country Development Cooperation Strategy for Afghanistan are not always supported by sufficient evidence. Problems with the strategy’s evidence base are detailed in Chapter 2.

Where the theories and assumptions underlying a strategy or program are shaky, M&E as practiced does a poor job of identifying and underscoring that core problem. This creates the risk of doing the wrong thing perfectly: A program supporting a flawed strategy may appear to be successful even if it does not ultimately contribute to strategic objectives. This is why a systematic review of the evidence base underlying USAID’s current strategy is vital. If such a review shows that the core logic of the strategy is flawed, a different set of interventions or even a new strategy may be required. The review should focus on available academic literature relating to the current strategy’s core development hypothesis that “all [development objectives] are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address key drivers of conflict.”¹

- 6. The Administrator of USAID should ensure that future portfolio reviews and midcourse stocktaking exercises for the current Afghanistan Country Development Cooperation Strategy maximize the potential to adapt programs in light of new information.**

SIGAR recommends that USAID conform closely to both the letter and spirit of its current M&E policies to ensure that strategic reviews fully embrace bad news without equivocation or obfuscation. This creates the opportunity to learn from failures and improve future programming.

- 7. The Administrator of USAID should ensure that project evaluations are properly adhering to USAID policy to provide “specific, concise” conclusions that can be “readily understood,” and to “objectively evaluate [a] strategy, project, or activity.”**

According to USAID policy, evaluations “should be readily understood and should identify key points clearly, distinctly, and succinctly.” Additionally, findings and conclusions should be “specific” and “concise,” and evaluations should “objectively evaluate the strategy, project, or activity.”² However, there are instances of evaluations which have seemingly contradictory findings and conclusions,

or which minimize significant qualifications. Those criteria should be more consistently followed.

Recommendations for State

- 8. The Secretary of State should ensure that regular progress reviews of contracted programs adequately and consistently “test theory of change hypotheses,” as encouraged by State’s M&E policy.**

To foster analysis and learning, State M&E guidance encourages “incorporating regular reviews of progress,” which can help program teams “test theory of change hypotheses by filling knowledge gaps and resolving uncertainties in the hypotheses with new research or syntheses of existing analyses.” The reviews are also intended to “inform future strategic documents . . . and budgetary decisions.”³ Because strategic alignment and evidence-based theories of change are articulated as critical aspects of program design, testing and periodically reassessing both alignment and underlying theory is an important component of periodic reviews. In practice, however, reviews do not always fulfill these objectives. The Secretary of State should ensure that periodic reviews better adhere to existing guidance.

- 9. The Secretary of State should ensure that evaluations of reconstruction programs and projects in Afghanistan and other contingency environments properly comply with standards of “usefulness” and “methodological rigor” articulated in State’s M&E policy.**

State M&E policy provides clear standards for evaluations. Those standards include concepts of “usefulness”—namely, that “evaluations should help the Department improve its management practices and procedures as well as its ongoing activities”—and “methodological rigor”—meaning evaluations “should be ‘evidence based.’”⁴ However, adherence to the written policy is uneven, and there are cases in which evaluations are less useful or rigorous than required by policy. Moving forward, the Secretary of State should ensure that evaluations properly adhere to these important standards.

Recommendations for DOD

- 10. The Secretary of Defense should ensure that the effectiveness of future contracted security capacity-building programs is adequately assessed, using appropriate metrics.**

Despite DOD’s extensive use of contractors, DOD did not always assess the effectiveness of contracted programs intended to increase Afghan government security institution capacity. This is an avoidable issue which DOD should address moving forward. This recommendation echoes prior SIGAR calls for DOD to ensure that security capacity-building contracts are achieving the results for which they aim.

- 11. The Secretary of Defense should ensure that a requirement is in place to assess the impact of all major reconstruction programs it implements in the future, including those that are not typically part of DOD’s core mission, such as those intended to benefit host-nation civilians.**

Applying counterinsurgency doctrine, DOD implemented a vast array of reconstruction and development projects that did not fall under its core mission. But despite the importance of such efforts, DOD struggled to implement M&E systems that could have measured whether desired effects were being achieved. Years later, the impact of some of these efforts on the overall conflict remains unassessed or uncertain. This represents a lost opportunity. In the future, DOD should ensure that a requirement is in place to assess impacts—even when projects fall outside of its core mission.

- 12. To reduce the prevalence of overoptimism in determining progress towards objectives, the Secretary of Defense should ensure that campaign-level operation assessments follow existing doctrine and incorporate “red teams” that challenge organizational biases, provide opposing points of view, and constructively critique proposed plans for accomplishing the mission.**

Operation assessment is similar to M&E in nature and goals, but it focuses on evaluating the progress of military operations or campaigns as opposed to specific projects or programs. When conducted at the campaign level, it is the U.S. military’s highest-order M&E-like mechanism for assessing progress.

A key flaw of operation assessment was a tendency towards overoptimism. One potential way to mitigate this tendency is to integrate red teams into operation assessments to ensure equal consideration of “glass half-full” and “glass half-empty” views. While not foolproof, this kind of devil’s advocacy may contribute to more objective interpretation of data, and curb overly optimistic assessment conclusions. Current joint military guidance recommends the routine employment of red teams. The Secretary of Defense should take measures to ensure this guidance is adequately implemented at the campaign level in Afghanistan and in other contexts in which the United States seeks to promote stability.

Matters for Consideration by the Congress

- 13. The Congress may wish to consider requiring the Secretary of Defense, Secretary of State, and Administrator of USAID to submit a report that describes in specific terms how oversight, monitoring and evaluation of contracted programs will continue in the event of a further drawdown of U.S. military and civilian personnel in Afghanistan. Should the report not satisfactorily explain how agencies plan to ensure proper oversight and**

M&E, the Congress should consider the efficacy of continuing spending at current levels.

As detailed in both this chapter and in other parts of this report, insufficient staffing can contribute to problems in adequately overseeing, monitoring, and evaluating reconstruction contracts. Most of the funds appropriated for Afghanistan reconstruction since 2002 could have been spent more wisely and cost-effectively—and achieved better and longer-lasting outcomes—with better oversight and M&E. But the drawdown of military and civilian personnel creates the risk of further lowering the ratio of program management staff to dollars spent, which could exacerbate perennial oversight and M&E challenges.

To help mitigate this risk, the Congress should consider requiring agencies to submit a report detailing in specific terms how they plan to maintain adequate oversight, monitoring, and evaluation of reconstruction contracts as the U.S. personnel drawdown continues. Should the Congress be dissatisfied with agency plans, it should consider the efficacy of continuing assistance at present levels.

- 14. The Congress may wish to consider appropriating funds to DOD, State, and USAID specifically for impact evaluation of both Afghanistan reconstruction programs and more broadly for U.S. foreign assistance, including security sector assistance. An alternative would be to mandate that a certain proportion of funds appropriated to these agencies be used for impact evaluation.**

Pursuant to Recommendation 3, the Congress may wish to consider appropriating funds to DOD, State, and USAID that could be used to evaluate the effects of future reconstruction assistance. The Congress may also wish to consider providing funds to conduct impact evaluations of other U.S. foreign assistance programs, including security sector assistance, implemented around the world.



The National Defense Authorization Act for FY 2008 (P.L. 110-181) established the Special Inspector General for Afghanistan Reconstruction (SIGAR).

SIGAR's oversight mission, as defined by the legislation, is to provide for the independent and objective

- conduct and supervision of audits and investigations relating to the programs and operations funded with amounts appropriated or otherwise made available for the reconstruction of Afghanistan.
- leadership and coordination of, and recommendations on, policies designed to promote economy, efficiency, and effectiveness in the administration of the programs and operations, and to prevent and detect waste, fraud, and abuse in such programs and operations.
- means of keeping the Secretary of State and the Secretary of Defense fully and currently informed about problems and deficiencies relating to the administration of such programs and operation and the necessity for and progress on corrective action.

Afghanistan reconstruction includes any major contract, grant, agreement, or other funding mechanism entered into by any department or agency of the U.S. government that involves the use of amounts appropriated or otherwise made available for the reconstruction of Afghanistan.

Source: P.L. 110-181, "National Defense Authorization Act for FY 2008," 1/28/2008.

A silhouette of a construction worker wearing a hard hat, standing on a building site. The background is a dramatic sunset with orange and yellow hues. The worker is positioned in the upper left quadrant, looking towards the right. The building's steel framework is visible, with several vertical rebar rods extending upwards. The overall scene is dark, with the worker and the building's structure appearing as black shapes against the bright, glowing sky.

WASTE, FRAUD, OR ABUSE MAY BE REPORTED TO SIGAR'S HOTLINE

By phone: Afghanistan
Cell: 0700107300
DSN: 318-237-3912 ext. 7303
All voicemail is in Dari, Pashto, and English.

By phone: United States
Toll-free: 866-329-8893
DSN: 312-664-0378
All voicemail is in English and answered during business hours.

By fax: 703-601-4065
By email: sigar.hotline@mail.mil
By web submission: www.sigar.mil/investigations/hotline/report-fraud.aspx

SIGAR-21-41-LL

**SIGAR
SPECIAL INSPECTOR GENERAL
FOR AFGHANISTAN RECONSTRUCTION**

2530 Crystal Drive
Arlington, VA 22202

www.sigar.mil