



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





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 Bentrutt, 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



EXECUTIVE SUMMARY VIII

CHAPTER 1 INTRODUCTION 1

Scope	5
A Brief History of Reconstruction Contracting: Over Time, U.S. Government Outsourcing Became More Prevalent	10
Contracting and M&E in Afghanistan: All the Usual Challenges, Plus a Wartime Environment	16



CHAPTER 2 USAID 27

Introduction	27
The Challenge: Selecting What to Measure, and How	30
Putting M&E to Use	52
Chapter Summary and Key Findings	74



CHAPTER 3 STATE 77

Introduction	77
State Reconstruction Programming in Afghanistan	79
Finding the Right Metrics to Assess Progress: Measuring the Measurable	83
Learning and Acting on Information	95
Chapter Summary and Key Findings	109



CHAPTER 4 DOD 111

DOD's Approach to M&E Has Evolved Over Time	111
Reconstructing the ANDSF: Measuring Operational Effectiveness Proved Difficult	114
Counterinsurgency through Economic Development: Was the Underlying Theory Valid?	124
Operation Assessment: DOD's Overarching M&E Analogue	144
Chapter Summary and Key Findings	154



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 unconducive 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, Totaling \$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 Szekely, "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.



CHAPTER 1 INTRODUCTION

USAID photo

The United States has appropriated more than \$144 billion for Afghanistan reconstruction since 2001.⁵ A large proportion of that total funded contracts, grants, and cooperative agreements through which third parties implemented projects that aimed to achieve reconstruction objectives.⁶ These contractors and nongovernmental 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.⁷ 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.⁸

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 may not have meaningfully advanced—the reconstruction effort. In its most recent calculation of cumulative waste, SIGAR estimated that total reconstruction losses have been approximately \$19 billion, representing 30 percent of the amount SIGAR reviewed.⁹ Among the losses are \$12 billion of funds spent on two whole-of-government efforts that appear to have failed entirely: **stabilization** and counternarcotics.¹⁰

Shortcomings with the effectiveness of contracting and the way in which reconstruction activities were monitored and evaluated has become a recurring theme of

Stabilization is defined as “a political endeavor involving an integrated civilian-military process to create conditions where locally legitimate authorities and systems can peaceably manage conflict and prevent a resurgence of violence.”

State, USAID, and DOD, “Stabilization Assistance Review: A Framework for Maximizing the Effectiveness of U.S. Government Efforts to Stabilize Conflict-Affected Areas,” 2018, p. 4.

reconstruction. There is roughly a decade's worth of literature documenting various kinds of contracting excesses in Afghanistan, and feedback from those media and inspectors general reports have led to increasing emphasis on contract monitoring and evaluation (M&E).¹¹ 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. Better and more candid M&E could have reduced the extraordinary magnitude of waste and reconstruction failure.¹²

M&E is the effort to determine whether a contractor or implementing partner fulfilled the terms of a contract, whether or not the contracted program or project accomplished its intended goals, and what useful lessons can be drawn. Although certain M&E concepts and practices overlap with those associated with oversight, there are important differences (see the callout box below). In theory, M&E allows U.S. agencies to measure performance, determine and learn from success and failure, make course corrections in ongoing programs in response to environmental changes or inadequate performance, 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 contracting officers and program managers 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.¹³ This trend massively 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.¹⁴ Yet contracting is essential: there is no practical way that U.S. agencies can do everything that contractors do. Many of the jobs contractors perform require specialized knowledge and skills.

M&E vs. Oversight

Both the Congress, particularly through committee staff, and the executive branch, through various inspectors general and the Government Accountability Office (GAO), engage in oversight of federal agencies. Agencies themselves oversee their respective contracted activities. While many of the methods employed in oversight work are similar to, and overlap with, those used by M&E, there are important differences of intent. M&E procedures are meant to determine whether a program or project is accomplishing a goal. By contrast, the primary aim of oversight is to ensure that programs and projects comply with relevant laws, regulations, and policies that exist to prevent waste, fraud and abuse.

As the U.S. government continues to rely extensively on contractors for reconstruction work, improving M&E and ensuring that officials act on M&E findings will remain critical needs. If the two-decade history of Afghanistan reconstruction is any guide, those needs must be urgently met. 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 (Table 1 below presents prominent examples).¹⁵ 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. It is intended for the Congress, for agencies involved in reconstruction contracting—especially DOD, State, and USAID—and for these agencies' contractors and implementing partners.

TABLE 1

A PATTERN OF POOR M&E: EXAMPLES FROM PRIOR SIGAR WORK					
Audit	Sector	Year	Agencies	M&E Issues Identified	Result of Poor M&E
Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion	Rule of law	2015	DOD, DOJ, State, and USAID	Lack of performance indicators ¹ No measurement against indicators ² Lack of baseline data	Outcomes and impacts unclear
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	Education	2016	DOD, State and USAID	Reliance on misleading data Claims of success lacked qualification	Magnitude of, and contribution to, outcomes and impacts unclear Misleading claims of success
Afghanistan's Health Care Sector: USAID's Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress	Healthcare	2017	USAID	Reliance on misleading data Claims of success lacked qualification	Magnitude of outcomes and impacts unclear Misleading claims of success
Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained	Infrastructure	2017	DOD, State, and USAID	No measurement against indicators	Outcomes and impacts unclear
Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program's Achievements and the Afghan Government's Ability to Sustain Them	Women's empowerment	2018	USAID	Lack of baseline data Inconsistent performance indicators	Outcomes and impacts unclear
USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure	Power	2019	USAID	Reliance on misleading data Inconsistent performance indicators Lack of baseline data	Outcomes and impacts unclear
Afghanistan's Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability	Power	2019	DOD, USAID	Lack of performance indicators ³ No measurement against indicators ⁴ Lack of baseline data	Outcomes and impacts unclear

¹ This audit's scope covered two different U.S. rule of law strategies. It was the U.S. Embassy's 2013 strategy that did not include performance measures.

² The 2009 rule of law strategy included performance indicators. However, no measurement was performed against them.

³ Both DOD and USAID implemented projects to increase access to electricity in southern Afghanistan. It was DOD that did not establish performance indicators for its projects.

⁴ USAID established performance indicators but did not report results against 9 of 18 indicators.

Source: SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015, pp. i–ii, 5; 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; SIGAR, *Afghanistan's Health Care Sector: USAID's Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017, p. i; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. ii; SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program's Achievements and the Afghan Government's Ability to Sustain Them*, SIGAR-18-69-AR, September 7, 2018, p. i; SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 4, 2019, electronic p. 3; SIGAR, *Afghanistan's Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, May 1, 2019, p. i.

Recent developments make this report especially timely. The withdrawal of U.S. troops from Afghanistan could be viewed as an implicit acknowledgement that U.S. objectives—which even now include “economic prosperity” and “stability”—have been unrealistic.¹⁶ The U.S. troop withdrawal is an opportunity to reset U.S. aspirations based on what can be reasonably accomplished, and to rethink how progress should be measured. Unless policymakers determine that a failed Afghan state is an acceptable outcome, the United States may have to provide significant assistance for years to come.¹⁷ Our hope is that this report helps policymakers and practitioners to avoid another two decades of poor M&E.

This report also has far-reaching application beyond Afghanistan in two important ways. First, it is almost axiomatic that the United States periodically becomes involved in large-scale reconstruction efforts such as that undertaken in Afghanistan. Should policymakers choose to launch another major reconstruction effort—several years or even decades from now—this report’s core lessons will likely still be highly relevant.

Second, even in the present, these lessons are likely generalizable to other country and regional contexts. U.S. efforts in Afghanistan are distinguished by their sheer scale. But 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, that logic undergirds U.S. efforts to counter instability in countries and regions around the globe.

The language of actual U.S. strategies illustrates the point. U.S. goals in Afghanistan include achieving “peace and stability,” fostering a “more stable, democratic, and accountable” government, “strengthening law enforcement,” and creating “economic prosperity based on private-sector-led exports, and job creation with social gains in education, health, and women’s empowerment.”¹⁸ All such goals “are grounded in the fundamental objective of preventing any further attacks on the [U.S.] by terrorists enjoying safe haven or support in Afghanistan.”¹⁹ Notably, the U.S. strategy in Pakistan follows a similar track: In order to promote “stability”—benchmarked in part against the extent to which terrorists are “using [Pakistani] territory to attack [external] targets”—the United States seeks to increase “private sector-led inclusive economic growth,” foster a “better educated, more skilled, inclusive, and healthy” Pakistani society, and increase “law enforcement capacity.”²⁰ Far removed from South Asia, the U.S. strategy for Niger, a regional “linchpin for stability,” follows suit: There, the United States seeks to “strengthen democracy, human rights, and good governance,” to professionalize Niger’s security forces to better “counter violent extremism,” and to “increase inclusive and sustainable economic growth.”²¹ Similarly, in Honduras, the United States aims to promote “sustainable economic growth,” to “support [the] effectiveness, transparency, accountability, and inclusiveness of government institutions,” and to improve security.²²

The list goes on. The implication is that the ability to properly apply M&E to the innumerable contracts, grants, and cooperative agreements that are intended to support these common elements of strategic logic extends far beyond Afghanistan’s borders.

In short, for as long as the United States continues to pursue its foreign policy interests through a combined “defense, development and diplomatic” approach that aims to promote stability, the findings, lessons, and recommendations of this report are likely to remain highly relevant.

SCOPE

This report examines M&E practices as applied to **contracts**, **grants**, and **cooperative agreements** for Afghanistan reconstruction directly administered by DOD, State, and USAID between 2002 and the present. (Figure 2 explains these mechanisms in further detail). Although Chapter 5 briefly surveys how other donor countries and multilateral organizations performed M&E, the report generally does not examine the Afghan government’s use of donor-provided funds or the M&E practices of public international organizations like the World Bank. Attempting to do so would have widened the report’s scope beyond what could be reasonably managed. Also excluded are contracts for materiel procurement, security, or for the construction, operations, and maintenance of Afghan security forces facilities and equipment. These activities have been covered in depth elsewhere, and would require examining a range of contracts that are too different to be meaningfully compared. Additionally, the report does not examine contracts for

FIGURE 2

CONTRACTING INSTRUMENTS

The main instruments used by U.S. government agencies for funding reconstruction activities are contracts, grants, and cooperative agreements—all of which are legally binding, though governed by different authorities. While DOD issues grants, cooperative agreements, and other non-procurement instruments for various purposes, it generally does not use these to implement foreign assistance in Afghanistan. Instead, it primarily uses contracts. Procurement is broadly divided into two categories: products and services. For the purposes of this report, DOD reconstruction contracting usually falls within the services category, which spans everything from advisory to construction management services. The table below defines each kind of award instrument and briefly describes the M&E and oversight requirements for each.

AWARD TYPES AND THEIR MONITORING REQUIREMENTS

Award Type	Definition	Monitoring and Evaluation and Oversight Requirements
Contract	A mutually binding legal instrument enabling the acquisition of goods or services for the direct benefit or use of the federal government. Contracts are inherently transactional in nature.	Requires frequent reporting. Exact time frames are written into individual contracts and are often determined by the type of activity.
Cooperative Agreement	A legal instrument allowing the transfer of money, goods or services to a recipient in order to accomplish a public purpose and authorized by the federal government. A federal entity is typically involved in implementation.	Reporting requirements determined by agreement.
Grant	A legal instrument providing for the transfer of money, goods, or services to a recipient in order to accomplish a public good, authorized by the federal government and not anticipating significant federal involvement.	Requires annual reporting to the relevant federal agency.

Table Source: USAID, “Glossary of ADS Terms,” revised April 18, 2018, pp. 55, 60, 118; Purdue University, “Federal Funding (Grants vs Cooperative Agreements vs Contracts),” n.d., accessed June 25, 2021, p. 1.

Source: Purdue University, “Federal Funding (Grants vs Cooperative Agreements vs Contracts),” n.d., accessed June 25, 2021, p. 1; Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling costs, reducing risks,” Final report to Congress, August 2011, p. 7; SIGAR, *Department Of Defense Spending on Afghanistan Reconstruction: Contracts Comprised \$21 Billion Of \$66 Billion In Total Appropriations, 2002–May 2014*, SIGAR-15-40-SP, March 2015, p. 8; SIGAR analysis.

goods and services to support the U.S. presence in Afghanistan, as they do not meet the criteria for “reconstruction contracting,” defined by the 2008 National Defense Authorization Act (NDAA) as:

Any major contract, grant, agreement, or other funding mechanism that is entered into by any department or agency of the United States government that involves the use of amounts appropriated or otherwise made available for the reconstruction of Afghanistan with any public or private sector entity for any of the following purposes: to build or rebuild the physical infrastructure of Afghanistan; to establish or reestablish a political or societal institution of Afghanistan; and, to provide products or services to the people of Afghanistan.²³

The report combines original research with findings from the considerable analytical work previously done by SIGAR and others. It will sometimes use the term “contracting” to generically refer not only to contracts but also to grants and cooperative agreements. The report looks at how U.S. agencies, primarily DOD, USAID, and State, have used monitoring and evaluation techniques to oversee contracting to ensure compliance and achieve reconstruction objectives.

This report primarily examines USAID, State, and DOD’s use of the following funds: Economic Support, Development Assistance, and Child Survival and Health, all of which finance mainly USAID programs; the Afghanistan Infrastructure Fund, jointly administered by DOD and State, with some monies flowing to USAID; International Narcotics Control and Law Enforcement, and Migration and Refugee Assistance, both of which are administered by State; and select programs under the Afghanistan Security Forces Fund, which is administered by DOD. The report builds on previous analyses of the Commander’s Emergency Response Program (CERP) and the Task Force for Business and Stability Operations (TFBSO). It also discusses the Afghan First Initiative, a cross-cutting contracting program financed by monies appropriated under several of the aforementioned funds.²⁴

Attempting to analyze the entire universe of U.S.-funded contracts, grants and cooperative agreements that have played a role in the Afghanistan reconstruction effort was beyond the scope of this report. Instead, this report took a sample of contracts, grants and cooperative agreements from DOD, State and USAID and attempted to compare each agency’s approach to managing projects of similar type and scope. The report also looks at stand-alone contracts for M&E. The report covers U.S., Afghan, and third-country contractors.

The wide-ranging terminology and definitions used by each agency reflect not only varying levels of M&E development but also different models of accountability, measurement, and planning. Appendix B provides further information on the terminology, definitions, and models of M&E at all three agencies. Overall, however, the core concept of M&E (explained further in the highlight on the next page) remains the same.

WHAT IS M&E?

In the context of this report, monitoring and evaluation is the attempt to objectively determine whether a reconstruction program—implemented through a contract, a grant, or a cooperative agreement—adhered to the terms of the award and achieved its intended outcomes. Ideally, agencies aim to establish causal attribution—that is, proof that the program caused an outcome or set of outcomes. In practice, they often loosely define the term “outcome” and stop short of establishing causation. There is a generally accepted nomenclature for the phases of, and the concepts related to, M&E. Like any specialized field, these unique terms and their sometimes complex definitions can make it difficult for those outside the field to comprehend how M&E works.

M&E is generally linked to a project (or program) “cycle.” The cycle begins with a theory—commonly referred to as a “theory of change”—that articulates a rationale for why an intervention (or set of interventions) will achieve desired goals.²⁵ The theory permeates every aspect of a project, informing its design and implementation, and defining the inputs, outputs, outcomes, and impact that will be monitored and evaluated. In general, the sequence of a project from inception to impact is supposed to look like this:



As a simple example, counterinsurgency theory might suggest that building or improving a road to connect a cluster of villages to a district center will increase support for the national government by showing the communities that the government has their interests at heart, creating a point of interaction between the local community and national government and increasing economic activity, including employment. The design process might involve:

- determining which villages to target and which local leaders to work with,
- identifying the specific stretch of road to be paved,
- ascertaining the right balance between construction efficiency and labor-intensive employment,²⁶
- determining compensation for private property disrupted by construction,
- and, finally, selecting a contractor to actually build the road.

Implementation—that is, project activities or “inputs”—would be the actual construction work. Outputs might be the paved road, the amount of short-term employment created, and increased interaction between government or coalition forces and local communities directly resulting from the construction.²⁷ Outcomes might be increased community engagement with the government resulting from the completed road (rather than just the process of construction), increased economic activity, lower consumer prices, and a longer-term increase in employment brought about by higher levels of local trade. If the original theory proved correct, the impact would be increased local support for the Afghan government. On the



Afghan men working on a road construction project in Badakhshan Province in 2009. (USAID photo)

other hand, the road could provide insurgents with an easy route to attack the district center, become a site for local warlords to set up roadblocks, result in increased violence from attacks on the road itself, or otherwise simply not catalyze increased support for the government—any of which would be detrimental to counterinsurgency goals and betray flaws in the underlying theory.

Monitoring is generally conducted during implementation.²⁸ It is supposed to collect and track data on inputs, outputs, and in some cases outcomes, to determine whether the desired results are occurring. Evaluations are often conducted at one or more points during or immediately after implementation, often using monitoring data as a starting point to understand how results are being achieved—or, if not, why not. Impact evaluations measure the change in an outcome or outcomes that is attributable to an intervention, and require a rigorously defined **counterfactual** in order to do so.²⁹ Information collected from M&E efforts is then supposed to inform the design of future projects.³⁰

A **counterfactual** articulates what would have occurred had an intervention not been implemented. In simple terms, the counterfactual for a road-building project would be the expected result had the project not been initiated and the road not constructed.

The highest standard of rigorous evaluation is establishing causation for, or at least plausible contribution to, a desired outcome. In a complex environment such as Afghanistan this can be extremely difficult. In the field, moreover, there are often numerous strands of effort, different inputs, and often multiple implementing partners. Real-life M&E models look more like a web than the line above. In addition, determining if a project had the desired outcome might include tracking public attitudes about the project and its outputs. This is also one of the major challenges of determining causation—measuring public attitudes involves polling, a notoriously inexact science. Moreover, the attitudes reported by survey respondents may not correlate with their actions.

Typically, lower-level results (such as activity completion and outputs) are more easily measured than higher-level ones (such as outcomes and impact). The latter may be hard to measure because they are abstract, are only recognizable over time, or are dependent on interaction with other projects or efforts.

To return to the road-building project above: The impact could be increased public support for the Afghan government—or the impact could be something completely unintended, unforeseen, or even negative. When a reconstruction project fails to produce its intended results, there are three common causes. It could be a flawed theory: The project planners failed to foresee how insurgents were likely to exploit a new road. It could be a failure in design: The road should have been built in a way that bypassed contested territory or motivated communities to support it. Or it could be a failure of implementation: It was built using inappropriate standards, was poorly constructed, or took so much time that the public came to see the project as a symbol of governmental incompetence and corruption.

There are also factors beyond the agency or contractor's control—for instance, a change in local laws governing a ministry's responsibilities. At times, it may be hard to draw a line between internal factors and external ones. For example, if a contractor or implementing partner cites predictable or known external factors such as “insecurity in project area” or “conservative social values and traditions” as a reason for not achieving objectives, one might question why those weren't addressed by the project's design. The goal of evaluation is to identify the causes of failure and use those observations to improve future contracts.

Although M&E is supposed to be a routine part of how government agencies oversee their contractors, the process rarely lives up to social science ideals. Establishing a cause and effect relationship between a given intervention and an impact of interest, for example, is notoriously difficult in environments like Afghanistan.³¹ Within an agency or implementing partner organization, there may also be a tension between proponents of rigorous M&E and practitioners who see M&E as a plodding, complicated activity that does not help them make decisions about their day-to-day implementation challenges, much less inform higher-level strategies. However imperfect, M&E is an attempt to ensure that government resources are used appropriately and effectively, particularly when a non-government entity is doing the actual reconstruction work.

A BRIEF HISTORY OF RECONSTRUCTION CONTRACTING: OVER TIME, U.S. GOVERNMENT OUTSOURCING BECAME MORE PREVALENT

The U.S. government's use of contractors in both wartime and peace is not a modern phenomenon. As the U.S. Superintendent of Finance noted in 1781, "In all countries engaged in war, experience has sooner or later pointed out that contracts with private men of substance and understanding are necessary for the subsistence, covering, clothing, and moving of an Army."³² More recently, the expanded use of contracting for expeditionary work has simply been part of the U.S. government's overall trend of outsourcing non-core functions. The only boundaries on outsourcing have been warfighting—in effect, a ban on hiring mercenaries to fight U.S. wars—and other "inherently governmental functions," those that cannot be performed by non-federal employees except in specific circumstances.³³ (A highlight at the end of this section discusses inherently governmental functions in more detail.)

The Vietnam War was a significant boon for so-called expeditionary stabilization and reconstruction contracting: An estimated 80,000 contractors supported American military operations, constructing everything from bridges and roads, to airfields and power plants.³⁴ More broadly, from the 1980s onwards, foreign aid shifted from government-to-government assistance, in which the foreign government was responsible for putting U.S. dollars to work under the oversight of U.S. government officials, to a system in which USAID and other U.S. government agencies contracted with the private sector to implement U.S.-funded projects.³⁵ The Reagan administration's efforts to shrink the size of government added another catalyst: Its general policy "to rely on commercial sources to supply the products and services government needs" was described by one author as "a free-market call to arms."³⁶ The Clinton administration's goal of "reinventing government" further encouraged the outsourcing of functions that had long been assigned to the federal workforce.³⁷ At USAID, then-Administrator Brian Atwood volunteered the agency as a "laboratory for reinvention," which accelerated the move toward "focusing on partnerships and private sector relationships."³⁸ For DOD, U.S. operations in the Balkans during the 1990s provided the next step in the evolution of expeditionary contracting: the ratio of contractors to military personnel there was roughly 1:1.³⁹ By the mid-2000s, the government-wide transformation was complete. As a 2007 report from a bipartisan commission on U.S. foreign assistance noted:

Over the last 20 years, the contract and grant award processes used by U.S. agencies have transformed from an occasional administrative support function into what is now often the agencies' main responsibility. Agencies currently accomplish their missions principally by awarding grants and contracts to non-federal entities.⁴⁰

Accelerating the general trend toward outsourcing, the U.S. military interventions in Afghanistan and Iraq following 9/11 greatly expanded the market for defense and development contractors. As researchers Richard Fontaine and John Nagl have observed, "The post-invasion reconstruction environments in Iraq and Afghanistan serve as the largest-ever market for private firms providing [expeditionary stabilization and reconstruction] services."⁴¹ This was accompanied by the dramatic expansion in scale and scope of DOD's reconstruction and development activities, based in part on the

perception that it was more capable than State and USAID.⁴² Between 2002 and 2005, the percentage of U.S. global development funding that went through DOD increased from 6 to 22 percent, while USAID's decreased from 50 to 39 percent.⁴³ At the same time, total U.S. official development assistance nearly tripled between 1999 and 2008 as development was elevated, in the words of one USAID-commissioned report, to an "equal footing with defense and diplomacy" as a critical and increasingly militarized element of the fight against global terror.⁴⁴

To enlist reconstruction activities as an instrument of warfighting, more flexible authorities such as the Commander's Emergency Response Program were created.⁴⁵ So that they could be deployed quickly, CERP funds were initially appropriated by the Congress with a "notwithstanding" clause that ensured they were unencumbered by procedures typically associated with federal procurement, including those required by the **Federal Acquisition Regulation (FAR)**.⁴⁶ The intent was to enable military commanders, in the words of the Congress, "to respond to urgent humanitarian relief and reconstruction requirements within their areas of responsibility."⁴⁷ Even in its early years, the program was thought of as a stabilization tool and eventually came to be viewed as a key component of the U.S. counterinsurgency effort in Afghanistan.⁴⁸ As part of that effort, the Afghan First Initiative was launched in 2006 by Lt. Gen. Karl Eikenberry, then-commander of the Combined Forces Command–Afghanistan. The initiative prioritized contracts with Afghan companies to promote economic development and create jobs, the theory being that better economic conditions would support stability and further counterinsurgency objectives.⁴⁹

The Federal Acquisition Regulation

The **Federal Acquisition Regulation** is the result of a 1979 statute that directed the Office of Management and Budget (OMB) to develop uniform procurement regulations for the federal government.⁵⁰ The FAR contains "the primary regulations for use by all executive agencies in their acquisition of supplies and services with appropriated funds."⁵¹

The FAR is granular in some places and broader in others. For example, it lays out in detail the process that a **contracting officer** must follow in order to properly enter into a contract, outlining standards ranging from required market research to the application of labor laws.⁵² (The figure on the following page defines "contracting officer" as well as the titles of other key contracting personnel.) More broadly, the FAR establishes an overarching framework for federal contracting by setting forth certain guiding principles, including that contracts must "fulfill public policy objectives."⁵³

Certain sections of the FAR are explicitly relevant to M&E. For instance, when contracting for services, there are specific requirements for the kinds of metrics agencies must include in performance work statements (core documents of performance-based contracts that describe the work contractors are supposed to perform).⁵⁴

Appendix B discusses the FAR further.

FIGURE 3

CONTRACTING PERSONNEL

The key positions for contract management are the contracting officer and the contracting officer's representative (frequently referred to in the literature as COs and CORs). The FAR gives contracting officers the responsibility for “ensuring performance of all necessary actions for effective contracting, ensuring compliance with the terms of the contract, and safeguarding the interests of the United States in its contractual relationships.” Contracting officers are empowered to enter into contracts on behalf of the U.S. government, and are the backbone of the contracting system.

As contract management can sometimes be a detailed, time-consuming affair, the FAR provides a contracting officer with the authority to designate a representative to assist with contract administration. The contracting officer's representative is responsible for most day-to-day management, and is an essential liaison between contract administration offices, which handle legal and monetary matters, and program offices responsible for ensuring that programming meets intended objectives. Contracting officer's representatives have less authority than contracting officers and, per the FAR, cannot “make any commitments or changes that affect price, quality, quantity, delivery, or other terms and conditions of the contract.”

For cooperative agreements and grants, the functions are similar, but performed by an agreement officer/agreement officer's representative and grant officer/grant officer's representative, respectively. (Acronyms follow suit: there are AOs and AORs, as well as GOs and GORs.) These officers perform essentially the same function as contracting officers and contracting officer's representatives, but within the structure of cooperative agreements and grants.

CONTRACT MANAGEMENT POSITIONS AND THEIR MONITORING RESPONSIBILITIES	
Management Position	Responsibilities
Contracting Officer	A contracting officer has the authority to enter into contracts on behalf of the U.S. government and the responsibility to “[ensure] performance of all necessary actions for effective contracting, [ensure] compliance with the terms of the contract, and [safeguard] the interests of the United States in its contractual relationships.”
Contracting Officer's Representative	A contracting officer's representative handles the day-to-day management of a contract or contracts, and typically resides closer to where the program or project is implemented. A representative is subordinate to a contracting officer, and cannot change the terms and conditions of a contract without the officer's authority.
Agreement Officer	An agreement officer acts within the scope of a duly authorized warrant or other authority to “enter into, administer, terminate and closeout assistance agreements, and make related determinations and findings on behalf of [the U.S. gov-ernment].” “Grant officers” are included under this term.
Agreement Officer's Representative	Sometimes called a technical representative, this role's functions are designated by the supervising agreement officer.

Table Source: Federal Acquisition Regulation, Part 1: “Federal Acquisitions Regulations System,” Subpart 1.6: “Career Development, Contracting Authority, and Responsibilities,” §1.602: “Contracting officers,” subsection (2), “Responsibilities”; USAID, “Glossary of ADS Terms,” revised April 18, 2018, pp. 15–16, 57, 60, 118; DOD, “Defense Contingency COR Handbook version 2,” September 2012, pp. 191–192.

Source: Federal Acquisition Regulation, Part 1: “Federal Acquisitions Regulations System,” Subpart 1.6: “Career Development, Contracting Authority, and Responsibilities,” §1.602: “Contracting officers,” subsections (1), “Authority,” and (2), “Responsibilities”; USAID, “Glossary of ADS Terms,” revised April 18, 2018, pp. 15–16, 57, 60, 118; DOD, “Defense Contingency COR Handbook version 2,” September 2012, pp. 72, 191–192.

In 2009, the Obama administration dramatically increased troop and reconstruction assistance levels and introduced an explicit stabilization strategy to counter an increasingly resurgent Taliban.⁵⁵ A July 2009 cable articulating the new approach to the delivery of U.S. assistance noted that increasing local procurement by expanding the Afghan First Initiative and contracting with more local firms would show U.S. commitment by “putting Afghans in charge of their country’s development.”⁵⁶ Accountability was to be improved by using “smaller, flexible, and faster contract and grant mechanisms to increase decentralized decision-making in the field.” This included simplifying acquisition procedures and shifting USAID personnel to regional platforms in order to better coordinate with the U.S. military, the Afghan government, and the myriad other actors operating in support of counterinsurgency objectives.⁵⁷ The Obama

administration's increase in resourcing and shift in strategy also changed the goals and conduct of many DOD reconstruction programs, with programs such as CERP explicitly shifting their goals from humanitarian aid and reconstruction to improving government legitimacy and displacing Taliban influence.⁵⁸

The infusion of money and personnel for stabilization activities led to massive increases in the number and value of development contracts. As part of that effort, the U.S. military pushed an understaffed USAID to develop a new set of development projects to complement expanded military operations.⁵⁹ The pressure to produce results in a short amount of time, sometimes without specific definitions of what those results might be or reliable means of measuring them, meant that the “burn rate”—the rate at which money was spent on a program—was often the only way to measure “success.”⁶⁰ In its most basic form, the metric relied on the questionable assumption that simply spending more money correlated with, or even directly contributed to, operational goals.⁶¹

With inadequate oversight, the torrent of money that had to be spent in a short period of time led, unsurprisingly, to **corruption**—some of which was extravagant. Media reports continued to reinforce the narrative of rampant corruption and document where U.S. funding was diverted to the Taliban and other insurgent or criminal groups.⁶² The media reports were confirmed by a number of official U.S. government reports, which criticized the diversion of USAID development funds.⁶³ One of the most notorious practices was “contract flipping,” whereby a primary contractor sold the contract to a subcontractor, after skinning a percentage of the contract's value. This was repeated successively until the final contractor could not adequately complete the work with what remained of the money.⁶⁴

What Is Corruption?

Throughout this report, the term “**corruption**” refers to “the abuse of entrusted authority for private gain.” This widely accepted definition appeared in USAID's 2005 anticorruption strategy and closely parallels definitions used by various international organizations.⁶⁶ Corruption takes various forms, including bribery, embezzlement, extortion, and fraud.⁶⁷ In addition to obvious permutations of corruption, such as making illicit payments to public officials in exchange for contracts, there are also gray areas, such as legally permitted nepotism—for example, the appointment of friends or family to official positions even though they may not be fully qualified.⁶⁸

Interest in accountability peaked during this period, in large part because corruption and ineffective programming were seen as undermining the counterinsurgency strategy. As a USAID report noted, “the ‘hearts and minds’ component of the counterinsurgency strategy relies heavily on successful development activities to ensure the areas that are cleared of insurgents by the military remain that way through development efforts.”⁶⁵ From 2008 through 2010, the United States formed a number of interagency anti-corruption groups “tasked with ensuring that the U.S. and international community's dependency on contractors does not ultimately undermine the progress made in

Afghanistan.” These included various groups housed within U.S. Forces – Afghanistan, as well as the International Security Assistance Force’s Combined Joint Interagency Task Force–Shafafiyat (“transparency” in Dari).⁶⁹ During 2009, State issued four strategies related to anti-corruption, including the “Anticorruption Action Plan for Afghanistan” and the “U.S. Government Anticorruption Strategy for Afghanistan.”⁷⁰

In the fall of 2010, USAID launched its internal Accountable Assistance for Afghanistan working group, which included personnel from the Office of Acquisition and Assistance, the Office of Afghanistan and Pakistan Affairs, and the Office of Security. According to the group’s report, “The team’s mission [was] to identify shortcomings, and to help ensure that the agency [was] taking the steps necessary to limit the likelihood assistance [was] . . . directly or inadvertently supporting malign groups or being diverted from their development purpose by extortion or corruption.”⁷¹

The working group’s June 2011 report described the steps that USAID was taking to enhance oversight and avoid funding the insurgency and other malign actors. These fell largely into four areas: reform of the contracting process, vetting of partners, improved financial controls, and increased oversight.⁷² In particular, the report noted new rules that would reduce the practice of contract flipping, requiring the prime contractor to perform at least 15 percent of the work and limiting the number of tiers of subcontractors to two.⁷³ The report noted that fixed price contracts—which set a price for work regardless of time or resources expended—posed the most risk “due to the limited documentation required, and projects with multiple layers of subcontractors, due to USAID’s limited visibility over subcontracts.” The report also mentioned the intention to limit the use of cooperative agreements (which had fewer oversight requirements), to create a contract management team, and to increase vetting of partners.⁷⁴

Several years later, in the run-up to the 2014 transition to Afghan authority, staffing reductions and the resulting narrower physical footprint of U.S. government personnel raised concerns about agencies’ ability to oversee reconstruction programs.⁷⁵ According to several USAID officials, experiments in trying to perform oversight remotely from other, safer countries created too many logistical and management issues, as “face time” and site visits (however limited) were crucial.⁷⁶ The U.S. government staff that remained in country were limited in their ability to actively manage contracts by increased restrictions on movement and reduced military support for travel to program sites.⁷⁷

THE CONCEPT OF INHERENTLY GOVERNMENTAL FUNCTIONS REFLECTS THE RISKS AND BENEFITS OF CONTRACTING

Decisions to contract often involve careful cost-benefit calculations. These vary, depending on the purpose of a contract, the contracted party, and the context within which a contract is expected to be implemented. On one hand, contracting can provide additional capabilities to the U.S. government, particularly when special expertise is required or when projects are implemented in less secure environments that government officials cannot easily access. On the other, contracting runs the risk of effectively outsourcing sovereignty, particularly in cases where private parties assume direct implementation responsibility for wide-ranging projects with far-reaching political consequences—like reconstructing a war-torn country. No single concept embodies the potential risks and benefits of contracting more than the term “inherently governmental functions.”

“Inherently governmental functions” are those that, by U.S. law and regulation, cannot be performed by non-federal employees.⁷⁸ In practice, however, aside from the legitimate use of violence, the borders are not always clear. Although there seems to be general agreement that certain functions should not be outsourced because they are so innate to government, what exactly those functions are is debatable.⁷⁹ The FAR lists 17 examples of broad functions that cannot be contracted—including core tasks such as “the command of military forces,” and “the conduct of foreign relations and the determination of foreign policy”—but also explicitly notes that the examples it provides are not all-inclusive.⁸⁰ As a 2009 report from the Center for a New American Security observed, “present statutes and regulations offer very limited guidance for determining precisely which functions fall into this category.”⁸¹

But even when outsourcing is clearly permitted, it may not always be wise—particularly in a wartime environment where the stakes are high. As the Commission on Wartime Contracting in Iraq and Afghanistan observed, “The ‘inherently government’ standard in law, policy, and regulation that reserves certain functions for government personnel” does not provide sufficient guidance for contracting in contingencies, “nor does it enable

officials to decide whether contracting for non-governmental functions is *appropriate* or prudent in contingency operations” (italics in original).⁸² Notably, the sheer number of contractors working alongside the military in contingency environments like Iraq and Afghanistan muddies the concept of inherently governmental functions.⁸³ While warfighting itself is not technically performed by contractors, the use of private security firms puts contractors in positions where they could encounter combat.

A wide range of risks can render contracting inappropriate under certain circumstances—not only because it might stretch legality to contract out certain governmental functions, but also because heavy reliance on contractors can pose more fundamental risks. According to the Commission on Wartime Contracting, these include:⁸⁴

- The *operational* risk that expansive contracting could lead to the loss of mission-essential government capabilities;
- The *political* risk that the full human cost of war is obscured by putting contractors, rather than U.S. government personnel, in harm’s way—particularly because contractor deaths often go uncounted and unpublicized;
- The *financial* risk created when a large portfolio of contracts cannot be adequately overseen, leading to waste, fraud, or abuse.

Other factors may make outsourcing a perfectly reasonable choice. As the commission also pointed out, contracting can:⁸⁵

- Help develop rapport with host-nation governments and communities and support the local economy when local contractors are used with discretion;
- Help free up the military to focus on its primary responsibility for warfighting rather than performing supplemental support functions;
- Lower the costs of contingency operations, particularly with the full use of competitive bidding processes.

CONTRACTING AND M&E IN AFGHANISTAN: ALL THE USUAL CHALLENGES, PLUS A WARTIME ENVIRONMENT

Afghanistan's complex environment had a number of characteristics that exacerbated the normal challenges of contract management and oversight: insecurity, unpredictability, limited mobility, lack of information about basic demographics (Afghanistan's last attempt at a census was in 1978), and corruption. Added to that were two factors that surface again and again in various inspectors general and think tank studies: inadequate human resources, and haste—specifically, political pressure to implement programs and spend money at a pace that outstripped administrative and oversight capacity.⁸⁶ With limited government M&E resources and capacity, many aspects of M&E are effectively outsourced to the primary contractor—a situation that creates the potential for bias.⁸⁷

In Afghanistan, numerous reports documented lackluster M&E, including the lack of performance metrics to assess if objectives were met, the absence of baseline studies, and inaccurate or suspect data collected from the field.⁸⁸ There is no clear evidence that U.S. agencies' attempts to develop and improve M&E systems, and to generate actionable information from them, consistently produced an accurate, objective answer to a fundamental question: whether reconstruction funds were having their intended impact.⁸⁹ Instead, M&E practices too often 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. Figure 4 on the next page presents examples of this risk. These examples are discussed in detail later in this report.

At the beginning of the U.S. involvement in Afghanistan in October 2001, the priority was implementation—getting funds out the door for humanitarian and reconstruction activities that would support the United States' political and counterterrorism mission.⁹⁰ Initially, the Bush administration's resistance to "nation-building" prevented the designation of clear reconstruction objectives to measure program progress against.⁹¹ Within USAID, much of the early work was construction of roads, schools, and clinics, and early efforts at monitoring were in reality basic contract oversight: visiting construction sites to check that proper materials were being used, usually in response to concerns raised through informal reporting rather than systematic monitoring.⁹² As one official noted, "Initially, there was not an overarching M&E plan in Afghanistan. There was no appreciation of how M&E can support decision-making. . . . Those on the ground were fighting for M&E, but they felt like they weren't being heard."⁹³ According to another USAID official, "Until the end of 2014, there were a lot of M&E plans, but they weren't closely followed."⁹⁴

Under Zalmay Khalilzad, the U.S. ambassador to Afghanistan from 2003 to 2005, a "whole-of-embassy" monitoring system was established under the leadership of the Afghanistan Reconstruction Group, which was created by the National Security Council and staffed by senior private sector executives and military officers. The Afghanistan Reconstruction Group was seen as Khalilzad's way of working around U.S. government aid mechanisms that he considered cumbersome.⁹⁵ In addition to providing a private-

FIGURE 4

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-poppo cultivation. However, during implementation the program dropped indicators that were directly related to reducing opium-poppo 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, Totaling \$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, p. 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, pp. 182; SIGAR analysis.

sector perspective on reconstruction to both the ambassador and Afghan ministers, the group functioned as an ad hoc, high-level M&E unit, setting metrics for activities and meeting regularly to review progress.⁹⁶ While the Afghanistan Reconstruction Group was valued for its strategic input, its unclear position within the embassy caused tension with other agencies: USAID and State officials argued that the group "focused its efforts on criticizing USAID programs rather than providing constructive advice."⁹⁷ It was dissolved in 2008.⁹⁸

The **Commission on Wartime Contracting in Iraq and Afghanistan** was created by the Congress in 2008 to examine the extent of waste, fraud, and abuse in contingency contracting in Iraq and Afghanistan and to improve policies for managing contracts in contingency environments. The commission no longer exists, having published its final report in August 2011.

Commission on Wartime Contracting in Iraq and Afghanistan, "Transforming Wartime Contracting: Controlling Costs, Reducing Risk," August 2011, p. iv; Commission on Wartime Contracting in Iraq and Afghanistan, "Wartime Contracting Commission closes its doors September 30," press release, September 28, 2011, p. 1.

Meanwhile, concerns mounted over whether reconstruction contracts were being properly overseen. In attempts to improve accountability in contracting, the Congress established the **Commission on Wartime Contracting** and the Special Inspector General for Afghanistan Reconstruction.⁹⁹ High-profile scandals reinforced congressional concerns. One of the most visible involved corruption by the Louis Berger Group connected with construction of the Kabul-Kandahar highway. That project was part of the USAID-funded Rehabilitation of Economic Facilities and Services Program, an ambitious effort to build infrastructure, support public services, and build capacity in the Afghan construction sector. Accusations against the contractor included bribing foreign officials and defrauding the U.S. government by billing inflated overhead rates. The legal process continued until 2015, when the Louis Berger Group's former chairman and CEO pled guilty to conspiring to defraud USAID.¹⁰⁰

Reports of poor contractor oversight and of diversion of U.S. funds to criminal or insurgent activity created the impetus for strengthened M&E.¹⁰¹ The Government Accountability Office also expressed concern that agencies were not adequately tracking contractor personnel in Afghanistan and Iraq.¹⁰² The Congress instituted a variety of new reporting requirements to promote transparency. Vetting of non-U.S. vendors to ensure that they did not have criminal or other compromised backgrounds and were not affiliated with insurgent groups started in 2010 for DOD, 2011 for USAID, and late 2012 for State.¹⁰³

In 2007, USAID and State developed a set of global metrics called the **standard foreign assistance indicators**, commonly referred to as the "F Indicators." Intended to establish a standard set of measurement across U.S. government agencies working around the world, the standard indicators had little impact on how monitoring and evaluation was done on the ground in Afghanistan, but did create a burden for officials who were required to use them. The number of standard indicators was daunting: In fiscal year 2016, there were nearly 2,300.¹⁰⁴

The military surge and the related influx of civilian contractors that began in 2010 created even more concern about accountability and the threat to the Afghanistan mission posed by contract waste, fraud, and abuse and by poorly performing projects. The Obama administration's new approach to the delivery of assistance was to propose more emphasis on oversight, in addition to procurement changes (such as one-year contracts subject to rigorous performance reviews).¹⁰⁵ President Obama stated this in his speech on U.S. strategy in Afghanistan:

Going forward, we will not blindly stay the course. Instead, we will set clear metrics to measure progress and hold ourselves accountable. We'll consistently assess our efforts to train Afghan security forces and our progress in combating insurgents. We will measure the growth of Afghanistan's economy, and its illicit narcotics production. And we will review whether we are using the right tools and tactics to make progress towards accomplishing our goals.¹⁰⁶

The U.S. embassy noted that civilian oversight would "increase dramatically," and even went so far as to note that it would be "working closer with our Special Inspector

Standard Foreign Assistance Indicators

Designed in 2007, the **standard foreign assistance indicators** were part of a reform initiative led by then-Secretary of State Condoleezza Rice. They were intended to provide a standard way for all U.S. agencies to measure the collective impact of foreign and host-government development work.¹¹¹ The indicators have been widely criticized, particularly for their focus on “what USAID and State were *doing* with their budgets rather than what they were *achieving* with their budgets.”¹¹² Many indicators, in other words, measure output data that contributes little to knowledge about a program’s impact on foreign assistance objectives—for example, metrics like the “number of individuals who complete [U.S. government]-assisted workforce development programs” and “the number of people in a host country who received [U.S. government]-funded counternarcotics training.”¹¹³ While the standard indicators were “intended to streamline monitoring and reporting” they are widely considered to be “extraordinarily rigid and labor-intensive,” and to take away time and effort from other responsibilities.¹¹⁴

General for Afghanistan Reconstruction (SIGAR) and Regional Inspector General (RIG) colleagues to provide adequate oversight.”¹⁰⁷ Procurement and M&E were seen as related: Shortcomings in both were diluting the impact of assistance and undermining confidence in both the United States and the Afghan state it was trying to build.

Although Afghanistan was a special case due to its political visibility and the scale of operations, the emphasis on measurement occurred in the context of the Obama administration’s global attempt to modernize foreign assistance efforts. Special emphasis was placed on increasing accountability and better understanding efficacy. A 2010 Presidential Policy Directive on Global Development announced the administration’s intention of bringing rigorous M&E practices to bear by “[adopting] metrics, appropriate to our objectives and the context, against which we can measure progress. . . . We must hold accountable all countries to which the United States provides assistance, including those to which we have provided substantial assistance over years or decades.”¹⁰⁸

Parallel efforts to better leverage M&E were afoot within agencies. In a sweeping review of U.S. diplomacy and development published in 2010, State noted, “Monitoring and evaluation is one of the most powerful ways of assuring accountability for performance, transparency, and the ability to modify programs for enhanced impact.” The review proposed reforms based in part on USAID practices, including “establishing new requirements for performance evaluations, designing rigorous impact evaluations, linking evaluations to future funding decisions, and promoting the unbiased appraisal of programs and the full disclosure of findings.”¹⁰⁹

Around the same time within DOD, the search was on to find the right methods to measure counterinsurgency effects. The result was a series of complex and mostly quantitative assessment systems that relied on a plethora of indicators—at least one of which included the color-coded district stability maps that later became a well-known marker of war progress.¹¹⁰ Defense entities of all stripes, from the U.S. military to U.S. civilian

agencies, struggled to adapt their bureaucracies to analyze and understand conditions on the ground.¹¹⁵ As one RAND researcher wrote, “At one point in 2010, all the interested parties were collectively producing nearly 50 assessments of all kinds, delivered at various intervals by staffs from the battalion to the NATO headquarters level.”¹¹⁶

The most notable attempt by the military to understand how the war was progressing was conducted by then-International Security Assistance Force Commander General Stanley McChrystal in 2009. The final product became one of the most well-known documents of the war, both for its criticisms of Afghan government corruption and because it was leaked to the press in an apparent attempt to influence then-imminent decisions about possible troop increases.¹¹⁷ In what had become a recognizable pattern, McChrystal asserted that with additional resources and a new strategy “success is achievable.”¹¹⁸ As former British diplomat and politician Rory Stewart summarized, “Each new general in Afghanistan from 2002 to 2011 suggested that the situation he had inherited was dismal; implied that this was because his predecessor had had the wrong resources or strategy; and asserted that he now had the resources, strategy and leadership to deliver a decisive year.”¹¹⁹ The scale of success envisioned by the 2009 assessment remained elusive.

To a great extent, that success hinged on the capabilities of the Afghan security forces, which the military also strove to measure. That proved challenging: By 2010, the initial rating mechanism had been replaced.¹²⁰ The various systems used over the course of reconstruction graded units along comparable spectrums of dependence on coalition support. But the systems shifted over time, emphasizing different inputs, analyzing different levels of command, varying in their thresholds for achieving a given score, and in many cases using different words to describe individual rating levels.¹²¹ Such changes reflected persistent questions about what data mattered, how to balance quantitative and qualitative information, and, more fundamentally, how exactly to measure capability in the first place. Adjustments to the systems to some extent reflected the military’s ability to adaptively answer these important questions. Nonetheless, the systems were criticized for their inconsistency.¹²²

USAID was nominally better positioned to fulfill the new mandate for measurement. Its first evaluation project in Afghanistan had begun in 2006 in response to the general need for M&E.¹²³ The number of evaluations performed through the program jumped from one in fiscal year 2008 to eight in fiscal year 2009, and peaked at 10 in fiscal year 2012, the program’s final year.¹²⁴ The focus on evaluation within USAID was further buttressed by the release of a new global evaluation policy in 2011.¹²⁵ Other agencies followed suit. The following year, State issued its first department-wide evaluation policy, based in part on USAID’s. Up to then, only about one-fourth of the 24 bureaus that worked with foreign assistance funds had ever performed evaluations of their contractors.¹²⁶

Widespread interest in M&E, combined with the sheer number of ongoing projects in Afghanistan, led to the development of new tools and systems. In 2010, USAID launched

its online Afghan Info database—an effort that was initially hampered by the lack of reliable input.¹²⁷ A recurring theme in this report is that the existence of M&E systems does not always translate into effective M&E.¹²⁸

The overall emphasis on measurement was often tied to efforts to better understand the effects of counterinsurgency and stabilization projects. Many such projects had somewhat abstract goals that were notoriously difficult to measure, such as building public confidence in the Afghan government. One of the most prominent of these efforts was USAID’s Measuring Impacts of Stabilization Initiatives (MISTI), initiated in 2012 to evaluate the impact of the stabilization programs then in high gear in key areas of the country. MISTI was a complex third-party monitoring and evaluation program that looked at a range of socio-economic indicators, violent incident counts, U.S. stabilization project performance indicators, and population perceptions to track stability effects at district, provincial, and regional levels.¹²⁹ A SIGAR audit of MISTI credited the program for meeting its objectives to track stabilization trends and provide best practices for stabilization programs, but noted that its recommendations were ultimately of limited value in Afghanistan because most of USAID’s stabilization programs had ended by the time MISTI had concluded.¹³⁰ (The development and use of MISTI is discussed in detail in Chapter 2.)

The interest in metrics, along with the massive budgets associated with counterinsurgency and the surge, as well as the development of new information and communication technology (see highlight on next page) contributed to the rise of analytics and fueled a cottage industry of international and Afghan research and M&E firms selling a variety of products and apps: public opinion polling, contract research, activity monitoring, and analysis services, including assessment and evaluation.¹³¹ In the years leading up to 2012, the estimated number of Afghan and international commercial firms conducting research in Afghanistan increased from three or four to somewhere between 25 and 30.¹³² The work they produced was of varied quality and reliability, especially when it came to the accuracy of the data collected from insecure areas.¹³³ Data quality assessments found that, even with the more reputable surveys, sampling was skewed towards the better off and better educated, responses were affected by social desirability bias (the tendency of respondents to provide answers that will be viewed favorably by others), and respondents’ understanding of the questions was different from that of the highly educated persons who designed them.¹³⁴ In addition, firms’ choice to pay surveyors on a per-interview basis set up perverse incentives to fill out forms quickly, perhaps even without actually conducting interviews.¹³⁵

As the international military presence reduced in the run-up to the 2014 transition, U.S. and international institutions optimistically described this as the transition “from counterinsurgency and stability operations toward more traditional diplomatic and development activities.”¹³⁶ But those “traditional” activities would have to be carried out in the midst of an ongoing war. Oversight and implementing agencies began to express concerns that the combination of declining security and the withdrawal of most international military forces would limit mobility and make M&E more difficult.¹³⁷

TECHNOLOGICAL ADVANCEMENTS CREATED NEW TOOLS FOR M&E, BUT POTENTIAL BENEFITS WERE SOMETIMES STYMIED

The U.S. engagement in Afghanistan coincided with rapid technological innovation, opening up new avenues for monitoring and evaluation. Perhaps the most significant was the smartphone, which quickly became an integral part of ground-based data collection.¹³⁸ Geo-tagging allowed agencies to confirm the location of monitors in the field, and attached time and location data to photos and videos. This added another layer of verification, theoretically providing more reliable information about ongoing projects.¹³⁹ Other innovations—for example software that could “kill” a phone remotely—helped protect personnel reporting on donor-funded projects in dangerous areas.¹⁴⁰

Technology could also be used to mitigate fraud. Under the USAID women’s empowerment program Promote, for example, reports emerged that participants were sending their friends or sisters to program events in order to collect the stipend provided to attendees. In response, the program started using biometric data like fingerprints to track the progress of participants and limit the potential for misuse.¹⁴¹

Of course, technological advancements did not come without risks and detractors. One former senior USAID official recounted how a clandestine business in Kandahar would, for a fee, provide contractors with generic photographs of projects customized with fraudulent geo-tags embedded in the digital photos.¹⁴² Some observers believed that technology quickly became something of a fetish. As one critic noted, “Technology can appear to be a silver bullet if you don’t have a good understanding of M&E. Too much data is like white noise. You can have all of the widgets, but you need the time to analyze and to specify a theory of change, rather than trying to measure everything. And, if you don’t have the hours to assess and review data, it’s useless to push for more of it.”¹⁴³

But when time was available, more data was sometimes better, particularly for outcome and impact assessment. Geospatial data derived from satellite imagery allowed for monitoring of crop growth across vast land areas—and

by extension, for analysis of the impact of alternative development programs on illicit opium cultivation.¹⁴⁴ Satellite imagery could also be combined with census data to track internal migration in order to better understand local economic and security conditions.¹⁴⁵ Meanwhile, the military created large datasets that could be used to help assess how the war was progressing. Most prominently, it tracked attacks against coalition forces—often referred to as “significant activities” or “SIGACTs.”¹⁴⁶

Various researchers used these and other data to ascertain the effects of reconstruction programs. For example, a RAND assessment of the Commander’s Emergency Response Program in Afghanistan drew from “a broad array of different types of data as outcome measures, including agricultural production, ambient light, SIGACTs, [International Security Assistance Force] vehicular movement, intelligence reporting, migration rates, nonmilitary vehicular movement, market activity and survey data.”¹⁴⁷

The contributions of such research were often complex, nuanced, and heavily qualified. For example, researchers for RAND’s study of CERP acknowledged that although their quantitative data sources were “diverse,” they functioned “only as proxies for the outcomes in which we are interested.”¹⁴⁸ Although important, these kinds of qualifications often meant that interpretations of research results could vary widely. For example, in the case of the RAND study, advocates of CERP could highlight the study’s conclusion that “cumulative CERP activity is associated with improved security and economic conditions for local populations.”¹⁴⁹ Other observers could point to the key result “that quantitative analyses cannot credibly identify the impact of CERP independent of overall counterinsurgency efforts.”¹⁵⁰

Overall, technology created opportunities for better M&E of reconstruction activities, even if bad actors could sometimes circumvent its potential oversight benefits and even if its efficacy was sometimes stymied by Afghanistan’s complex environment.

These concerns occurred against the backdrop of the demand for increasing levels of accountability for the money spent in Afghanistan, with much of the focus being on Afghan government corruption.¹⁵¹ Official documents of the period, which were steeped in the language of accountability and measurement, reflect this concern. For example, the 2012 Tokyo Mutual Accountability Framework—an agreement signed at the July 2012 Tokyo Conference on Afghanistan defining the partnership between the international community and Afghanistan—predicated \$16 billion of donor funding for the following four years on commitments by the Afghan government to a variety of governance reforms, with periodic reviews of specific progress indicators.¹⁵²

U.S. agencies attempted to meet these challenges by introducing new or strengthened methodologies, such as multitiered and third-party monitoring. Multitiered monitoring, introduced by USAID in early 2013, was supposed to improve information reliability by checking multiple data sources from different actors against each other.¹⁵³ It was widely considered to be ineffective—partly because a team created by USAID to monitor and analyze information produced through multitiered monitoring lacked access to relevant data, consisted of USAID employees with other full-time jobs, and received insufficient guidance regarding their roles and responsibilities.¹⁵⁴ Third-party monitoring—contracting of an independent entity to monitor projects in the field—had been adopted from the start of reconstruction by USAID, the World Bank, and the European Commission in the health sector. But particularly during the 2014–2015 transition, it came to be viewed as a crucial source of information about projects given ongoing insecurity.¹⁵⁵ (Multitiered monitoring and third-party monitoring are discussed in detail in Chapter 2.) In October 2013, USAID established a working group to develop a compendium of best practices in implementing and monitoring projects.¹⁵⁶

In an attempt to strengthen its M&E regime, in 2015 State rolled out the core of its current M&E system: the Managing for Results (MfR) framework. Created by State’s Office of U.S. Foreign Assistance, MfR attempted to formally link strategy, resource management, program design, and M&E, with monitoring and evaluation integrated into each step.¹⁵⁷ Concurrently, State’s Bureau of International Narcotics and Law Enforcement Affairs (INL) implemented State’s first M&E-specific contract in Afghanistan, the Flexible Implementation Assessment Teams program. The program provided third-party monitoring of INL’s contracts, cooperative agreements, grants, and task orders throughout the country.¹⁵⁸

In July 2016, reflecting ongoing concerns about contract accountability, the Congress passed the Foreign Aid Transparency and Accountability Act, which was seen as “a clear signal about the importance Congress places on aid measurement, accountability and transparency.”¹⁵⁹ The 2017 National Defense Authorization Act established a statutory requirement for DOD to create a global program of assessment, monitoring, and evaluation (AM&E) for security sector cooperation, provide public summaries of its evaluations, and submit an annual report to the Congress on lessons learned and best practices.¹⁶⁰ Two years later, the 2019 National Defense Authorization Act set a target for investment in AM&E for security cooperation at \$12 million.¹⁶¹ According to DOD, it

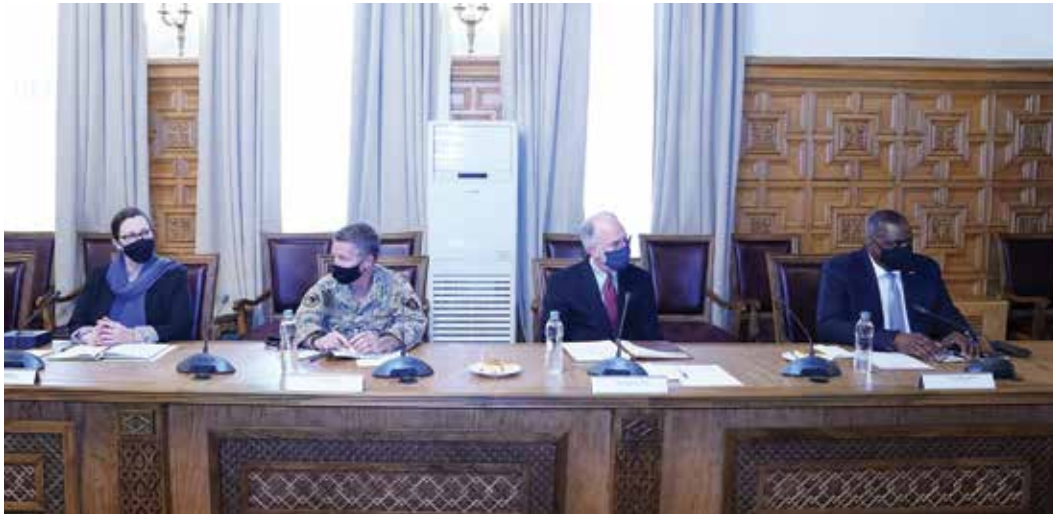
is currently implementing the global AM&E program and has spent \$20 million annually on AM&E since fiscal year 2019.¹⁶²

While most agency and implementing partner personnel would agree that M&E has value, critics believe that accountability has focused excessively on contract compliance—the result of both demands from the Congress and the prevailing corporate culture in the United States, which tries to measure everything. In the critics’ view, this deflects attention from long-term impacts and outcomes, and, in the words of former USAID Administrator Andrew Natsios, has created a kind of “Obsessive Measurement Disorder.” The result has been to foster a culture of risk aversion at government agencies, as well as misguided attempts to isolate and measure the impact of individual interventions.¹⁶³

Still, M&E remains as important as ever, as fast-moving developments in Afghanistan suggest that 2021 may actually live up to perennial predictions of a critical turning point. 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 (with a cooperation agreement signed between the United States and the Afghan government the same day); the continuation of slow, tense discussions between the Afghan government and the Taliban; and increasing levels of violence.¹⁶⁴ In the meantime, President Joseph R. Biden has announced that all U.S. military forces will withdraw by September 2021, and the devastation of the COVID-19 pandemic has increased the level and complexity of Afghanistan’s social and economic needs.¹⁶⁵

The pandemic has also affected contract oversight. For example, in September 2020, USAID/Office of Agriculture personnel based in Afghanistan reported that third-party monitoring was continuing, though at a “reduced effort.”¹⁶⁶ In some cases, movement restrictions imposed by the Afghan government precluded travel to certain project sites.¹⁶⁷ Similarly, the Combined Security Transition Command–Afghanistan reported that the pandemic had “indirectly impacted some of [the Army Corps of Engineers’] standard construction oversight processes.”¹⁶⁸ For example, oversight personnel were unable to conduct routine site visits or meet with contractors in person.¹⁶⁹ Both USAID and DOD reported that disruptions have so far been mitigated through the use of third-party monitors and substituting tele- and videoconferencing for in-person meetings.¹⁷⁰ The extensive insecurity of recent years has had the unintended effect of preparing agencies for pandemic conditions: As USAID’s Office of Education (OED) reported, “Even before the COVID-19 pandemic, OED was not allowed to do monitoring visits of . . . training sites here in Kabul and [the] provinces.”¹⁷¹

In the midst of the COVID-19 crisis, U.S. policy towards Afghanistan continues to evolve. In April 2021, President Biden announced, “While we will not stay involved in Afghanistan militarily, our diplomatic and humanitarian work will continue.”¹⁷² This decision 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



Resolute Support Commander General Austin S. Miller, Charge d'Affaires Ross Wilson, and Secretary of Defense Lloyd Austin gather to meet with Afghan President Ashraf Ghani in 2021. (Afghan government photo)

how U.S. funds will be spent in Afghanistan moving forward. Afghanistan's social and economic reconstruction needs remain significant, with the additional possibility of the return, perhaps involuntarily, of refugees from Pakistan and Iran, as well as the need for support for hundreds of thousands of internally displaced Afghans.¹⁷³ The extent of U.S. support and what form it will take is still far from clear. As U.S. agencies continue to draw down their presence, the U.S. government's ability to oversee its reconstruction activities may become even more difficult than in recent years. Preventing waste may require more creativity in using remote monitoring and other techniques for activity-level monitoring, and more rigor in strategic-level monitoring.

The U.S. government will have to consider a number of key questions, including: What level of staffing will be feasible at the embassy? Will available staff be able to adequately oversee projects? Would an eventual intra-Afghan peace agreement—even one granting the Taliban considerable governing power—actually benefit U.S. M&E efforts by reducing violence, potentially increasing third-party access to project sites? In the meantime, should the U.S. continue to fund activities which may indirectly bolster support for the Taliban in areas not controlled by the Afghan government—and, if so, based on what rationale?

Although the road ahead may be very different from the one travelled 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 M&E can be improved, is a necessary starting point for U.S. agencies attempting to help rebuild a country devastated by 40 years of war.



CHAPTER 2

USAID

USAID photo

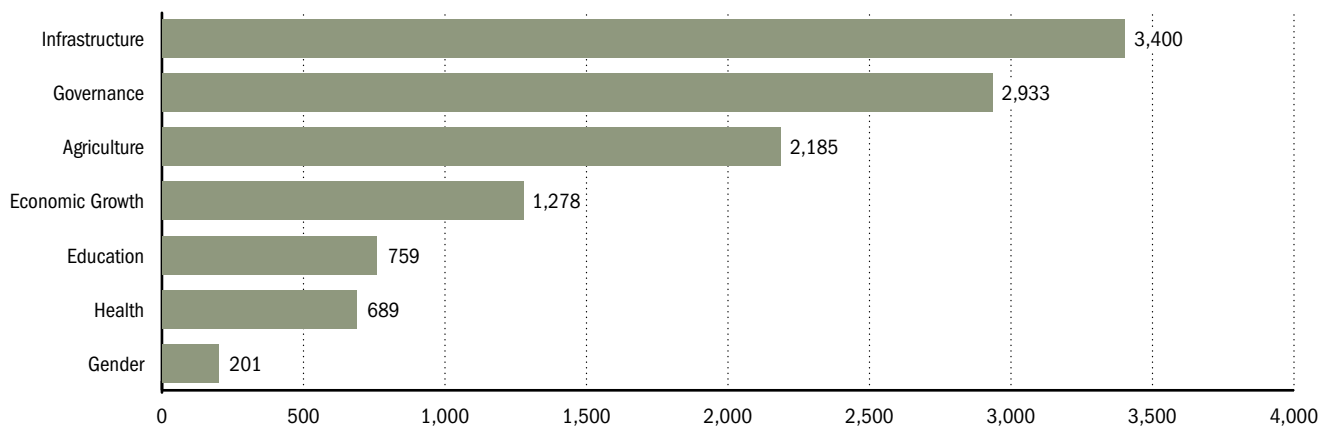
INTRODUCTION

USAID policies set a very high standard for M&E.¹⁷⁴ The agency's M&E systems are complex and highly elaborated (see Appendix B). Nevertheless, in Afghanistan, where USAID has cumulatively managed more than \$11.8 billion in contracts, grants, and cooperative agreements, the agency's actual practice of M&E often fell short of the high bar set by its policies.¹⁷⁵ Prior SIGAR work, including audits of USAID's sector-level efforts in health, education, rule of law, women's rights, and infrastructure, paints a picture of an agency struggling to effectively measure outcomes and impacts while sometimes relying on shaky data to make claims of success.¹⁷⁶

This chapter discusses USAID's use of M&E to understand how its reconstruction programs, projects, and activities were progressing, and to improve upon them. The chapter begins by summarizing the evolution of the agency's approach to M&E within the context of reconstruction in Afghanistan. It then discusses the various challenges USAID faced in deciding what and how to measure, and how to put information produced by M&E to use. Figure 5 on the following page presents USAID disbursements on contracts, grants, and cooperative agreements by sector, providing an overview of where USAID focused its spending.

FIGURE 5

USAID SPENDING ON CONTRACTS, GRANTS, AND COOPERATIVE AGREEMENTS BY SECTOR (\$ MILLIONS)



Note: This figure presents disbursements made by major USAID technical offices since 2002, disaggregated by sector. As used in the title of the figure, the phrase “contracts, grants, and cooperative agreements” includes all disbursements made to implementing partners categorized in USAID’s financial data as for-profit or not-for-profit entities. The calculations exclude disbursements made to U.S. or foreign government agencies, Afghan government entities, and public international organizations such as the World Bank. Where data fields were missing, SIGAR conducted additional research to confirm that implementing partners did not fall into one of these excluded categories. Sectors were assigned according to the USAID technical office that managed the program, project, or activity for which the funds were disbursed. For example, if USAID’s financial data indicated that a program was managed by the Office of Agriculture, disbursements made towards the program were categorized as spending on agriculture. The following technical offices were included in the calculations: the Office of Agriculture, the Office of Democracy and Governance, the Office of Education, the Office of Economic Growth, the Office of Health and Nutrition, the Office of Gender, and the Office of Infrastructure. All other USAID offices are excluded: Categorizing their programming into the sectors above would have been overly cumbersome for the purpose of this analysis. The calculations include only USAID mission-managed funds.

Source: SIGAR analysis of USAID, response to SIGAR quarterly report data call, April 13, 2021.

A Brief History of USAID’s M&E Practices in Afghanistan

In the chaotic and rapidly changing environment immediately following the U.S. invasion of Afghanistan, USAID, like other U.S. government agencies, largely focused on implementation.¹⁷⁷ The reconstruction effort commenced even as the administration of President George W. Bush struggled to clearly define its objectives in Afghanistan.¹⁷⁸ By July 2003, USAID had awarded seven major reconstruction contracts worth hundreds of millions of dollars before it had produced a complete strategy, making it difficult for the agency to be held accountable for results.¹⁷⁹

Initial M&E efforts that aimed to track USAID’s early health, education, and infrastructure programming were not overly systematic.¹⁸⁰ Regarding the quality of early project monitoring, accounts differ. One former mission director stated that USAID emphasized M&E even in the initial years of reconstruction, noting that a monitoring team consisting of Afghan staff was in place when he arrived and that there was an embassy-wide oversight effort centered around the Afghanistan Reconstruction Group.¹⁸¹ However, a June 2004 Government Accountability Office report described a chaotic situation in which overworked USAID staff members were unable to provide the precise location of projects the oversight entity asked to visit.¹⁸²

What is clear is that even in the initial years of reconstruction, the dangerous and fast-paced environment of Afghanistan militated aggressively against USAID’s institutional emphasis on M&E.¹⁸³ As security deteriorated between 2004 and 2006, the inability to

visit project sites became a growing source of frustration for senior staff.¹⁸⁴ Around that time, as the Afghanistan Reconstruction Group closed down and it became obvious that the United States was making little headway in Afghanistan, USAID's approach to M&E began to shift.¹⁸⁵ In October 2006, USAID signed a **task order** with Checchi and Company Consulting, Inc., for the Services under Program and Project Offices for Results Tracking (SUPPORT) project, the agency's first dedicated evaluation project in Afghanistan. After eight contract extensions, SUPPORT concluded in August 2012.¹⁸⁶ The successor project, SUPPORT II, also awarded to Checchi, ran from July 2012 to July 2019.¹⁸⁷

According to USAID's Automated Directive System, a **task order** is a "document that procures specific products and services through a master contract. It also sets the negotiated price at which the merchant must provide the products and services. A task/delivery order must comply with the terms and conditions in the master contract."

USAID, "Glossary of ADS Terms," April 18, 2018, p. 268.

Concern about oversight increased during the 2009 to 2012 military and civilian surge, when program funding rose significantly, emphasis on providing that funding directly to the Afghan government increased, and USAID's heavy reliance on contractors emerged as a key reconstruction theme.¹⁸⁸ Reports that developments funds were being siphoned off to the Taliban and other groups that threatened the incipient Afghan state were particularly vexing.¹⁸⁹ In response to ongoing concerns, in 2010 USAID launched its internal Accountable Assistance for Afghanistan working group, which issued a report on steps that could improve oversight and reduce corruption.¹⁹⁰ The same year, in order to better analyze data and make program recommendations to senior leadership, USAID launched the Afghan Info database (discussed below) as "the central repository for monitoring data and the principal mechanism for project managers to examine and analyze them."¹⁹¹

Concern with measuring performance and effectiveness was reinforced in 2011 by the agency's release of its new global evaluation policy, "Evaluation: Learning from Experience," issued in response to a perceived "decline in the quantity and quality of evaluation practice in the agency," and to internal and external "critiques regarding methodological quality, objectivity, access to evaluation findings, and use of evaluation recommendations for decision-making."¹⁹² The new policy was released at a time when it was increasingly difficult for USAID to properly oversee its reconstruction contracts. According to USAID's then-director of the Office of Acquisition and Assistance, the agency would have had to send nearly its entire overseas workforce to Afghanistan in order to meet the U.S. government's average ratio of dollars to contracting officers.¹⁹³

The policy was finalized at the height of the stabilization era (roughly 2009 to 2012), with its emphasis on winning hearts and minds and on measuring the Afghan population's commitment to its government. Evaluation in this environment required assessing progress toward difficult-to-measure achievements, such as changes in public attitudes or improved governance capacity. To try to capture progress on stabilization, in 2012 USAID introduced the Measuring Impacts of Stabilization Initiatives program, an ambitious venture that involved collecting, synthesizing and analyzing a wealth of data. (MISTI is discussed in detail on p. 48). Meanwhile, USAID's Stabilization Unit (which oversaw programming closely integrated with the military) used a number of contractors to evaluate individual programs.¹⁹⁴

The drawdown of U.S. troops starting in 2011 meant fewer U.S. military or civilian personnel to directly observe projects, and forced USAID to become increasingly dependent on its implementing partners and third parties for information.¹⁹⁵ Responses included multitiered monitoring, beginning in early 2014, and expanding and stepping up third-party monitoring in the middle of the following year.¹⁹⁶ Both mainly consisted of formalizing practices which had already existed in some form. Throughout the period between 2002 and 2020, new technology, such as smartphones with cameras, geo-tagging capacity, and geo-spatial mapping, provided new tools for M&E.¹⁹⁷

THE CHALLENGE: SELECTING WHAT TO MEASURE, AND HOW

Some types of **projects** and **activities** lend themselves to relatively straightforward measures.¹⁹⁸ Health care projects, for instance, can be measured in the number of clinic visits, or changes in maternal mortality rates. Yet, depending on wider strategic goals, these metrics may not be completely relevant. Whether people are going to a clinic to receive health services, for example, is an important data point at the project level, but it may not adequately capture progress towards the broader security or stability outcomes to which healthcare programming in unstable environments is intended to contribute.¹⁹⁹ In Afghanistan, the U.S. government's ultimate objectives were military and political; USAID projects were assessed using development indicators that were not always well suited to those objectives.

The following describes USAID's challenges in measuring progress, attributing impact, determining how much information to collect through its implementing partners, and developing the right mechanisms for M&E.

Defining Key USAID Terms: Projects and Activities

A **project** is a set of complementary activities with an established timeline and budget aimed at achieving a discrete development result.²⁰⁰ For example, USAID implemented the Power Transmission Expansion and Connectivity (PTEC) Project in order to increase access to electricity in Afghanistan. PTEC was a large, ambitious project that involved a suite of interrelated activities aimed at constructing energy infrastructure, commercializing Afghanistan's national electric utility, and improving the legal and regulatory environment for the private sector.²⁰¹

An **activity** is a mechanism for an intervention such as a contract, grant, or agreement that is a subcomponent of a project and that contributes to a project's purpose.²⁰² For example, one activity within PTEC involved improving the operations and maintenance capabilities of Afghanistan's power utility.²⁰³

This report will differentiate between these terms where appropriate—for instance, to accurately refer to a specific project or activity or to preserve the integrity of a quote from an interviewee. However, parts of this report may use these terms interchangeably, as “activities” are often, in common terms, understood to be “projects.”



An Afghan farmer harvests wheat in 2009. In order to curtail opium poppy cultivation, USAID implemented programs that provided farmers with alternative, licit crops. However, the metrics USAID used to measure the performance of these programs were sometimes flawed. (USAID photo)

Finding the Right Indicators to Measure Progress

Measuring such concepts as governance, stabilization, and capacity building is challenging: Not only are the concepts abstract, but there are difficulties in how to define them, what progress looks like, and what metrics should be used—questions made even more difficult in a complex conflict environment. A 2017 USAID-funded review of research on Afghanistan stabilization found “nearly 200 different indicators used in various combinations to measure and track implicit or explicit definitions of stabilization.”²⁰⁴ Without any clear definition of what “stability” consisted of, projects such as Stability in Key Areas (which aimed to reduce the impact of the insurgency and increase confidence in the Afghan government) relied on public *perceptions* of stability or of Afghan citizens’ faith and confidence in their government, collected via public opinion surveys.²⁰⁵ (The following highlight discusses the strengths and weakness of public opinion polling in Afghanistan.)

Sustainability became more of a program focus after the 2009 announcement of the planned 2014 drawdown of troops—but, as with stabilization, it was a difficult concept to measure or even define. Did it refer to a project’s fiscal sustainability, for instance, the organizational stability of a government ministry, or the economic sustainability of a particular agricultural initiative or even the entire economy?²⁰⁶ In some cases, evaluators tried to retrofit analyses, driven by the changing political needs and the reporting demands of the Congress and other institutions. For example, one retrospective meta-evaluation of 35 USAID projects conducted between 2010 and 2015 was intended to assess how the agency used findings on the efficiency and sustainability of its projects—when the original evaluations did not always include measurements of efficiency

PUBLIC OPINION POLLING WAS A KEY SOURCE OF INFORMATION, BUT DATA WERE SOMETIMES QUESTIONABLE

As the concepts of counterinsurgency and stabilization began to influence the logic of development projects, the views of the Afghan population were considered important measures of whether the government and its international partners were winning the battle for hearts and minds. The main method for ascertaining this was public opinion polling. Although many of the polls were not commissioned specifically for one project, civilian and military agencies used their results to gauge the success of programs, especially in governance, as perceptions of government effectiveness and legitimacy were selected as key indicators of success. Some of the polls were available to the public. Others, especially those conducted for the military, remained restricted.²⁰⁷

The most visible poll was The Asia Foundation's Survey of the Afghan People, carried out first in 2004 and annually from 2006.²⁰⁸ It is the longest-running nationally representative survey of the attitudes and opinions of Afghan adults, with over 17,800 respondents in 2019.²⁰⁹ Trends in responses over time, such as perceptions of the Afghan National Police, satisfaction with Afghan government performance, and opinions about whether the country was moving in the right direction were viewed as useful gauges of reconstruction progress.²¹⁰

Still, the Asia Foundation's yearly survey was not without flaws. A data quality assessment of the 2006 to 2009 surveys carried out for the UK Department for International Development (DFID) revealed a number of concerns surrounding the reliability of responses.²¹¹ The assessments observed that there were more problems with accuracy in insecure areas, and in some areas it was not possible to conduct interviews at all.²¹² Moreover, respondents were more likely to be better-off Afghans, raising questions about the representativeness of the results.²¹³ There were other sources of potential error: The assessment cautioned that social desirability bias (the tendency of respondents to say what they think will be viewed favorably by others) "contaminat[e]" the Asia Foundation survey as well as other public opinion polls, especially responses relating to contentious issues like democracy, justice, gender, and the quality of the Afghan government.²¹⁴ Even so, "with nuanced interpretations," the DFID study concluded, "much of the trend information remains usable."²¹⁵

More broadly, critics have questioned whether polling is even feasible in an environment like Afghanistan. There is little private space that would allow respondents to answer frankly; there are incentives to exaggerate, minimize, or even lie, depending on the situation (after 2001, many Afghans associated surveys with aid projects); and Western concepts are implied in many of the questions.²¹⁶ In rural areas, there is frequently a lack of familiarity with percentages or scales, which could affect respondents' understanding of survey questions written by urban and educated persons, as well as "drive-by polling"—surveys conducted rapidly by enumerators who would barely get out of their vehicles to pose questions.²¹⁷

Polls also struggled to navigate gender. Cultural restrictions on the mobility of women and their interactions with strangers limited the number of female enumerators, and therefore the extent to



A surveyor interviews an elder in Helmand Province. (USAID photo)

which surveys could reflect female views.²¹⁸ According to one report, “some informed respondents were skeptical about the extent to which research incorporated a true gender analysis or lens, rather than simply aiming for female enumerators and a 50 percent female sample in order to show that gender had been taken into account. This was characterized as “add women and stir.”²¹⁹

Much of the Afghan population was skeptical about these polls. As one report noted, “There is widespread suspicion, well founded or not, that polls and surveys reflect the positions of the donors who sponsor them.”²²⁰ Reflecting such skepticism, a 2010 ABC/BBC/ARD poll was questioned after it showed what was considered an unlikely increase in positive attitudes towards international forces and then-President Hamid Karzai.²²¹

The thirst for polling data and other monitoring information produced a cottage industry of international and Afghan firms of variable reliability and quality.²²² In the years leading up to 2012, the number of commercial polling firms increased from three to 30. One report noted:

Staff from private research firms readily acknowledged that research fraud was extensive, although they typically ascribed it to other firms or to their own firms in the distant past, and described their own protocols and procedures for avoiding fraud. A recurrent explanation offered for this was that some of the owners of research (contractors) didn’t care about accuracy and therefore turned a blind eye.²²³

A report attributed the problem in part to the commercial focus of the firms—that it was “more about generating contracts than knowledge.”²²⁴ In other words, survey firms believed, correctly or not, that responding to donors’ appetite for good news would be rewarded with future survey work.

Although public opinion polling often proved to be a fraught exercise in Afghanistan, it is unreasonable to dismiss the method entirely—particularly when local perceptions may be relevant to understanding program effectiveness. Even though prominent surveys like The Asia Foundation’s suffer from certain flaws, they can still prove useful in assessing changes in perceptions over time, provided that contextual limitations (such as less visibility into insecure areas) are fully acknowledged and a system of checks and balances (such as data quality assessments) is in place.



Carpets from the Kabul Carpet Export Center are displayed for sale in Charleston, South Carolina, in the summer of 2020. (USAID photo)

and sustainability.²²⁵ The report's first recommendation was that the concepts of sustainability, effectiveness, and efficiency be standardized.²²⁶

What was easiest to measure did not always equate to what mattered most: The number of schools built, for example, did not necessarily indicate improving literacy rates.²²⁷ In a 2015 strategic-level **performance management plan**, one of the indicators for increasing regional trade was “person hours of training completed in trade and investment enabling environment”—an easily quantifiable output rather than a more meaningful (but also perhaps more difficult to measure) outcome.²²⁸ Reliable numbers were hard to come by even on concrete indicators such as job creation and employment rates, and differences in definitions made interpretation difficult.²²⁹ At times, USAID and its implementing partners came up with creative calculations. One agricultural program, for instance, assumed that one hectare of cultivated land created .17 days of employment.²³⁰ Similarly,

Defining Key USAID Terms: Performance Management Plan

The **performance management plan** is “a tool used by a USAID Mission/Office and assistance objective team to plan and manage the process of assessing and reporting progress towards achieving an assistance objective.”²³¹ Although the term “PMP” is also sometimes casually used to refer to project-level (or activity-level) M&E plans, according to the Automated Directives System, the official term for that is “project MEL plan,” or Project Monitoring, Evaluation, and Learning Plan, which has a similar function to the mission-level PMP but is focused on an individual project.²³²

Adding to the confusion, within State’s M&E system, the term “performance management plan” refers to “programs, not strategies,” making them functionally equivalent to USAID’s project MEL plans.²³³ By contrast, USAID performance management plans are explicitly intended to monitor “strategic progress.”²³⁴

a carpet export project came up with a numerical estimate of the amount of employment created with every square meter of carpet shipped.²³⁵

A common assumption was that such quantitative indicators were inherently more rigorous and therefore more desirable than qualitative ones.²³⁶ But that was not a universally held opinion. As a former senior USAID official stated, “The utility of quantitative M&E practices is overstated; these practices often erroneously link correlation with causality and reach ridiculous conclusions.”²³⁷ The official’s critique mirrored broader misuse of quantitative data in Afghanistan, which were sometimes constructed by assigning ordinal numbers to qualitative responses (for example, asking for an opinion on a scale of one to 10). Doing so not only conveyed a false sense of precision, but also resulted in nonsensical results when additional arithmetic (such as averaging) was performed on the ordinal numbers.²³⁸ But the fact that some users did not employ quantitative data in informed ways did not make it inherently useless. SIGAR itself relies on certain quantitative indicators—such as electric power generation output, quantitatively summarized Asia Foundation survey data, and violence levels—to assess the effectiveness of reconstruction.²³⁹ The key was to understand how the data was produced, to avoid overestimating the degree of point-estimate precision, and to appreciate its limitations.

But even a careful approach to interpretation could not compensate for irrelevant indicators. In some cases, USAID and its implementing partners resorted to proxy metrics that were tangential to the main goals of a project—and in so doing missed an opportunity to use quantitative data to objectively monitor results. For example, the Kandahar Food Zone project, which was intended to reduce opium poppy cultivation in the province, measured progress through 27 indicators, including kilometers of irrigation canals and drainage ditches rehabilitated, and the amount of land with increased high-value crop production.²⁴⁰ However, it did not measure the change in the amount of land devoted to opium-poppy cultivation—a reduction of which was the project’s purpose.²⁴¹ Although better irrigation and cultivation of higher-value crops can theoretically contribute to a reduction in opium-poppy cultivation, improved irrigation can also increase it.²⁴²

Similarly, a 2012 USAID Office of Inspector General audit criticized a five-year, \$150 million program, Incentives Driving Economic Alternatives for the North, East, and West (IDEA-NEW), which aimed to dissuade Afghans from growing opium poppies, for a number of M&E shortcomings.²⁴³ Notably, the program dropped from its performance management plan any reference to indicators that were directly related to reducing opium poppy cultivation. The indicators, which tracked assistance provided to individuals and communities affected by opium poppy eradication efforts, were reportedly dropped because USAID had shifted the program’s emphasis from eradicating illegal crops to encouraging legal crops.²⁴⁴ The audit concluded that omitting information about counternarcotics efforts deprived the mission and its implementing partners of vital information they could have used—such as whether the program might have had the unintended effect of actually increasing poppy cultivation in some areas.²⁴⁵

The central critique was that the revised performance management plan lacked any performance indicators related to opium poppy, despite the program's aim to reduce opium poppy cultivation.²⁴⁶

This conclusion was reinforced by a subsequent evaluation of project activities in Nangarhar Province, requested and funded by IDEA-NEW. Evaluators found that dropping the counternarcotics indicators not only took away information that USAID and the implementing partner could have used to make programming decisions and provide evidence of opium poppy reduction, but—as the previous audit had noted—reduced their awareness of whether interventions might even have contributed to an increase in production.²⁴⁷ The evaluation also noted that monitoring reports referred to opium poppy only briefly, and that the program's final survey did not cover districts in Nangarhar that had seen increasing poppy cultivation during the life of the program. All of this left unanswered the basic question of how effective the program had been in the areas where cultivation and the insurgency were most active.²⁴⁸

“There is a ginormous gap between what happens at the activity level and mission-level indicators.”

—Evaluation expert

If the challenges of measuring individual projects' outcomes were serious, the next step was even harder: “rolling up” to the sector or country level—meaning, consolidating data on individual projects that are part of a broader portfolio and trying to understand whether the projects are having a cumulative impact that contributes to sector- or country-level objectives. Some observers were skeptical that it was possible to do this in a meaningful way.²⁴⁹ USAID's mechanism for such data consolidation was the strategic-level performance management plan.²⁵⁰ But that required combining disparate types of data from various projects into a meaningful big-picture view—a very difficult task. The problem, said one evaluation expert, is that “there is a ginormous gap between what happens at the activity level and mission-level indicators. All the action is at the activity level, where little or nothing neatly rolls up into an activity theory of change.”²⁵¹ A former USAID official added, “the Holy Grail is finding a way to portray impact beyond the project level.”²⁵²

But pressure to do just that came from the standard foreign assistance indicators, which were intended to provide a standard way for all U.S. government agencies to measure the collective impact of their development work.²⁵³ Standardization meant that indicator values could be aggregated across different countries, in order to “measure and illustrate what foreign assistance accomplishes.”²⁵⁴ The standard indicators were widely regarded as often not very meaningful, the paperwork was labor intensive, and the benefit was unclear.²⁵⁵ In particular, many of the indicators measured input or output data that contributed little to knowledge about a program's impact on foreign assistance objectives—for example, metrics like the “number of individuals who complete [U.S.



An Afghan man works at a milk processing center funded through USAID's IDEA-NEW program in 2012. (USAID photo)

government]-assisted workforce development programs” and “the number of people in a host country who received [U.S. government]-funded counternarcotics training.”²⁵⁶ Questioning both the methodology and level of rigor required to accurately aggregate such metrics, one USAID contractor told SIGAR, “Thinking of standard indicators gives me hives.”²⁵⁷

Selecting the Right Goals to Measure Against: Does USAID Assess Progress Towards the Desired End State of a Stable Afghanistan?

With USAID's development work in Afghanistan occurring within the context of an ongoing war, development and security were often inextricably linked.²⁵⁸ Although USAID officials and implementing partners were frequently motivated by the desire to improve the lives of Afghans, such sentiments did not always comfortably coexist with the *raison d'être* for the U.S. presence in Afghanistan: to prevent a reoccurrence of 9/11, a goal that was grounded inherently in stabilizing Afghanistan.

But the concept of “stability” defied easy definition. As a programmatic category, “stabilization” projects were generally thought of as transitional, shorter-term initiatives intended to help “legitimate” authorities “peaceably manage conflict and prevent a resurgence of violence,” in the words of a 2018 U.S. government definition.²⁵⁹ Yet this conception of stability was frequently belied by the language of USAID's strategies for Afghanistan, based on the belief that stability was also the result of *longer-term* development. According to the “U.S. Foreign Assistance for Afghanistan Post Performance Management Plan (2010–2015),” security and stability goals were woven into initiatives as diverse as public health, governance, education, and agriculture.

These goals included building active support for the government (health), supporting and reinforcing efforts to improve security in Afghanistan (governance), increasing the legitimacy of the Afghan government (education), increasing confidence in the Afghan government and “undercutting the appeal of the insurgency to potential recruits by offering economic alternatives and providing stability to communities that are on the frontlines of the war” (agriculture).²⁶⁰

Although the specific language shifted in subsequent strategies, the general sentiment remained. USAID’s “Plan for Transition,” which guided the agency’s efforts from 2015 to 2018, was based on the theory that USAID-funded development projects would help the Afghan government better deliver key services, which, “along with security and the equitable rule of law, will ultimately generate increased confidence in the legitimacy and effectiveness of the government and, in turn, will foster stability.”²⁶¹

However, empirical evidence often clashed with the theory that development programming could support stability. In 2015 and 2016, SIGAR reported on the work of the Measuring Impacts of Stabilization Initiatives program, a three-year, \$19 million effort to measure and map changes in stability over time as well as evaluate the impact of USAID stabilization programs in key areas of Afghanistan.²⁶² MISTI’s findings raised worrying questions. The MISTI program reported, for example, that villages receiving USAID stability projects actually scored lower on stability—an aggregate measure of whether the projects strengthened perceptions of good governance and effective service delivery—than similar villages that received no such assistance. Some villages reportedly under Taliban control that received USAID stability projects subsequently showed *greater* pro-Taliban support.²⁶³

As SIGAR reported, the findings pointed to a general weakness in the theory behind using economic tools to achieve stability. If the United States aims to increase stability as part of a whole-of-government effort to defeat an extremist threat, SIGAR said, project impacts must be assessed against these political outcomes.²⁶⁴ Still, USAID seemed to be largely indifferent to the implications of the MISTI findings, and made little effort to assess political outcomes for non-stabilization programs.²⁶⁵

By 2016, USAID had gotten out of the business of stabilization programming.²⁶⁶ In 2018, SIGAR concluded that stabilization efforts appeared to have failed to yield intended outcomes.²⁶⁷ Yet stability objectives remained highly relevant, leaving lingering questions about whether USAID projects should be assessed against stability or against the development objectives preferred by USAID. Such questions invited strategic contradiction. For example, USAID’s current strategy, which will remain in effect until 2023, underscored that USAID was “shifting away from quick-impact stabilization programs to focus on longer-term, broad-based development efforts.”²⁶⁸ Yet in the same document, USAID said that all of its development objectives “are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address the security and development challenges that have made Afghanistan a safe



A midwife presents an illustration of healthy food for pregnant women. (USAID photo)

haven for terrorists and violent extremist organizations and have driven the conflict with the Taliban.”²⁶⁹ Stability, in other words, remains a core aspect of USAID’s intervention logic.

In the end, USAID never resolved the tensions inherent in implementing development projects in the middle of ongoing conflict. In fact, via semantical maneuvering, it has managed to sidestep the inconvenient challenge of measuring the extent to which individual development projects, or portfolios of those projects, are achieving stability effects. USAID policy defines a development objective as “the most ambitious result that a mission, together with its development partners can contribute through its interventions.”²⁷⁰ Within USAID’s strategic framework, development objectives contribute to a strategic goal.²⁷¹ However, per its policy, USAID “is not solely accountable for achieving the [country-level strategic] goal” because the goal is theoretically also advanced by a range of other actors, such as multilateral organizations and partner-country governments.²⁷² Consequently, performance indicators for the goal are not required.²⁷³ As a result, USAID is presently able to claim that all of its development objectives in Afghanistan contribute to stability without actually having to assess, via performance monitoring, whether that is in fact true.²⁷⁴

TO WHAT EXTENT IS USAID’S STRATEGY GROUNDED IN EVIDENCE?

USAID policy requires each mission to integrate evidence into strategic planning. Such evidence should provide rigorous justification for USAID’s selected development approach in a country.²⁷⁵ A mission’s best understanding of the causal relationships underlying its strategic approach is supposed to be presented in a Country Development Cooperation Strategy (CDCS).²⁷⁶ Evidence that supports perceived causality between development interventions and the desired outcomes or impacts articulated by the strategy is particularly important, as evidence-based relationships form the foundation of effective programming.²⁷⁷

Conversely, presumed causal relationships based on weak evidence create the risk of doing the wrong thing perfectly—even if programming successfully produces intended outputs. For example, the assumption that providing jobs to young unemployed males will cause an increase in stability by keeping young men from joining an insurgency may sound like a good reason to create jobs: Increasing employment should combat violent extremism. But if the link between unemployment and insurgent recruitment is tenuous—as some studies suggest—even effective job creation may serve no stability-related purpose.²⁷⁸ In this case, an M&E system may detect effective job creation, but go no further to test whether the job creation had an effect on counterinsurgency or stabilization outcomes. In short, a fragile link between project-level objectives and strategic goals can render M&E less effective.

To test whether causal statements made in USAID’s current CDCS for fiscal years 2019 through 2023 had sufficient evidentiary support, we closely examined the strategy’s footnotes. Of the 136 footnotes in the strategy, we identified 21 instances where a footnote appeared to support a sentence that outlined or touched on a causal relationship or assumption. USAID generally used the remaining 115 footnotes to support facts. Of the 21 footnotes associated with causal statements or assumptions, we selected the 11 that seemed to have the greatest implications for programmatic decisions—in other words,

FIGURE 6

COUNTRY DEVELOPMENT COOPERATION STRATEGY EVIDENCE GRADING SCALE

A	The evidence provides full support for the statement.
B	The evidence provides some support for the statement, or provides full support for the statement but without transparency.
C	The evidence provides limited support for the statement.
D	The evidence provides limited and flawed support for the statement.
F	The evidence provides no support for, is unrelated to, or contradicts the statement.

for how and why funds would actually be spent to achieve objectives. For these 11 footnotes, we reviewed the cited sources in detail and assigned each source a grade based on how well it appeared to support the associated causal statement in the strategy using the grading scale shown in figure 6 on the previous page.

This highlight presents the grades for some of the 11 sources in the table below (for a full version of the table with all 11 sources, see Appendix C). While there are examples of strong sources that support their associated statement in the strategy, there are also cases where the strategy falls short of USAID's policy regarding the use of rigorous evidence to support causal hypotheses and assumptions. A number of key issues common to multiple sources emerged, and can be broadly grouped into the following categories:

- The source provides weak and flawed support for the statement.
- The source only identifies a correlation while the statement makes a causal assertion.
- The source makes claims without supporting evidence.
- The source contradicts other assertions or evidence presented in the strategy.
- The source is too general and may not apply to Afghanistan.

Table 2, on the next page, presents a summary of our grades. Examples of a source falling into each category are discussed in detail below the table.

Weak and Flawed Support

The Afghanistan Country Development Cooperation Strategy claims on page 13 that “the struggle to find a job can lead to economic marginalization and instability, which, in turn, makes one susceptible to violent extremism,” and cites a 2017 Brookings Institution report titled: “How do Education and Unemployment Affect Support for Violent Extremism?”²⁷⁹

Overall, this source provides weak and flawed support for the statement:

- The source's own literature review acknowledges that “previous empirical work has failed to demonstrate any link between unemployment and radicalization.”²⁸⁰
- The source's thesis was that relative deprivation, or the absence of opportunities relative to expectations (specifically, the lack of jobs for educated people in the labor market), is one driver of support for violent extremism.²⁸¹ The problem of educated workers being unable to find jobs is distinct from the problem of poverty and instability experienced by unemployed low-skilled workers with little education. Although the economic marginalization experienced by the very poor could also be linked to increased support for violent extremism, these two narratives require different interventions.
- The conclusions from the source may not apply well to Afghanistan, since the source examines a collection of Middle Eastern and North African countries that does not include Afghanistan, and its literature review finds high variability in results across those countries.²⁸²
- One of the regression results in the source actually contradicts its initial hypothesis, calling into question the robustness of the findings.²⁸³
- Even for the statistically significant regression results that support the initial hypothesis, the magnitude of the results is relatively small.²⁸⁴ If one scales down the magnitude further to adjust for the fact that *claiming* to support violent extremism is a weaker threshold than *engaging* in violence, the results may not hold any practical significance.

TABLE 2

COUNTRY DEVELOPMENT COOPERATION STRATEGY EVIDENCE GRADES (SELECTED)

Sentence from the CDCS	Grade	Source	Evidence Cited
"The struggle to find a job can lead to economic marginalization and instability, which, in turn, makes one susceptible to violent extremism" (p. 13).	D	Brookings Institution, "How do Education and Unemployment Affect Support for Violent Extremism?," 2017.	The source document provides some evidence that relative deprivation (the absence of opportunities relative to expectations) in the job market can drive radicalization (p. 3).
"Another driver of extremism among youth is poor social status; evidence shows that improving young Afghans' sense of respect in their communities appears to lower their risk of supporting violent groups and causes" (p. 13).	D	Mercy Corps, "Does Youth Employment Build Stability?," 2015.	The source document states, "Higher respect within the community was the only social factor found to be positively associated with a decrease in support for political violence" (p. 21).
"All development objectives are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address key drivers of conflict that have enabled the Taliban to make gains, including corruption, unemployment, and a lack of government legitimacy. Achieving these objectives will help create the conditions necessary for peace" (p. 26).	B	Stephen J. Hadley, Andrew Wilder, and Scott Worden, "Four steps to winning peace in Afghanistan," <i>Washington Post</i> , 2017.	Policy experts specializing in Afghanistan make this claim directly (p. 2).
"Economic growth is constrained by crippling government bureaucracy and trade policies, poor infrastructure, and infant industries that lack the enabling environment to grow" (p. 28).	A	SIGAR, <i>Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan</i> , 2018.	Research supports each of these conclusions (pp. 3, 165–166).
"Corruption also severely impedes economic growth and development progress, affecting the credibility of the government and fueling widespread resentment" (p. 52).	C	Integrity Watch Afghanistan, "National Corruption Survey," 2016.	The source states, "Corruption also retards investment, making it a significant drag on the economy," (p. 1) and that "more than 70% of respondents said that corruption is now worse than it was two years ago" (p. 1).

Source: SIGAR analysis of USAID, "Afghanistan Country Development Cooperation Strategy FY 2019–2023," September, 2018, pp. 13, 26, 28, 52; SIGAR analysis of Katika Bhatia and Hafez Ghanem, "How do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries," *Global Economy and Development Working Paper 102*, Brookings Institution, March 22, 2017, pp. 1, 3, 5; SIGAR analysis of Mercy Corps, "Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan," January 2015, pp. 20–21, 23, 26; SIGAR analysis of Stephen J. Hadley, Andrew Wilder, and Scott Worden, "Four steps to winning peace in Afghanistan," *Washington Post*, April 27, 2017, p. 2; SIGAR analysis of SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, pp. 3, 165–166; SIGAR analysis of Integrity Watch Afghanistan, "National Corruption Survey," December 2016, p. 1.

Correlation, Not Causation

In the very next sentence, the Country Development Cooperation Strategy claims that "another driver of extremism among youth is poor social status; evidence shows that improving young Afghans' sense of respect in their communities appears to lower their risk of supporting violent groups and causes." It cites a Mercy Corps impact evaluation of a vocational training program in Afghanistan titled: "Does Youth Employment Build Stability?"²⁸⁵ While the statement in the strategy implies a causal relationship, the source makes no attempt to support such a claim. The source did not conduct a rigorous enough analysis to support any causal hypothesis, and is careful instead only to suggest a correlation, simply concluding that, generally, greater respect for youth within the community was "positively associated" with a decrease in support for political violence.²⁸⁶ Additionally, the authors of the source were surprised to find that some markers of social status, such as having more social connections or higher personal confidence, were in fact associated with *greater* support for political violence.²⁸⁷ In selectively highlighting the one conclusion about respect within the community and misrepresenting the relationship as causal, the strategy risks influencing USAID toward interventions that may be ineffective or could potentially even compound the problem.

Contradictory Context	Further flaws
However, the source also says that “previous empirical work has failed to demonstrate any link between unemployment and radicalization” (p. 1), and that “there is also no consensus on the relationship between education and violent extremism” (p. 3). Furthermore, the data do not include Afghanistan (p. 5).	The Mercy Corps source, which is also used to support the CDCS, states that “being employed was not significantly associated with either a willingness to engage in or support violence against the state,” and that “[more] business connections significantly increased the probability of being willing to use violence to fight an unfair law or state decision” (p. 20).
The cited sentence is cherry-picked from a paragraph that states, “The analysis finds little evidence that social outcomes can decrease support for political violence” (p. 21). The study also states, “personal confidence, number of friends, and identifying as an Afghan were found to be highly significant and perversely associated with propensity towards political violence” (p. 23). The analysis makes no claims of causation, restricting itself to noting correlations (p. 21).	N/A
Apart from their own status as “experts” who held meetings with various stakeholders, no outside evidence is presented to support the claims made.	According to the Mercy Corps source, “No relationship was found between confidence in Afghan institutions – including the national, provincial, and local government – and [respondents’] propensity towards political violence” (p. 26).
N/A	N/A
Polling data is presented out of context, and there is no supporting data for a causal link between corruption and investment (despite the fact that such a link is intuitively plausible).	N/A

Lack of Transparency

There are some examples where the cited evidence strongly supports the statements in the Country Development Cooperation Strategy, but the cited source does not itself present evidence demonstrating that its assertions are empirically accurate. One such example is the statement on page 26 of the strategy: “All [development objectives] are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address key drivers of conflict that have enabled the Taliban to make gains, including corruption, unemployment, and a lack of government legitimacy. Achieving these objectives will help create the conditions necessary for peace.”²⁸⁸ The source is a 2017 op-ed from the *Washington Post* authored by prominent policy experts with knowledge of Afghanistan, who assert that their conclusions came from discussions with senior officials and civil society and business leaders in Afghanistan.²⁸⁹ While the authors have credibility on this topic, the article alone does not provide any of the evidence upon which their statements are based, and the reader therefore has no ability to review or evaluate that evidence independently.



Participants in a tailor training program practice sewing techniques at a bazaar in Kajaki District, Helmand Province, in 2014. (USAID photo)

Contradiction with Other Statements

The Mercy Corps source not only provides weak support for its associated statement in the Country Development Cooperation Strategy, but also directly contradicts other statements elsewhere in the strategy. For example, the Brookings Institution source was used to suggest the causal relationship between unemployment and radicalization, a relationship referred to multiple times throughout the strategy.²⁹⁰ However, the Mercy Corps source noted in its conclusion that, “taken together, these results suggest that young people’s current economic circumstances are not a major driver of propensity towards political violence within the context of southern Afghanistan.”²⁹¹

It goes on to caution against assuming that increased youth employment will increase stability, and specifically recommends that implementing agencies “decouple employment and stabilization interventions.”²⁹² The Mercy Corps source did allow for the possibility of a valid relationship between long-term unemployment and violence by noting that “economic optimism was found to be significantly related to a lesser acceptance of the use of violence,” and extrapolating that “improving youth’s perceptions of their future economic prospects may hold potential to decrease their support for or participation in violent movements.”²⁹³ However, Mercy Corps also went on to explicitly recommend that additional empirical research is “needed to fully assess the roles of long-term employment and improved economic conditions on political violence.”²⁹⁴ In short, the drivers of political violence are far more complex than employment, and employment alone does not appear to have a significant impact on a person’s likelihood to engage in violence.

Another example of a claim with contradictory evidence cited in the Country Development Cooperation Strategy is that low government accountability and legitimacy is a driver of conflict.²⁹⁵ The cited version of this claim points to a similar assertion in the *Washington Post* source, but the Mercy Corps source once again raises questions about this hypothesis by finding “no relationship” between “confidence in Afghan institutions” and “propensity towards political violence.”²⁹⁶ The hypothesis that government legitimacy is related to conflict is not necessarily unreasonable. But evidence that perceptions of government performance are *uncorrelated* with a proclivity to commit violence calls into question the approach of pursuing higher government accountability for the purpose of increasing stability through violence reduction—a common element of U.S. strategy. These examples demonstrate how the CDCS may provide evidence in some footnotes that appear to contradict causal assumptions made elsewhere in the document.

Too General

The CDCS states on page 29, “Increasing international trade and connectivity will accelerate private sector-driven and export-led economic growth because it provides business access to more inputs and information and technology, increasing their efficiency and productivity.” It cites a webpage from the World Trade Organization (WTO) promoting the benefits of free trade.²⁹⁷ While there is an intuitive link between trade and growth, and while there may be countries for which this statement is completely accurate, this source does not address any specific country and it is too general to justify promoting free trade in Afghanistan as a strategy to stimulate economic growth. The source even includes numerous caveats that discourage blindly applying its content to specific countries without further analysis:

- First, the source openly acknowledges the complex relationship between trade and jobs, noting that opening to trade will inevitably cost some workers their jobs.²⁹⁸
- Second, critically, the source mentions that it is often more difficult for developing countries than for developed countries to open their markets.²⁹⁹
- Third, the source specifically warns that the impact of trade on employment “cannot be assessed in a vacuum,” noting that “the impact of competition from foreign producers varies across firms in a sector, across sectors of the economy as well as across countries.”³⁰⁰
- Fourth, the source implies that opening to trade without parallel government social programs to help workers through that transition may prevent a country from maximizing the benefits associated with trade.³⁰¹

Overall, of the 11 sources that we reviewed and graded, only two received an A for fully supporting the statement referenced in the CDCS. If the footnotes are an accurate representation of the total evidence base used to support the mission’s development hypotheses and proposed theories of change, the mission’s implementation of the Afghanistan CDCS fell short of USAID’s policy that “a CDCS must be grounded in evidence and analysis.”³⁰²

Looking for Impact: Success Stories and Overattribution Crowded Out Rigorous Measurement

Given the enormous pressure to show success and contribute to the overall narrative of progress, USAID and its implementing partners displayed a tendency to overattribute impact to their activities. What was labeled “outcome” or “impact” had often been upgraded from the more pedestrian “activities” or “outputs.”³⁰³ For example, in response to a SIGAR query on USAID’s most successful reconstruction programming, USAID said it had built more than 600 schools, funded teacher training, and developed university teaching degree programs (among other claimed achievements).³⁰⁴ These were not outcome metrics. Yet, USAID claimed to be highlighting “assistance programs that have contributed to measurable positive impacts on Afghanistan’s development and stability.”³⁰⁵

The reality that Afghanistan could not be quickly transformed clashed with the political hunger for quick successes.³⁰⁶ As one USAID contractor observed, “Everyone wants big results in a short time frame, but programs are trying to change social attitudes, which takes a long time. . . . [You] won’t change society in four years.”³⁰⁷ These dynamics created incentives for implementing partner and USAID officials to overattribute positive outcomes, ignore negative observations, or err on the side of reporting success.³⁰⁸ At times, the problem extended from the implementation level all the way up reporting chains to the Congress, where, as Inspector General John Sopko has expressed, agency officials seemed more eager to provide good news stories than candor.³⁰⁹ In some cases, inaccurate and incomplete information was published for public and official use.³¹⁰

“Everyone wants big results in a short time frame, but programs are trying to change social attitudes, which takes a long time. . . . [You] won’t change society in four years.”

—USAID contractor

The best-known examples were in the education and health sectors.³¹¹ For example, lacking other illustrative high-level indicators of progress, USAID consistently pointed to the number of Afghan children enrolled in school as a significant success story.³¹² Even today, the agency claims that “over 9 million children are enrolled in school (including over 3.5 million girls).”³¹³ However, as SIGAR continues to report, these figures are misleading. Afghanistan’s Ministry of Education counts students who have been absent for up to three years as enrolled because, it says, they might eventually return to school. This means attendance rates are far lower than enrollment rates. In December 2016, then-Minister of Education Assadullah Hanif Balkhi said that after adjusting school records to deduct registered but permanently absent students, only 6 million students were actually attending classes in Afghanistan—2.4 million fewer students than were enrolled at the time.³¹⁴

Similarly, in the health sector, USAID did not disclose limitations in the data behind achievements claimed in increased life expectancy and in decreased child, infant, and maternal mortality. For instance, USAID’s public documents cited a huge decrease in maternal deaths from 1,600 to 327 per 100,000 live births between 2002 and 2010.

SIGAR Has Audited Agency Efforts in the Health and Education Sectors

SIGAR's audits of U.S. efforts to improve Afghanistan's education and health sectors inform the analysis above. Like much of SIGAR's work, the audits showed that claims of success made by USAID and others were sometimes based on misleading or anecdotal data, a recurring theme of reconstruction that is reinforced by this report.³²⁹

SIGAR's review of USAID's data found that the 2002 information was based on a survey conducted in only four out of 360 districts. USAID's internal documentation acknowledged the limitations, but they were not shared with the public. USAID funded a new health survey in 2015, and the agency now draws its health data from this survey.³¹⁵

At the funding level, USAID has to respond to its political masters in the White House and the Congress, and to respond to the perennial skepticism among the American public about the efficacy of U.S. foreign aid in general.³¹⁶ This need to “feed the beast” resulted in an overreliance on success stories in lieu of more rigorous measurement.³¹⁷ For their part, implementing partners naturally feel the need to show their work is effective in order to secure additional funding.³¹⁸

The perverse incentives to report positive results have led some critics to refer to aid organizations as part of a “success cartel” because they “believe their survival relies on convincing their political leadership that taxpayer funds are making an outsized difference and fulfilling foreign policy objectives.”³¹⁹ In this “do-or-die market environment,” explained one such critic, “development contractors . . . cannot afford to report poor results.”³²⁰ A 2009 global study on M&E in U.S. government agencies concluded that “bureaucratic incentives do not support rigorous evaluation or use of findings.”³²¹ Two-thirds of a group of external evaluators polled for the study “agreed or strongly agreed that USAID cares more about success stories than careful evaluation.”³²² One evaluation expert stated that some people were hired for the sole purpose of writing success stories.³²³

Especially in the area of stabilization, there emerged some “darling districts” (such as Nawa) that were held up as symbols of success.³²⁴ But as one former USAID staffer noted, “anecdotal evidence and stories crowded out more sober, systematic evaluation efforts.”³²⁵ Anecdotes of local or individual personalities' success were generalized to a national level because of organizational incentives to demonstrate results, a preference for interesting narratives over abstract measures such as indexes of stability or resilience, and, most importantly, the lack of other ways to measure progress.³²⁶

Others put the need to report success as a part of USAID's institutional culture. As one international development professional put it, “There's always so much pressure to present a good story, including leaving things out when they go wrong. It's not just an M&E issue, it's the entire relationship.”³²⁷ As the head of one implementing partner put it:

Project people are good at making lead look like gold—polishing things up through vague language and obfuscation. . . . There is a culture of ‘nothing can go wrong’ that does not work in a war zone. Lots of factors are outside the manageable interest. There are many outside pressures that interfere and delay a project, but all of this puts nonprofits into a situation where they need to polish things.³²⁸

THE MEASURING IMPACTS OF STABILIZATION INITIATIVES PROGRAM: A UNIQUE ATTEMPT AT ASSESSING IMPACT



A MISTI surveyor interviews an elder in Samangan Province in 2014. (USAID photo)

A significant exception to USAID's lack of rigorous measurement was its Measuring Impact of Stabilization Initiatives program. MISTI was one of several mixed-methods approaches to measuring outcomes and impact during the surge period. As an umbrella that looked over several USAID stabilization programs, MISTI had three primary goals:

- Providing independent monitoring, evaluation and impact assessment of USAID stabilization programs;
- Collecting, synthesizing and analyzing data at the district, provincial and regional levels to track changes in stability over time and help shape U.S. and Afghan government policy and practice related to transition; and
- Contributing to the larger body of knowledge on stabilization activities within a counterinsurgency context.³³⁰

MISTI took a quasi-experimental approach to measuring impact by comparing stabilization-related changes in villages where USAID stabilization projects had been undertaken (treatment villages) to control villages selected by MISTI whose key characteristics matched the treatment villages. MISTI collected a vast amount of data (nearly 200,000 interviews in more than 5,000 villages in 23 provinces), which it used to create indices to measure stability and resilience.³³¹ It also produced a number of critical observations about U.S.

government stabilization programming, including the need for flexibility in programming, the confirmation that stabilization projects may have been destabilizing in some cases, and the finding that isolated projects had a weak impact on stabilization.³³²

MISTI was subjected to a number of criticisms, however. On a methodological level, critics argued that the proxy measure used to gauge local support for the Taliban was unreliable and produced conclusions that were not credible.³³³ After the release of the final analytical report in November 2015, SIGAR and the *New York Times* both noted the “perverse” finding that in some areas support for the Taliban increased *because* of U.S. stabilization assistance.³³⁴ The contractor subsequently reissued the report with a two-page addendum that tried to downplay those findings, attributing them to results from a few villages that were not representative of areas that had received stabilization programming.³³⁵

On a more prosaic level, the lack of accurate data from USAID about its programs made it difficult to track activities. Simply verifying project locations and clarifying poor data took up 60 percent of MISTI's time at the start of the project.³³⁶ Part of the problem was insufficient buy-in from implementing partners. According to an evaluation expert, “One of the tragedies of MISTI was that the implementing partners didn't feel invested, which . . . directly contributed to data quality issues.”³³⁷ Moreover, Afghanistan's chronic lack of reliable data on village and population characteristics made it difficult to create the control and treatment pairs.³³⁸

These criticisms were acknowledged by those who had been involved in MISTI. According to one former USAID official, “the way MISTI captured data needed to be as complex as it was, but [the outputs from MISTI] should have been presented in much simpler way for policymakers. You need to give them actionable information.”³³⁹ Despite these challenges, MISTI produced important observations about the efficacy of USAID's stabilization programming, including those described earlier in this chapter.



Afghan girls hide behind their school books. (USAID photo)

Measurability and Visibility Sometimes Influenced the Choice of Activities

In some cases, the preference for measurability and visibility could drive the choice of activities. One implementing partner staffer explained that during a round of budget cuts in 2017 that required a need to reduce activities, “we were asked to identify things that would show tangible results—and things we could guarantee. We prioritized based on ability to measure and show improvement with causality.”³⁴⁰ In this case, planned innovation grants were dropped; health financing activities that led to increased revenue and international publicity were retained.³⁴¹ On a larger scale, USAID continued and sometimes even increased its funding for the American University of Afghanistan despite persistent problems with the university’s management, processes, and internal controls because USAID viewed the university as an important symbol of U.S. commitment.³⁴²

But measurability and visibility did not always equate with effectiveness. Reporting on trade shows was attractive because the events were visible to the public and because they generated reports on the value of goods sold or the number of deals made. But confirmed deals did not always translate into sales, which was the more important metric.³⁴³ It was also true that one-time sales at a trade show were not as useful as ongoing commercial relationships.³⁴⁴ An external evaluation of the September 2017 “Passage to Prosperity” trade show in New Delhi, while being “cautiously positive” about the benefits of the event, reinforced some of these criticisms: doubts about counting memorandums of understanding as actual deals, and a lack of confidence in USAID’s sales data, which lacked “clarity, completeness, and transparency.”³⁴⁵

Attributing change to specific projects was also difficult: more than one activity could contribute to the same impact, many development activities have delayed impacts, and some projects generated only intermediate outcomes. Attributing impact was even more complicated for projects that were funded from more than one budget. For example, the Initiative for Hygiene, Sanitation, and Nutrition (IHSAN) project was funded out of both health and water/sanitation budgets.³⁴⁶ Aside from the conceptual difficulty of making

a Solomonic assessment of which budget made which impact, this made it nearly impossible to invoice and submit bills.³⁴⁷ Above all else, in a dynamic environment such as Afghanistan, there were often simply too many confounding factors to be able to isolate effects of a given activity.³⁴⁸

One reasonable way to understand impact while solving for such challenges is through a concept called “plausible contribution” (see callout box below).

Struggling to Find the Right Level of Information and Reporting

Some implementing partners reported that USAID erred on the side of requiring too much information—what one former USAID administrator referred to as “obsessive measurement disorder.”³⁴⁹ One implementing partner chief of party described a municipal-support

Plausible Contribution May Be a Reasonable Substitute for Impact Assessment

In some if not many cases, the link between activities and impacts in an environment like Afghanistan can only be evaluated on the basis of a judgment, as opposed to a firm conclusion on attribution—especially where impact is measured at the national level. As one former senior USAID official said, “At the end of the day, you have to make a judgement on attribution, because there is usually a lot going on in the environment.”³⁵⁰

Implementing partners and USAID personnel had a range of perspectives on the feasibility of doing true impact evaluations in an unstable environment such as Afghanistan. Rigorous evaluation methodologies such as randomized controlled trials and quasi-experimental design can show impact because there is a “treatment” and a “comparison” (or “control”) group. But these are especially difficult to do, due to lack of resources, the skill levels required, maintaining the distinction between treatment and control groups, and insecurity or other factors that may cut off access to one or more of the groups.³⁵¹ The studies of the National Solidarity Program and Community-Based Education are two prominent examples of randomized controlled trials conducted in Afghanistan.³⁵²

Evaluation professionals were especially skeptical. As one noted, “Impact evaluations are passé because they are too complex and also not relevant. They don’t make sense for what USAID does because often you can’t get a control for comparison. Having a control environment where you don’t do a program is a wonderful idea. But how does USAID apply it in all sectors?”³⁵³ Moreover, the potential

use of these methodologies also raised ethical concerns: measuring the impact of, for example, health care delivery would require providing health care to one village while withholding it from another.

Of course, it may be more unethical to provide healthcare to no villages, and at least equally unethical to implement unproven interventions on a wide scale. For example, as SIGAR has reported, U.S. stabilization efforts often exacerbated conflicts, enabled corruption, and bolstered support for insurgents.³⁵⁴ Additionally, to some extent, there can be a mythology of difficulty and complexity surrounding randomized controlled trials that may discourage their use.³⁵⁵ As one former USAID official commented, “Sure, RCTs are technical, but this isn’t black ops.”³⁵⁶

But debates surrounding how best to assess impact need not be either-or. In evaluating some of its work, the UK’s Department for International Development (the UK’s USAID-equivalent that was merged into the Foreign, Commonwealth and Development Office in 2020) uses a “contribution analysis” approach that examines whether the intervention could have made a “plausible contribution” to results.³⁵⁷ This approach says that “if an evaluator can validate a theory of change with empirical evidence and account for major external influencing factors, then it is reasonable to conclude that the intervention has made a difference.”³⁵⁸ Under this approach, a full randomized controlled trial is not needed to measure impact.

project that had so many assessments and surveys that they delayed actual implementation, generating complaints from local counterparts that they were only doing assessments and plans.³⁵⁹ A former USAID official said that “overmeasurement comes from a genuine desire to do well and achieve data-driven, smart programming.” The problem, the official added, was learning how to “pick three good indicators rather than 500.”³⁶⁰ Still, a few implementing partners regarded some of the reporting as simply a “box-checking exercise.”³⁶¹

On the other hand, as one current USAID official stated, “The implementing partners get a lot of money to do the data collection, and they choose to bid on the contract or grant. Honestly, I don’t have much sympathy for implementers who complain about the reporting.”³⁶² Whether it was to relieve implementing partners from an onerous reporting burden or to reduce their own workload, USAID officials did at times make an effort to reduce reporting requirements.³⁶³ As one official summarized, “you have to curate the M&E process; if not, you risk being buried under data.”³⁶⁴

The level and type of interaction and reporting was partly dependent on individual USAID personnel, which meant that it could change with the arrival of a replacement for the contracting officer’s representative—something that happened frequently.³⁶⁵ The frequent turnover of contracting officer’s representatives and agreement officer’s representatives could be disruptive. According to one implementing-partner project’s chief of party, “there were three sequential [contracting officer’s representatives], each with a different vision of report writing and communication intimacy. Every year, [the implementing partner] had to revamp their reporting because the [contracting officer’s representative] wanted something different. . . . Sometimes the monthly reports were three pages, sometimes they were 30. Sometimes they were more than 1,000. It varied by the [contracting officer’s representative].”³⁶⁶

A key question for the agency was where to store all of these reports so that project managers could efficiently examine and analyze them.³⁶⁷ Afghan Info, a web-based tool, was created in 2010 as a central repository for all project-related information, including monitoring data.³⁶⁸ It was intended to be accessible to all implementing partners, who would be required to regularly update their entries. It was also intended to contain information from all U.S. government agencies working on Afghan reconstruction. This broader function was never achieved, due to a combination of technical shortcomings and data quality issues.³⁶⁹

Five years after it was established, Afghan Info still contained no mechanism for entering the monitoring data it was supposed to centralize, and instead functioned primarily as a repository for project documents.³⁷⁰ Even in this area it was not always useful: Implementing partners would upload reports that then sat in the system for long periods of time before being approved by the contracting officer’s representatives or agreement officer’s representatives charged with managing the awards. Of a sample of 27 reports uploaded for the quarter ending December 31, 2014, and reviewed by the USAID inspector general, only four had been approved.³⁷¹ As a result of low rates of document approval, the inspector general said, “There was little evidence that the accuracy of data in Afghan Info had been verified.”³⁷²

According to USAID, the database has improved substantially in recent years.³⁷³ This claim is broadly consistent with both the observations of those interviewed for this report, and SIGAR's own experience with the system.³⁷⁴ For example, staff from one USAID implementing partner stated that Afghan Info was a good system, noting that USAID was working to ensure that contracting officer's representatives uploaded all required information.³⁷⁵ Staff from another implementing partner stated that as long as an activity's M&E plan was reflected in Afghan Info, the system worked well.³⁷⁶ Still, some shortcomings remain. For example, the absence of granular financial data limits the type of analysis that can be performed.³⁷⁷

PUTTING M&E TO USE

On paper, USAID stressed the use of M&E activities for learning and adaptation. The agency's 2016 evaluation policy emphasized that "USAID bases policy and investment decisions on the best available empirical evidence, and uses the opportunities afforded by project implementation to generate new knowledge for the wider community."³⁷⁸ The policy noted, however, that evaluation "must be embedded within a context that permits evidence-based decision-making, and rewards learning and candor more than superficial success stories."³⁷⁹

Various respondents and reports raised questions about whether USAID effectively used M&E information to make decisions, at least commensurate with the agency's stated focus on monitoring and evaluation.³⁸⁰ However, both implementing partners and USAID personnel said evaluations were useful in documenting needed changes in activities, in communicating with program stakeholders, and providing a platform for the implementing partner and USAID to talk about issues.³⁸¹ Midterm evaluations were also considered useful in order to proverbially "kick the IP" if it was failing or to respond to some negative event.³⁸²

The following section discusses the political and organizational constraints that USAID faced in learning from and using information to make evidence-based decisions. These included political and military pressure, the lack and unpredictability of time and resources, the complexity of the operating environment, and a range of human resource and organizational issues. In short, USAID faced many challenges.

Seeking the Right Mechanisms for M&E

Even as early as 2003, the security situation restricted the ability of USAID personnel to adequately monitor projects through regular site visits.³⁸³ Afghanistan's insecurity remained a key obstacle to adequate oversight throughout the early years of reconstruction.³⁸⁴ As U.S. troop levels declined significantly between 2011 and 2016, movement for U.S. government personnel became even more limited, with civilian agencies losing enabling mobility, life support, and medevac functions—losses that raised concerns USAID would become even more dependent on implementing-partner and third-party monitoring reports.³⁸⁵ In response, USAID developed a number of mechanisms to improve M&E. Some seemed to consist of a rebranding of standard M&E practice, and the various mechanisms had mixed results.

Multitiered Monitoring: An Attempt to Triangulate Data that Failed to Live up to Expectations

In 2014, in response to concerns that the reduction in the military presence would limit the completeness and accuracy of monitoring information, USAID introduced multitiered monitoring.³⁸⁶ The system was supposed to use multiple data sources that would allow USAID officials to compare and triangulate information.³⁸⁷ In 2017, USAID acknowledged that “multitier” was a bit of a misnomer, since there was no defined hierarchy of sources, but the agency elected to keep the terminology due to its familiarity with stakeholders.³⁸⁸ Multitiered monitoring seemed in some ways like a rebranding of what should be standard practice: collecting and verifying information on projects.³⁸⁹

There is little evidence that multitiered monitoring was implemented in anything close to an adequate fashion, in part because the needed supporting information systems and management processes were not in place. A highly critical 2015 USAID inspector general report noted a variety of flaws: a lack of guidance to the technical offices on the level of monitoring and the mix of data from each of the five tiers; the inability of the Afghan Info database to accommodate all of the data; a lack of monitoring plans; and the inadequate functioning of the team charged with making recommendations based on review and analysis of the data.³⁹⁰ The main reasons for the team’s limited functioning were the workload of team members and insufficient guidance from mission officials.³⁹¹

The report found that only one of 127 contracts, grants, and cooperative agreements had used multitiered monitoring as had been intended.³⁹² According to one evaluation expert, “[US]AID was under pressure to do it, and whether or not they ever applied it, I don’t know.”³⁹³ In September 2017, USAID went from five to three tiers.³⁹⁴

Third-Party Monitoring: An Attempt to Outsource Data Collection That Achieved Mixed Effects

Third-party monitoring was exactly what it sounded like: It was based on the principle that disinterested observers provided a more objective view of what was (or was not) happening on the ground. Third-party monitoring had been introduced to Afghanistan in the early years of reconstruction as part of the international collaboration on the Basic Package of Health Services (see highlight box on the following page). While it had been used since 2002, its perceived importance as a source of information on project progress grew during the 2014–2015 transition period as USAID lost access to regional facilities and the number of agency personnel in Afghanistan declined.³⁹⁵

Levels of Multitiered Monitoring

Tier 1: Hands-on monitoring by USAID or other U.S. government officials

Tier 2: Required reporting by implementing partners

Tier 3: Feedback from the government of Afghanistan and international donors

Tier 4: Local civil society organizations and project beneficiaries personally interested in a project’s progress

Tier 5: Independent third-party monitors³⁹⁶

THIRD-PARTY MONITORING IN THE HEALTH SERVICES SECTOR ILLUSTRATES THE STRENGTHS AND LIMITATIONS OF OUTSOURCING DATA COLLECTION



A midwife supported by USAID tends to a newborn at a medical facility in Kandahar Province. (USAID photo)

In early 2002, the Afghan Ministry of Public Health agreed with its donor partners to a system of contracting with NGOs for the delivery of a Basic Package of Health Services that would be available to all citizens. Although a significant percentage of Afghanistan's health services were already being provided by NGOs due to years of conflict, the agreement to contract out services was a radical break with historical government policy that the government was the sole officially recognized provider of health care. The three main donors to the health sector—The World Bank, USAID, and the European Commission—all adopted performance-based partnership agreements, each with slight variations in approach.³⁹⁷

To compensate for the lack of a functioning health management information system, and to monitor progress on overall healthcare delivery and the performance of the individual NGO providers, the Ministry of Public Health and its international partners contracted for a system of third-party monitoring of health services in all provinces via Johns Hopkins University, in cooperation with the Indian Institute of Health Management Research.³⁹⁸

The scorecard that monitoring produced consisted of 29 core indicators and benchmarks intended to capture six domains of health services: patient perspectives, staff perspectives, capacity for service provision, quality of services, financial systems, and overall vision for the health sector (which captured whether the system was

providing equitable service to different demographics). It also incorporated two composite scores composed of the percentage of benchmarks achieved. While information was collected on individual facilities, the main unit of analysis was the province.³⁹⁹ This system quickly developed a reputation as a useful tool for the ministry, the donors, and even the NGOs in oversight of the health sector. Among its positive characteristics were:

- Standardization of monitoring results, which created objective indicators for how overall performance was going, and simplified discussion at the national level between provincial health officials and the central ministry, including negotiations on staffing and other types of support;
- Identification of specific geographic areas and medical conditions in special need of attention and performance improvement; and
- A gauge for performance that helped to justify performance bonuses, the continuation or termination of contracts.⁴⁰⁰

But the system also had limitations, including:

- Limited information on service coverage and on actual health outcomes;
- Lack of information on facilities not covered by the Basic Package of Health Services (including some hospitals); and
- An inability to show the reasons for underlying differing performance by provinces, or to accommodate insecurity and other contextual factors in measuring performance.⁴⁰¹

Critics have also pointed to the high costs associated with contracting with international organizations for its implementation. Defenders insisted that the costs were actually reasonable, considering Afghanistan's difficult conditions (insecurity, high transportation costs, the need for both male and female staff) and the fact that setting up the system involved a lot of technical assistance and capacity building.⁴⁰²

USAID's most ambitious third-party monitoring initiative was the Monitoring Support Program, the objective of which was "to provide USAID/Afghanistan with additional data on project implementation, allowing technical teams to compare and verify information from their own monitoring efforts with additional sources of monitoring data, and make effective management decisions on project performance."⁴⁰³ The Monitoring Support Program covered 56 projects (contracts, grants, and cooperative agreements), as well as activities implemented by the Afghan government through on-budget assistance.⁴⁰⁴

The Monitoring Support Program ended in August 2017, three years before its planned end date. According to implementing partner staff with direct knowledge of the program, its basic problem was its structure, which divided Afghanistan into three regions, each assigned to a different monitoring contractor.⁴⁰⁵ This meant that national projects implemented in all three regions of the country could be visited by three different monitors.⁴⁰⁶ Consequently, USAID and the implementing partners spent a large amount of time coordinating the scope and design of data collection.⁴⁰⁷ Moreover, USAID did not establish a hierarchy among the monitors, so the regional contractors were left to sort that out themselves.⁴⁰⁸

In addition, there was a skepticism about the Monitoring Support Program both within the mission and among the implementing partners. Implementing partner staff reported that because there was no ongoing M&E capacity in the mission, they had to sell staff on the program's utility. Added to these problems were the ever-present issues of short tours, the rotation of responsibilities of staff during their tour, and work overload.⁴⁰⁹

Finally, given the time and experience constraints within the mission, it may have been unrealistic to expect anything more from the Monitoring Support Program than simply verifying that an activity took place. As one of the regional contractors observed, "Many projects start with verification, but you need an additional effort (analytics) to figure out what it means."⁴¹⁰

Of course, this observation applied to third-party monitoring more broadly. In general, monitors could check project activities only in the most superficial way, partly because they were unable to draw credible conclusions based on seeing a small slice of activities within the limited time frame available. Dropping in on a meeting, for example, does not allow a qualitative assessment of what the meeting does. As one evaluation consultant noted, monitors "can see the number of people who signed into a meeting, but could not assess attendees' opinion of the training or if they learned from it. . . . [USAID] could give GPS stamps, photos, and dates of events, but because the qualitative side is not assessed, it has limited value."⁴¹¹ That viewpoint was echoed by a senior official at an Afghan implementing partner, who said, "Third party monitoring is looking at things like numbers of people at trainings and contract deliverables. It's only form filling and a box checking exercise. It's for compliance, not integrity."⁴¹²

Indeed, the integrity of third-party and implementing-partner monitoring reports was sometimes questionable. In the course of a 2012 SIGAR audit of a USAID stabilization



A healthcare worker treats a mother and infant. (USAID photo)

project, for example, contractor Development Alternatives, Inc., reported “allegations of gross employee misconduct that included the falsification of monitoring and site visit reports.”⁴¹³ As SIGAR reported in February 2021, remote management can lead to inaccurate project data and reporting, as well as to fraud, corruption, and failure to report corruption.⁴¹⁴ The central strength of third-party monitoring remains the opportunity to monitor projects implemented in areas inaccessible to U.S. government personnel.⁴¹⁵ If carefully managed, third-party monitors can provide useful verification of basic project outputs, such as whether a building has been constructed.⁴¹⁶ As the United States continues to draw down its presence in Afghanistan, third-party monitoring may emerge as an even more important mechanism for ensuring that U.S. funds are spent in service of their intended purpose.

After ending the Monitoring Support Program, USAID consolidated its monitoring activities into the existing Services under Program and Project Offices for Results Tracking II project.⁴¹⁷ The subsequent Afghanistan Monitoring, Evaluation, and Learning Activity covered both monitoring and evaluation.⁴¹⁸ Because the Afghanistan Monitoring, Evaluation, and Learning Activity covered the entire country, it reduced the need for the contractor to coordinate with other partners.⁴¹⁹

MONITORING WHETHER DEVELOPMENT PROGRAMMING BENEFITTED THE TALIBAN: USAID SENDS MIXED MESSAGES

In Afghanistan, development programming implemented in Taliban-controlled areas carried the potential to actually harm stabilization efforts.⁴²⁰ For example, USAID's Measuring Impacts of Stabilization Initiatives program (described earlier in this chapter) found increased support for the Taliban when USAID stabilization programs were implemented in villages under Taliban control.⁴²¹

Later research further underscored the risk that insurgents could benefit from development projects. USAID viewed education and health service delivery as a way of generating "increased confidence in the legitimacy and effectiveness of the [Afghan] government" that would "foster stability."⁴²² But in 2017, the World Bank found that some Taliban were simply co-opting Afghan government schools, rather than attacking or closing them as they had done in the past.⁴²³ The Bank also described instances where insurgents actually protected health services: In the words of one of the Bank's informants, "Ambulances can easily travel around [in Wardak Province], because the Taliban and militants need health services."⁴²⁴ A prominent researcher of Taliban governance summarized the key risk: "The Taliban leadership realized that instead of attacking government schools and aid projects, it could gain much more by co-opting them. In doing so, it could take credit for providing services and win over the local population."⁴²⁵

Direct diversion of U.S. funds to the Taliban—described by some as an "open secret" of the reconstruction and warfighting efforts in Afghanistan—was also a concern.⁴²⁶ Media reports in 2009 and 2010 alleged that the Taliban siphoned off a proportion of funds spent on contracts.⁴²⁷ For example, an article published in 2009 by *GlobalPost*, a U.S.-based digital journalism company, described the existence of a Kabul-based Taliban "contracts officer" who negotiated with major Afghan contractors for a percentage of the value of contracts funded by donors.⁴²⁸ A USAID

inspector general review prompted by this and other allegations found that Afghan subcontractors working on USAID's \$349 million Local Governance and Community Development project, which aimed to foster political, economic, and social development in communities located in insecure areas, may have used USAID funds to pay Taliban insurgents up to 20 percent of the total value of subcontracts, in exchange for protection from Taliban security guards and Taliban promises not to attack subcontractor personnel.⁴²⁹ Interviews indicated that subcontractors often recouped the funds paid to insurgents by including the amount of anticipated protection payments in the total cost of subcontracts.⁴³⁰

Most USAID officials interviewed for the review believed it was impossible to properly monitor whether USAID funds were being diverted to the Taliban "in a war zone and in insurgency stronghold areas," according to USAID's inspector general.⁴³¹ Eight years later, in 2017 testimony to the Congress, Greg Huger, then a senior official in USAID's Office of Afghanistan and Pakistan Affairs, described safeguards put in place by USAID "to better ensure that funds supporting projects do not benefit or further the legitimacy of the Taliban or other violent extremist organizations."⁴³² Huger's statement implied that USAID had adopted an active role in ensuring funds were not inadvertently spent in ways that legitimized the insurgency.⁴³³ But in a January 2019 response to SIGAR queries regarding how USAID was using monitoring to determine whether its projects were being implemented in Taliban-controlled areas, USAID deflected responsibility. "USAID defers to implementing partners (and third-party monitors) to use the sources that they deem appropriate to determine whether a site is accessible and stable enough for programs to operate," the agency told SIGAR. "USAID does not make this assessment on behalf of the implementing partners."⁴³⁴

Shifting responses from USAID regarding the degree to which it played an active role in ensuring its projects were not being implemented in Taliban-controlled areas may signal an irreconcilable dilemma for development professionals operating amid armed conflict. On one hand, helping the Afghan government provide key services such as health and education was intended to help legitimize the Afghan government and foster stability.⁴³⁵ But, as a senior USAID official explained to SIGAR in 2017, the mission of ensuring that children are getting an education and that Afghans have access to health care remained a priority even if USAID funds were spent on such services in Taliban controlled areas.⁴³⁶

To explore this issue further, in February 2019, SIGAR asked five USAID technical offices to provide information on where their implementing partners were working, by district and province.⁴³⁷ SIGAR received USAID's responses on March 21, 2019.⁴³⁸ SIGAR compared the districts where USAID implementing partners were working with the last available district stability assessment produced by the U.S.-commanded NATO Resolute Support mission in Afghanistan. The district stability assessment was current as of October 22, 2018.⁴³⁹ The purpose of the comparison was to ascertain whether there was a reasonably high probability that some USAID implementing partners were working in areas controlled or influenced by the Taliban.

The analysis showed that, as of March 2019, USAID was implementing multiple projects in districts that, five months earlier, were categorized by Resolute Support as controlled or influenced by insurgents.⁴⁴⁰ For example, an implementing partner for USAID's Initiative for Hygiene, Sanitation and Nutrition project, which aimed to improve the nutritional status of young children and women of reproductive age, was working in five districts in Helmand Province categorized by Resolute Support as being under the control of insurgents as of October 22, 2018.⁴⁴¹ Similarly, an implementing partner of a USAID community-based education project was working in two districts in Uruzgan Province that, according to Resolute Support, were under insurgent influence as of October 22, 2018.⁴⁴²

The five-month time lag between Resolute Support's district stability assessment and the date on which SIGAR received the data compiled by USAID raises the possibility that the aforementioned districts in Helmand and Uruzgan Provinces changed hands between October 2018 and March 2019. USAID has told SIGAR that it is difficult to tell from day to day whether an area is controlled by the Afghan government or the Taliban.⁴⁴³ However, the control status of these districts did not change over the 12 months during which unclassified district stability data were provided to SIGAR.⁴⁴⁴ Although there was no definite proof, there appeared to be a strong possibility that, as of March 2019, some USAID projects were being implemented in areas controlled or influenced by the insurgency.

Presently, U.S. law prohibits USAID from using funds appropriated in certain fiscal years to finance projects that risk legitimizing the Taliban.⁴⁴⁵ However, international organizations are not subject to similar prohibitions. In December 2020, the United Nations Children's Fund, which has been a USAID implementing partner, reportedly struck a deal with the Taliban to establish 4,000 community-based education classes across four provinces largely controlled or influenced by the insurgency.⁴⁴⁶ Along similar lines, an intra-Afghan peace deal may eventually prompt U.S. agencies to explicitly and overtly fund projects that benefit an Afghan government co-run by the Taliban. Senior U.S. government officials have increasingly referenced using future U.S. foreign assistance as leverage for influencing Afghanistan's post-peace political and human rights landscape.⁴⁴⁷ Until an intra-Afghan peace agreement materializes, it will remain an open question whether USAID has sufficient monitoring in place to ensure that development projects do not indirectly legitimize the Taliban, and thus work against the stability objectives they are supposed to support.

Third-party Evaluation: An Attempt to Independently Assess Project Performance that Demonstrated Mixed Utility

In 2006, USAID established the Services under Program and Project Offices for Results Tracking project, the agency's first dedicated evaluation project in Afghanistan.⁴⁴⁸ These evaluations were outsourced to third-party contractors—supposedly disinterested entities who evaluated implementing partners' activities. Emphasis on evaluation rose during the 2009 to 2012 military and civilian surge, when program funding significantly increased and USAID's heavy reliance on contractors emerged as a key reconstruction theme.⁴⁴⁹ Although SUPPORT (and subsequently SUPPORT II and the Afghanistan Monitoring, Evaluation, and Learning Activity) were USAID's main mechanisms for evaluation, they also performed a number of other management functions that would normally have been done in-house, such as producing strategic communications and information products, facilitating meetings, and providing translation services.⁴⁵⁰

But they also presented a number of challenges. Both USAID and its implementing partners raised questions about the quality and utility of evaluations in general, and in specific cases. A former mission director noted that “as an institution, the majority of [US]AID people welcome rigorous evaluations and criticism, but many staff don't have confidence in the quality of evaluations.”⁴⁵¹ Interviewees expressed a range of complaints—that evaluators misunderstood the reality on the ground, that an analysis was superficial or outside the scope of the evaluation, that facts were misconstrued, or that evaluators relied on anecdotal evidence.⁴⁵² In some cases, implementing partners disparaged the evaluation team's qualifications or experience, or the way it had conducted itself or organized the work, or they accused evaluators of bias.⁴⁵³ But as one former USAID mission director noted, “people who are personally invested have a hard time being objective. Human nature is to avoid negative findings on something that you've spent a lot of time on.”⁴⁵⁴

“People who are personally invested have a hard time being objective. Human nature is to avoid negative findings on something that you've spent a lot of time on.”

—Former USAID/Afghanistan mission director

Such tensions seemed to produce tortured language. In March 2017, for example, USAID conducted a midterm evaluation of its \$142.0 million Power Transmission Expansion and Connectivity Commercialization Activities project, which aimed to turn Afghanistan's national power utility into a more commercially viable business entity.⁴⁵⁵ The assessors, Checchi and Company Consulting, Inc., stated, “Overall, the team found . . . that achievement of the component's intended results was mixed.”⁴⁵⁶ Yet later in the report, the assessors stated that “overall, the [project's] commercialization component *did not achieve* its intended results” (emphasis added)—leaving the reader confused about just how effective or ineffective the project really was.⁴⁵⁷

Likewise, it was difficult to parse the phrasing of a mid-course stocktaking exercise (essentially the strategic-level equivalent of an evaluation) examining USAID's 2015 to 2018 strategy for Afghanistan. According to the document produced by the exercise, "Key conditions of the development hypothesis [the theory about which specific actions would help achieve strategic goals] did not prove to be valid, and most critical assumptions that underpinned it have not held true throughout implementation." Yet the same document said that key components of the strategy's **results framework**—defined by USAID policy as "representing the development hypotheses underlying the mission's strategy"—*"are still valid"* (emphasis in original source).⁴⁵⁸

Defining Key USAID Terms: Results Framework

M&E systems at USAID begin with strategy—specifically, a Country Development Cooperation Strategy, which articulates a mission's objectives for a given country and the planned approach for achieving them.⁴⁶² The strategy is mapped out in the **results framework**—a visual representation of a hierarchical sequence of outcomes that are presumed to produce strategic effects.⁴⁶³ It describes how intermediate outcomes, if achieved, will result in the attainment of increasingly broader objectives, stretching all the way along an assumed causal hierarchy towards an overarching, country-level goal.⁴⁶⁴ For example, in Afghanistan, the results framework relies on the assumption that:⁴⁶⁵

- increasing the commercial viability of Afghan businesses and the productivity of key agricultural crops will
- improve the competitiveness of Afghan value chains, which in turn will
- accelerate exports and economic growth
- thereby rendering Afghanistan a more "economically viable" country that is better positioned to be a national security partner of the United States.

If the underlying assumptions hold true, any activity aiming to achieve the initial outcome of increasing commercial viability should ultimately contribute towards the strategic goal of improving the security partnership between the United States and Afghanistan. In general, the closer an outcome is linked to actual implementation, the more accountable USAID holds itself for results. Appendix B explains the concept of a results framework in greater detail.

The assessors attributed "this somewhat contradictory finding" to "some modifications" to the results framework.⁴⁵⁹ However, they added, "some senior USAID officials . . . questioned the validity of the framework goal, 'Afghan-led, sustainable development,' noting that in mid-2017 it seems to be 'aspirational.'"⁴⁶⁰ According to USAID's July 2020 M&E policy, "the goal should be practical."⁴⁶¹ Given the serious questions about the achievability of USAID's strategic goal expressed by some senior USAID officials at the time the stocktaking exercise was conducted, it was difficult to understand the assessors' insistence regarding the supposed validity of the strategy's results framework. Table 3 on page 68 presents additional examples of apparent logical flaws or contradictions in USAID evaluations.



An Afghan farmer inspects his pomegranate harvest. Aiming to put Afghanistan on a track of sustainable development, USAID implemented programs that provided Afghan farmers with high-value crops that could be exported from the country. (USAID photo)

Still, in at least some cases, evaluators did not appear to have given the mission and its implementing partners the good—or the heavily qualified bad—news that they wanted. Where agreement couldn't be reached, the mission, sometimes with the collaboration of the implementing partner, appended a “statement of differences.”⁴⁶⁶ Ultimately, an evaluation's utility depends on a number of factors, including speed, timing, and the credibility of the process and the people involved. The highlight on page 70 explores these factors through the eyes of knowledgeable interviewees.

Political and Military Pressure and the Hectic Pace of Execution Sometimes Undermined Evidence-Based Approaches

The high visibility nature of the U.S. engagement in Afghanistan and USAID's vulnerability to political pressures often made it difficult for officials to make decisions based solely on the evidence.⁴⁶⁷ For example, the Afghanistan Small and Medium Enterprise Development Project had to respond continuously to requests from U.S. government sources, including provincial reconstruction teams, to carry out additional activities that diverted “significant staff resources due to their urgent and high-profile nature.” Since the project was also considered part of counterinsurgency strategy, it was sometimes asked to shift its activities to areas the military considered important, and to do so under an “unrealistic” time frame.⁴⁶⁸

The hectic pace of execution affected the quality of M&E in another way: Some projects were implemented without crucial performance indicators. A 2017 SIGAR audit report on the Afghanistan Infrastructure Fund (AIF) concluded that USAID, as well as State and DOD, had not developed the performance metrics that would have allowed an assessment of whether the fund had achieved its counterinsurgency objectives, despite the fact that AIF projects were supposed to support the U.S. counterinsurgency strategy.⁴⁶⁹

TABLE 3

ILLUSTRATIVE CONTRADICTIONS OR LOGICAL FLAWS IN USAID EVALUATIONS

Project/Program	Evaluation Year	Project/Program Purpose	Statement One
Assistance in Building Afghanistan by Developing Enterprises (ABADE)	2018	Create jobs by providing small and mid-size businesses with capital equipment and technical assistance	“ABADE’s purported contribution to jobs was 17,777” (p. viii). Evaluators concluded that ABADE’s “provision of [machinery and equipment] to successful business owners with solid expansion plans did generate substantial and sustained employment” (p. xi).
Passage to Prosperity (P2P): India-Afghanistan Trade and Investment Show	2018	Develop trade and financial relations between Afghanistan and India	“There is sound quantitative evidence of [increased exports] generated by the P2P Show, both in comparison of before and after export performance and involving with-without analysis” (p. x).
Promote Women’s Leadership Development (WLD)	2017	Empower Afghan women to pursue meaningful professional careers in all sectors (public, private, and civil society)	“By all accounts, from participants to principals to community leaders and government representatives to subcontractors and local sub-grantees, the WLD program is doing a good job of preparing a new generation of Afghan women leaders” (p. 9).
Power Transmission Expansion and Connectivity (PTEC) Commercialization	2017	Improve electrical efficiency and capacity in Kabul and four provinces, provide technical assistance to support critical power infrastructure, and improve management of the power industry	“Achievement of the component’s intended results was mixed” (p. ii).
Maternal and Under-Five Nutrition and Child Health (MUNCH)	2016	Improve the health and nutrition of infants, children and mothers	“The project aims to... influence the outcomes of maternal, newborn and child nutrition, through gender-sensitive implementation” (p. 8). “The overall goal of the MUNCH project is to contribute to the basic health needs and to reduce the vulnerability of the people of western Afghanistan, with a focus on women and girls” (p. 1). “MUNCH was generally successful in meeting its goals and objectives” (p. 3).
Support to Subnational Governance Structures (SNG)	2013	Provide support to provincial councils (PCs) and encourage positive relationships between PCs and constituents	“The [National Democratic Institute] project had a positive effect on PCs’ performance” (p. 2). “PCs have good relationships with constituents” (p. 2).

Source: Checchi and Company Consulting, Inc. “Performance Evaluation, Support to Sub-National Governance Structures.” February 2013, pp. 2, 6, 7; USAID Office of Inspector General, “Audit Of USAID/Afghanistan’s Assistance in Building Afghanistan by Developing Enterprises Program,” Audit Report No. F-306-16-002-P, March 14, 2016, pp. 2, 9, 18; Checchi and Company Consulting, Inc. “Final Evaluation of the Maternal and Under-Five Nutrition and Child Health Project,” November 2016, pp. 1, 3, 8, 10–11; Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation Power Transmission Expansion And Connectivity (PTEC) Commercialization Activities,” October 2017, pp. ii, iv; Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation of the Promote Women’s Leadership Development Program,” September 2017, pp. 9, 20–22; Checchi and Company Consulting, Inc., “Final Performance Evaluation of Assistance In Building Afghanistan by Developing Enterprises,” December 2018, pp. viii, xi, 7–9, 12, 20, 60; Checchi and Company Consulting, Inc., “Assessment of the Passage to Prosperity: India Afghanistan Trade and Investment Show,” December 2018, pp. x, 14–15, 24.

MONITORING AND EVALUATION OF RECONSTRUCTION CONTRACTING

Statement Two	SIGAR Analysis
“[An] OIG audit ... concluded that ABADE was “not tracking actual results” and that reporting was “of low quality and the numbers inflated” (p. 20). “The program reported the number of jobs created before partnerships received the equipment, although the equipment was the primary catalyst for generating new jobs” (p. 2).	While the evaluation included the important qualifier “purported” in front of the jobs figure in its opening findings, subsequent statements raised so many questions about the validity of that number (which was taken directly from the implementing partner), that there may have been cause to omit it from the evaluation’s findings altogether. Because it is difficult to assess whether this figure is accurate, it is similarly difficult to know whether the analysis based on the data is sound. Ultimately, although the evaluation claimed that ABADE “generate[d] substantial and sustained employment,” the implementing partner estimated the number of jobs created before beneficiaries received the equipment that was supposed to create the jobs.^a
“Based on the survey, there is little quantitative evidence of [increased exports] generated by the P2P Show” (p. 24). “When looking at changes in median Indian exports per firm ... none of the differences achieve statistical significance at traditional levels” (p. 15). “All that can be said is that this change is greater than zero” (p. 14–15).	The assessment initially claims success and then backtracks to clarify that only one of multiple data analyses shows statistically significant increases. One analysis notes that an increase in export values exists but is not statistically significant. Another analysis reports no change in export value from before to after the trade show, and yet another reports a statistically significant increase of an unknown greater-than-zero value. This suggests that there may be some evidence suggesting increased export value, but that the evidence is quantitatively inconclusive.
“There are a few areas of WLD program implementation that can be improved, including scaling up enrollment to meet program targets (mid way through the program, only 20 percent of the Jawana target and 16 percent of the Royesh target have been met)” (p. 9).	The first statement asserts that WLD is unambiguously achieving desired results. The follow up reveals that the phrase “by all accounts” is an inaccurate characterization, as two key training components were significantly off track to meet life-of-program targets midway through the program’s implementation.
“Overall, the PTEC’s commercialization component did not achieve its intended results” (p. iv).	Results cannot simultaneously be mixed and unachieved. While results could technically be mixed but on balance, unachieved, the first statement does not make that distinction.
“[World Vision’s] Annual Work Plans and Progress reports were not transparent in their explanation of monitoring and reporting methodology and many inconsistencies were found across years and indicators; moreover, many interventions did not have clear targets. Additionally, disaggregation by province and/or gender was rare and inconsistent leading to difficulties in analyzing secondary data.” (p. 10). “The evaluation relied largely on qualitative methods to collect data from participants who were willing and able to travel from their villages to the provincial centers.” This limited “the ability to make statistically valid conclusions and to generalize the results to the entire MUNCH project area” (p. 11).	In a project explicitly designed to focus on women and girls, and emphasizing gender-sensitive implementation, data provided by the implementing partner was rarely disaggregated by province or gender. This and other data limitations made it very hard for evaluators to draw statistically valid conclusions. Because of a lack of granular outcome data, it was difficult to actually know whether “MUNCH was generally successful in meeting its goals and objectives.”
“[A] limitation was linked to the scarcity of comprehensive baseline data required to measure the degree of improvement in PC performance during the project’s life to date” (p. 7). “...it was not possible to directly obtain citizens’ perceptions of PC effectiveness, so the performance of the PCs was not able to be adequately measured in terms of their impact on direct service recipients” (p. 6).	The report claims that provincial councils and community constituents have good relationships, without having interviewed any constituents. Improvements to performance were claimed without any established baseline data to compare against. The report claims a positive impact, despite significant limitations that made it difficult to obtain evidence of measureable results.

^a USAID’s Office of Inspector General described ABADE’s job-creation data as “not valid or reliable.” Specifically, the data was based on beneficiary self-reporting and was premature, as job-creation figures were collected before beneficiaries had actually received the machinery and equipment that was supposed to create the jobs. This and other problems with ABADE led the Office of Inspector General to recommend an immediate pause in creating new public-private alliances. Evaluators attempted to confirm the number of jobs ABADE claimed to have created with their own representative survey of beneficiaries. Curiously, the survey found that the average number of jobs created per beneficiary business was 64, similar to ABADE’s claim of 60. However, like ABADE, the evaluation team’s survey relied on beneficiary self-reporting and therefore also may have been unreliable. Even though evaluators noted “considerable discrepancies among job numbers,” they concluded that ABADE “did generate substantial and sustained employment.” Their recommendation was far different than the one offered by USAID’s Office of Inspector General: “Offer more programming like ABADE.”

WHAT MAKES FOR A GOOD EVALUATION IN AN ENVIRONMENT SUCH AS AFGHANISTAN?

The ultimate utility of an evaluation depends on a number of factors, including speed, timing, and the credibility of the process and the people involved. This highlight explores these factors through the eyes of knowledgeable interviewees.

Finding the Right Balance: Navigating Trade-offs in Quality, Speed, and Cost

Evaluations can be rigorous, rapid, or cost-effective, but there are tradeoffs between these qualities. The length and complexity of some USAID performance evaluations were driven in part by an attempt to be rigorous and therefore credible. A number of interviewees recommended that evaluations be better focused and that reports be shorter, more readable, and have more actionable recommendations, in part because the findings are so local and context specific.⁴⁷⁰ As one evaluation expert said, “The programs are so large and complex, there is a need to focus on the narrative, because often the data doesn’t tell the story.”⁴⁷¹

One USAID official noted that “if I could change one thing, it would be to cut down the number of questions. The scope of an evaluation is often bigger than it should be, and it therefore takes longer to carry out evaluations. This holds things up. There should be clearer questions on the evaluation, so that the evaluators ask only what is important, and then USAID and the implementing partners can act on the findings.”⁴⁷²

M&E practitioners generally agreed on the limited ability to achieve higher-level outcomes or impact within three to four years, and therefore on the difficulty of measuring such outcomes.⁴⁷³ Although USAID has the ability to commission evaluations or assessments at any time—including after a project has been completed—one senior evaluation advisor asserted that the dispersion of project staff to other organizations or countries can make ex post facto evaluations difficult.⁴⁷⁴

Finally, a level of participation and ownership can reduce defensiveness among implementing partners vested in seeing the success of their activities.⁴⁷⁵ As one evaluation expert put it, “I’m a strong believer in doing whatever we can to make the implementing partner feel invested and part of the process, not a passive victim of whatever the

external evaluation finds.”⁴⁷⁶ A hybrid model would be to use sector-wide evaluations, such as MISTI, that would look at overarching outcome indicators, while assigning activity-level evaluation to the implementing partner, subject to external verification from the mission’s M&E mechanism.

Timing Can Determine Whether Programmatic Adjustments Are Realistic

Part of the justification for a mid-term performance evaluation was to identify needed changes. However, if a mid-term evaluation occurred too late in the project cycle, making course-corrections was difficult, especially if they required time-consuming contract modifications.⁴⁷⁷ By the time the evaluation is produced, the project had moved on—and employees who had worked on the project under evaluation had either left for other activities or did not want to talk about their work.⁴⁷⁸

Especially in cases where there were delays in project startup or a time-consuming subcontract, midterm evaluations were considered premature; chronologically, it was a midterm, but programmatically it was not.⁴⁷⁹ A few implementing partners complained that the evaluation didn’t account for first year start-up difficulties—for example, new U.S. visa restrictions that undermined one project’s planned academic exchanges and scholarships.⁴⁸⁰

Additionally, finalizing an evaluation can be a lengthy process. As one evaluation expert explained, “People take issues with small items and there’s no big push to move the process along.” They added that their recommendation would be to move the approval process along more quickly—“otherwise [the evaluation] gets held up for no real reason.”⁴⁸¹

Given these various constraints, some respondents questioned their utility. “In general, evaluations are not designed in a way that allows for course corrections, and are limited in value,” said one former mission director.⁴⁸² However, conducting midterm evaluations more efficiently (and perhaps with greater consideration of delays in start-up activities), while streamlining the process by which evaluations are finalized, may be better solutions than eliminating them entirely.

USAID officials told SIGAR at the time that “in 2011, the U.S. government’s focus was on implementing projects as fast as possible.”⁴⁸³ SIGAR concluded that the agencies’ lack of performance metrics meant that taxpayers may never know whether the AIF projects would ever achieve their intended counterinsurgency effects.⁴⁸⁴

“The political demands on USAID were so high that the agency shifted an increasing part of its development effort into short-term, quick-response, quick-impact programs in conflict areas.”

— Former USAID/Afghanistan mission director

The pressure was not isolated to economic-growth and infrastructure projects. SIGAR’s lessons learned report on stabilization documented numerous cases where USAID was forced to take on activities against its will and better judgement. This included the Afghanistan Vouchers for Increased Production in Agriculture program, which had originally been designed to assist wheat farmers during the 2007–2008 drought, but was refocused and expanded in spring 2009 as a cash-for-work activity to support counterinsurgency activities in the contested Kandahar and Helmand Provinces. At the direction of then-Special Representative for Afghanistan and Pakistan Richard Holbrooke, the budget was doubled over the objections of USAID, which protested that spending more money would be “ineffective and wasteful.”⁴⁸⁵

Moreover, the short-term objectives of stabilization and counterinsurgency were often in tension with the longer-term ones of economic and social development. According to a former mission director, “the political demands on USAID were so high that the agency shifted an increasing part of its development effort into short-term, quick-response, quick-impact programs in conflict areas.” The former mission director cited four



An Afghan receives wheat seeds through USAID’s Afghanistan Vouchers for Increased Production in Agriculture program in 2010. (USAID photo)



Dr. Rajiv Shah, then-USAID Administrator, visits Afghanistan in 2010. (USAID photo)

programs in particular: Stabilization in Key Areas, Afghanistan Vouchers for Increased Production in Agriculture program, the Afghanistan Social Outreach Program, and the Strategic Provincial Roads Southern and Eastern Afghanistan program.⁴⁸⁶ He added, “USAID was asked to undertake a range of work it would normally not undertake—all on an urgent basis and with extraordinary scrutiny from several U.S. government inspectors general and from Congress.”⁴⁸⁷

Given its mission as a development agency, USAID’s relationship with the military and with larger security objectives was complicated. Many officials were reluctant to play a role in counterinsurgency, or to be associated with the military in a way that might jeopardize project security or long-term development activities. This resistance was mostly drowned out by political concerns: the need to be a team player in the integrated “whole of government” approach, and the need to answer criticism that USAID was not nimble enough or sufficiently focused on counterinsurgency.⁴⁸⁸ The official position was reflected in Administrator Rajiv Shah’s testimony to the Commission on Wartime Contracting: “In the most volatile regions of Afghanistan, USAID works side by side with the military, playing a critical role in stabilizing districts, building responsive local governance, improving the lives of ordinary Afghans, and—ultimately—helping to pave the way for American troops to return home.”⁴⁸⁹

The enormous political pressure to show results sometimes had negative consequences.⁴⁹⁰ A 2018 SIGAR audit of Promote, USAID’s largest and most ambitious project for Afghan women and girls, found that some of the target indicators were highly unrealistic—the result, according to an implementing partner official, of the need “to justify money to Congress.”⁴⁹¹ As one former senior USAID official put it, “You can’t go to Congress and say we want less money.”⁴⁹²

ARE THERE LIMITATIONS TO THE UTILITY OF USAID'S STRATEGIC REVIEW PROCESSES?

USAID policy states that the agency “has a responsibility to monitor strategy implementation progress, performance, and operational context” and that M&E should be “adequate to facilitate strategic learning.”⁴⁹³ Two of the most important USAID M&E processes that contribute to learning are a strategy’s annual portfolio review and its mid-course stocktaking exercise. According to the June 2019 version of the agency’s Automated Directives System, “Portfolio reviews are opportunities for missions to periodically examine all aspects of the mission’s strategy, projects, or activities. Missions must conduct at least one portfolio review per year that focuses on progress toward strategy-level results.”⁴⁹⁴ At least once during the course of a strategy’s implementation, missions must carry out a mid-course stocktaking exercise to better align programs with changes to context and with emerging knowledge and lessons learned.⁴⁹⁵

In theory, these are strong processes. In practice, many obstacles stand in the way of their full implementation. For example, USAID policy states that both portfolio reviews and mid-course stocktaking exercises should focus on learning from evaluations.⁴⁹⁶ But the clarity and analytical quality of USAID’s evaluations were not always strong, raising questions about whether they could consistently generate meaningful learning.⁴⁹⁷ Information gleaned from monitoring activities is another key input to strategic M&E, but monitoring was not always implemented as intended.⁴⁹⁸ Above all else, persistent pressure to demonstrate that progress was being made undermined evidence-based decision-making.⁴⁹⁹

The tension between the theory and practice of strategic-level M&E is in many ways reflected in USAID’s current Country Development Cooperation Strategy for Afghanistan. Published in 2018, the strategy opens with a letter from Herbie Smith, then USAID’s mission director in Afghanistan. The letter displays some signs of institutional learning, noting that USAID is “shifting away from quick-impact stabilization programs to focus on longer-term, broad-based development.”⁵⁰⁰ Although it describes several key obstacles to success, the letter’s tone is generally positive:

It asserts that the Country Development Cooperation Strategy would help the Afghan government progress toward self-reliance, despite the perennial challenges of corruption and dependence on donor aid.⁵⁰¹ It ends on a note of optimism, stating that USAID is “encouraged daily by the new generation of Afghans poised to take their future into their own hands.”⁵⁰² By endorsing the strategy, Smith signed off on a five-year, \$2.5 billion effort to continue development assistance to Afghanistan.⁵⁰³

However, shortly after the Country Development Cooperation Strategy was finalized and Smith retired, he immediately undercut both the strategy and the broader reconstruction effort in an op-ed arguing it was “time to leave” Afghanistan. He added that, despite “tons of development dollars . . . the reality is little has changed” and there was “little to show” for the effort.⁵⁰⁴ The op-ed went on to argue against specific aspects of the same Country Development Cooperation Strategy he had endorsed as USAID mission director, including the goal of “gender equality and female empowerment increased.”⁵⁰⁵ In the op-ed, Smith implied that “protecting the advancement of women” was not “sufficient to justify our continued engagement in Afghanistan,” because “there is scant evidence Afghanistan’s leaders pay much more than lip service to women’s equality issues.”⁵⁰⁶ And although the USAID strategy’s opening letter said that the agency “looks forward to continued collaboration with the Afghan people,” Smith’s op-ed asserted, “This is no longer our fight, and it is evident that we want peace and progress more than the Afghans. . . . The Afghan people need to take responsibility for their own destiny.”⁵⁰⁷

The unusual candor of a former official raises critical questions about the bureaucratic incentives surrounding strategic formulation and review, and, further, about the limitations of what M&E can reasonably accomplish. Smith concluded his op-ed by asking rhetorically, “Three U.S. administrations have tried unsuccessfully to find [a solution for Afghanistan]. The constant policy refrain is that we must remain to prevent a collapse of the country. Why?”⁵⁰⁸

Another problem, one former USAID mission director noted, is that M&E reporting is not linked with budgeting decisions. The earmarks and directives issued by the Congress “don’t necessarily correlate with a quality program. It can go either way: you can have a bad evaluation and end up having more money for an activity, or else you can have a good evaluation and end up with less money for the activity. Good mission directors can use appropriations and congressional priorities creatively. But money is not always aligned with a partner country’s priorities.”⁵⁰⁹ It was one of the perverse realities of the war, however, that money—at least, how fast it was spent—could become a proxy for “success.” Especially during the surge, when there was pressure to spend large amounts of money, the “**burn rate**” became an important metric for signaling to the embassy and to Washington that the USAID mission in Afghanistan was making progress.⁵¹⁰

Burn Rate: The Amount of Money Spent Became an Indicator of Progress

The term “**burn rate**”—sometimes euphemistically called the “pace of execution”—refers to the speed with which money can be spent. It was often used to describe a phenomenon in which agencies and implementing partners focused more on spending money than on delivering results.⁵¹⁴ In the words of a former USAID senior official, “there was the assumption that if you spend more money, you get more results. . . . The [United States] was under pressure to spend this money. Given this pressure, organizations and contractors choose sub[contractors] and grantees who shouldn’t have received the money.”⁵¹⁵ Equating burn rate to success resulted in projects that poured money into a fragile environment with no concept of whether those projects achieved any actual goals, or even necessarily where all the money was going.⁵¹⁶

Given the impossibility of monitoring all of the huge number of activities that were taking place, a bias towards positive news emerged, which made some people involved in the massive effort uncomfortable. A former USAID evaluation expert noted that in Afghanistan, the pressure to show promising metrics was so intense, some people refused to work there.⁵¹¹ As one evaluation contractor asserted, “The major challenges have been speaking truth to power, as in presenting results that [program funders] do not want to hear, and the security challenge. There’s a conscious bias for positive news.” This bias, combined with insecurity that meant some areas were inaccessible, could cause “profound and sometimes deliberate obfuscation of results.”⁵¹² A chief of party of one USAID implementing partner asked rhetorically, “How does something look on paper versus reality?”—adding that upper management relied “on phones with Skype capacity, USAID spot checks, and Checchi compliance people to know what’s going on. But people are smart, and they have tricks.”⁵¹³

Political sensitivity extended even to using language that might unsettle the Afghan government or could imply that USAID was supporting bad or compromised actors. For example, one evaluation consultant reported that a draft governance-sector assessment used the term “mafia” to describe the Afghan government—a use of the term that, according to the consultant, conformed to how many observers spoke about the government. The consultant said that USAID stripped the word from the assessment,

and that the term had been retained in the final report only by presenting a direct quote from an Afghan source using the term rather than directly attributing the view to the assessment team itself.⁵¹⁷

“The major challenges have been speaking truth to power, as in presenting results that [program funders] do not want to hear, and the security challenge. There’s a conscious bias for positive news.”

—Evaluation Contractor

Complexity and Factors beyond Anyone’s Control

In many cases, factors beyond the control of the implementing partner or even USAID made meeting the performance indicators in the **activity monitoring, evaluation, and learning plan** difficult, and the inflexibility of contracting made adjusting to those changes even harder. As one respondent put it, “In the field, sometimes information is seen as a negative thing if an individual doesn’t have the power to act on it. Most people see ignorance as better than knowing about a problem and not acting, but sometimes there is no mechanism to address the problem.”⁵¹⁸

Defining Key USAID Terms: Activity, Monitoring, Evaluation, and Learning Plan

An **activity monitoring, evaluation, and learning plan** describes the monitoring approaches to be applied during activity implementation and defines relevant performance indicators consisting of outputs and outcomes.⁵²⁷ The plans fulfill many of the same functions as performance management plans, though activity MEL plans are applied to individual awards, as opposed to a strategic portfolio of contracts. Pages 204–205 of Appendix B discuss activity monitoring, evaluation, and learning plans in more detail.

Changes in the external environment sometimes rendered the original activity design obsolete, which meant that results and indicators that had been set out in the original request for proposals were irrelevant by the time work actually began.⁵¹⁹ As one M&E expert put it, in such a case, implementing partners “were implementing things that didn’t exist.”⁵²⁰

One example of events outrunning program design was the Strong Hubs for Afghan Hope and Resilience (SHAHAR) project, designed to support accountable and effective municipal-level governance.⁵²¹ The project began implementation in November 2014, at a time when mayoral and municipal council elections required by the Afghan constitution were expected to take place the following year.⁵²² Implementation of the project reflected USAID’s hope that “municipalities will become laboratories for localized democracy.”⁵²³ However, even by the time a June 2016 midterm evaluation of the project was finalized, municipal elections had not been held.⁵²⁴ Evaluators euphemistically noted that because the anticipated elections did not take place, SHAHAR was ultimately implemented “in a different context than originally envisioned.”⁵²⁵ In other words, a project that aimed to support locally accountable governance was instead applied to officials who were “appointed by, and answer[ed] to, authorities in Kabul.”⁵²⁶



Workers assemble a transmission line tower. The Engineering Support Program provided construction management services and quality assurance for many of USAID's power-infrastructure projects. (USAID photo)

Similarly, the University Support and Workforce Development Program, awarded in December 2013, had envisioned partnering with U.S. institutions through exchanges and scholarships.⁵²⁸ Not long after implementation started in 2014, USAID stopped sending Afghan faculty to the United States for advanced degrees.⁵²⁹ The reason: U.S. government policy changes made it nearly impossible for Afghans to acquire visas.⁵³⁰ Additionally, the project was negotiated at the end of President Karzai's administration, and the subsequent National Unity Government decided that the associate degrees that the project was going to provide were inadequate.⁵³¹ Such factors were beyond the project's control, but the evaluation was still critical of its failure to achieve results.⁵³²

In other cases, the lack of clarity on division of responsibilities between multiple parties complicated evaluation and accountability. For example, USAID's Engineering Support Program, intended to provide construction management services for infrastructure projects, received a highly critical midterm evaluation which found that the project had not been responsive to USAID, the Afghan government, or to private sector contractors.⁵³³ Specifically, it found weaknesses in capacity building of Afghan government staff, subpar implementation of projects, and foreseeable and preventable delays in implementation of some job order deliverables.⁵³⁴ The implementing partner, Tetra Tech, argued that it had been put in an impossible position: caught between USAID, the Afghan government, and the contractors, with maximum responsibility but minimal authority. Tetra Tech was essentially functioning as USAID's quality control agent, yet appeared to have little leverage over the relevant Afghan government institutions. It was also not clear who was

USAID'S POWER TRANSMISSION EXPANSION AND CONNECTIVITY PROJECT: A POSTER CHILD FOR FAILED M&E

The Power Transmission Expansion and Connectivity project, which began in 2011, was initially imagined as a \$861.7 million **on-budget** effort that would be implemented by Da Afghanistan Breshna Sherkat (DABS), the Afghan government's national power utility.⁵³⁵ Phoenix Information Technology Solutions Ltd. was contracted by DABS to provide equipment and technical assistance.⁵³⁶ The complex project had three components: infrastructure construction, commercialization of DABS operations, and clean energy (the last was later eliminated).⁵³⁷

In early 2018, USAID, citing credible allegations of serious corruption in procurement involving DABS' then-chief executive officer, shifted nearly \$400 million **off-budget**—essentially, taking that money out of the control of the Afghan government.⁵³⁸ However, USAID did not cut funds for several significant power-infrastructure contracts that had already been awarded by DABS, including a transmission line and a substation—leaving \$316.7 million at significant risk because of DABS' demonstrated institutional vulnerability to corruption.⁵³⁹

A 2019 SIGAR audit was highly critical of the project. To begin with, it noted that PTEC's power infrastructure activities were significantly delayed—several so much so that some of the funding attached to them had expired.⁵⁴⁰ It also questioned whether they were sustainable, noting that USAID had not conducted sustainability assessments for most of the infrastructure to be constructed under PTEC. Those sustainability assessments USAID did conduct relied on faulty assumptions—such as that Afghanistan would increase its supply of energy to Kabul and that DABS's ability to operate and maintain new infrastructure would live up to USAID's predictions.⁵⁴¹ Additionally, the audit criticized USAID's limited oversight.⁵⁴² In particular, USAID's on-budget monitors were unaware that Phoenix Solutions contractors embedded in DABS's finance department were approving Phoenix's own invoices.⁵⁴³ Across DABS, the frequency with which Phoenix contractors performed core tasks prompted one USAID inspector general investigator

Callout Box: On-budget and Off-budget Assistance

On-budget assistance refers to donor funds that are aligned with Afghan government plans, included in Afghan government budget documents, and included in the budget approved by the parliament and managed by the Afghan treasury system. On-budget assistance is primarily delivered either bilaterally from a donor to Afghan government entities, or through multi-donor trust funds, such as the Afghanistan Reconstruction Trust Fund.

Off-budget assistance refers to donor funds that are excluded from the Afghan national budget and not managed through Afghan government systems.⁵⁴⁸ Most programs and projects discussed in this report are off-budget.

to comment that Phoenix was “essentially running DABS, as it lacked any capacity to function on its own.”⁵⁴⁴

SIGAR's audit also cited a number of violations of USAID's ADS requirements related to project planning, implementation, and M&E. More significantly, the audit found that USAID could not monitor or evaluate PTEC's performance because the majority of indicators were dropped, other indicators lacked baselines or end-of-project targets, and other data was simply missing.⁵⁴⁵ Other indicators were not measureable because they were predicated on intermediary progress, such as the installation of electric meters. When that failed to happen, there was no way to assess the number of megawatt hours of electricity supplied or DABS revenue loss reduction.⁵⁴⁶ A senior official at Tetra Tech, which was responsible for providing third-party quality assurance for the project, expressed concerns that DABS and its contractors were reporting misleading information to USAID. The official attributed the inaccurate reporting in part to pressure from USAID.⁵⁴⁷

working for whom; according to Engineering Support Program staff, “DABS [the Afghan national electrical utility] thought that we were their engineers.”⁵⁴⁹

Implementing partners reported varying experiences when changing conditions required a change in their contracts. Substantive changes usually required modifying the project’s results framework and the Activity Monitoring Evaluation and Learning Plan. One implementing partner chief of party noted that the results framework for the project for which they were responsible was “impossible” because the environment had changed since the project was designed and the population data on which planned health activities were based was wrong. However, they added, the implementing partner was able to reformulate the results framework in a contract modification.⁵⁵⁰ Another noted that “results frameworks are too rigid,” and that often working through changes in contracts takes too much time.⁵⁵¹ One respondent observed that in some cases, contracting officers were reluctant to change scope within one year of the award start date for fear of prompting a protest from a losing bidder accusing USAID of “bait and switch” in the procurement. This was especially the case with younger, inexperienced contracting officers, who were typically less confident and more concerned about maintaining a clean reputation.⁵⁵²

As a senior evaluation expert and former USAID consultant said, “USAID’s [monitoring, evaluation, and learning] system is conceptually strong and thoughtful. The awareness and thinking on the type of monitoring and the complexity needed is thoughtful and good. Where it breaks down is when you deal with complex things. . . . It breaks down when you try to implement it.”⁵⁵³

High Workloads, Constant Turnover, and Not Enough People Made Rigorous M&E Difficult

At this point, it is a truism that a constant churn in personnel has led to shortcomings in the delivery of reconstruction assistance and its oversight. Every SIGAR lessons learned report, as well as the numerous other reports referenced in Chapter 1, is replete with examples of this. Rapid personnel turnover also affected the ability of contracting officer’s representatives to perform oversight and management functions, as did limited technical knowledge and experience, and general overwork.

Contracting officer’s representatives generally oversaw a large portfolio of activities, and were typically too overloaded to advocate for individual projects or programs. According to one evaluation consultant, “Their compliance responsibilities are so significant and time consuming that they don’t have a chance to step back and manage.”⁵⁵⁴ An implementing partner staff person noted that they would like to meet weekly with their project’s contracting officer and contracting officer’s representative, but due to workload at USAID, the meetings were held monthly.⁵⁵⁵ The same individual added that their current project produces approximately 12,000 pages of documentation per month, leaving no time to examine M&E data.⁵⁵⁶

A 2012 SIGAR audit of the Local Governance and Community Development project underscored the magnitude of the challenge. The audit noted the volume of

administrative paperwork performed by the project's contracting officer's representative, including certifying that contractor services had been provided, submitting payment forms to USAID, and reviewing invoices and their often voluminous supporting documentation. In theory, certifying services would require traveling to project sites; for the Local Governance project, there were 608 subprojects in 22 provinces. The administrative checklist for the contractor's August 2010 invoice potentially covered up to 214 subprojects. Clearly, on-site visits for all of these were impossible.⁵⁵⁷ Meanwhile, a 2011 USAID presentation called into question the project's underlying assumptions, referring to them as "myths."⁵⁵⁸ This raised questions about whether the administrative requirements described above were reducing the amount of time available for sober assessments of efficacy and, in so doing, increasing the risk of doing the wrong thing perfectly.

Overwork was exacerbated by rapid turnover, as contracting personnel were often learning their jobs. One implementing-partner chief of party reported having had a new agreement officer's representative "every eight to 10 months."⁵⁵⁹ To mitigate the effects of the typical one-year turnover of contracting officer's representatives, USAID turned to appointing Afghans. However, in recent years, emigration of Afghan employees who met the requirements of the Special Immigrant Visa program made that less effective, as Afghans rationally used the program to relocate to the United States. USAID estimated the annual turnover of Afghan contracting personnel at 25 percent.⁵⁶⁰

Short rotations also resulted in a lack of continuity and ownership. As one USAID official noted, "When staff often inherit mechanisms from predecessors, and no one who originally designed the mechanism [or] award is present throughout the entire project, staff often don't have the insight to own or implement midterm or final evaluation recommendations. Because of this, there is no sense of ownership of a mechanism [or] award on the part of USAID."⁵⁶¹ According to implementing partner staff, this explained part of the problem with the Monitoring Support Program, described earlier in this chapter: The program had been designed by staff who had left their posts by the time of the award, and the new staff did not understand the rationale for the program's structure.⁵⁶²

As noted by one chief of party, changes in contracting officers and their representatives "can be disruptive. The [implementing partner or] contractor often has to familiarize the new [contracting officer's representative] with the contract, which could take three to four months." More importantly, another chief of party noted, this meant a risk of "continual redesign."⁵⁶³ In some cases, implementing partners attributed modifications and changes directly to turnover in the contracting officer's representative. For example, IDEA-NEW, a previously mentioned program aimed at reducing poppy cultivation, had an initial focus on food security and wheat production. When a new contracting officer's representative arrived, the program's focus shifted.⁵⁶⁴

One international development professional shared that the most effective contracting officer's representatives were those with enough experience and technical knowledge to



A poster explains USAID's Incentives Driving Economic Alternatives for the North, East and West program. (USAID photo)

be flexible and decisive. The more senior, they added, were more open to having a frank and open discussion about what was not working and how to change it; inexperienced contracting officers and contracting officer's representatives tended to be afraid of taking risks.⁵⁶⁵ According to a former senior USAID official, "it takes seven years to master the FAR [Federal Acquisition Regulation] and develop a good procurement officer."⁵⁶⁶ Another former USAID official added:

Remote management coupled with junior employees was a big problem. You can't successfully manage and evaluate a program if you haven't actually seen the country, or the program operating in it, [because] you don't know the right questions to ask. It was like taking someone who had never seen an elephant before, blindfolding them, and asking them to assess a real elephant based only on what they'd learned secondhand.⁵⁶⁷

CHAPTER SUMMARY AND KEY FINDINGS

Although USAID policies set a high bar for M&E within the U.S. government, the agency struggled to find ways to measure results and to balance political demands to show success with the principles of accurate and honest reporting. Many respondents and reports were skeptical about the extent to which the agency was using M&E information to make decisions, at least on a scale commensurate with the expense and time spent collecting that information. It was clear that the mission typically made decisions based on a variety of information sources, many of them informal. At the same time, the agency was constrained both by external political pressures and by internal strains on human resources.

Key Findings

- USAID faced the challenge of selecting the right indicators with which to measure progress. This was a difficult task, particularly for objectives that defied easy measurement, such as stability. As a result, it often used proxy indicators that were not necessarily relevant to projects' objectives.
- Faced with political pressure to show results, USAID displayed a tendency to overattribute impact to its activities and to rely on unrepresentative success stories to show progress.
- USAID struggled to determine how much information to collect, and sometimes ended up collecting too much. This burdened the agency and its implementing partners with unnecessary reporting requirements.
- In response to changing conditions and external scrutiny of its activities, USAID introduced a number of innovations for M&E (most prominently, multitiered monitoring and third-party monitoring), which met with varied levels of success.
- Among other factors, political pressure and the hectic pace of activity made it difficult to make evidence-based decisions.
- It was difficult to hold individuals or implementing partners accountable for activities that might no longer make sense in a changed environment, where multiple institutions were jointly responsible for activities, or where there were factors beyond the control of either USAID or its implementing partner.
- A constellation of human resource and organizational issues made it more difficult to use information. These included overworked and inexperienced contracting staff, frequent personnel turnover, and the lack of incentives or requirements to use M&E information.

But by 2009, as the war entered its eighth year and a new administration refocused on Afghanistan, policymakers in Washington intensified demands for tangible measures of results for the billions of dollars being spent.⁵⁷⁷ Concurrently, pressure was growing to submit U.S. foreign assistance as a whole to more extensive M&E: A 2010 presidential directive on global development stated that the U.S. would “undertake a more substantial investment of resources in monitoring and evaluation, including . . . a focus on rigorous and high-quality impact evaluations.”⁵⁷⁸

At the time of the directive, M&E at State was still relatively informal and unstandardized. For example, a 2010 Government Accountability Office report on U.S. drug control efforts in Afghanistan noted that although State program officers were said to have “conducted informal evaluations of [justice reform] program activities, they did not document any of these evaluations.”⁵⁷⁹ By December 2011, State had issued department-wide performance management guidance.⁵⁸⁰ However, its use was not mandatory and the unstandardized nature of M&E at State up to then meant that the terms, policies, and procedures in department-level guidance were at times misaligned with bureau-level guidance.⁵⁸¹ In a gradual move towards systemization, State tightened its M&E systems over the next five years, releasing new M&E-related policies in 2012, 2015, and 2017.⁵⁸²

State’s 2012 program evaluation policy was modeled on USAID’s.⁵⁸³ The new policy required bureaus to submit a clear statement of strategic goals, and to develop plans for assessing whether programs were achieving them.⁵⁸⁴ To help bureaus fulfill these requirements, State published guidance, and offered training and technical assistance.⁵⁸⁵ The results were mixed. Of the 39 State bureaus the policy covered, 16 had not fully complied with evaluation requirements by the end of fiscal year 2014.⁵⁸⁶ Of the bureaus which did comply, only half produced evaluations of “high” or “acceptable” quality, according to a GAO report.⁵⁸⁷

In 2015, State rolled out the core of its current M&E system, the Managing for Results framework. Created by State’s Office of U.S. Foreign Assistance, the framework aimed to formally link strategy, resource management, and program design, with monitoring and evaluation integrated into each step.⁵⁸⁸ In November 2017, State published an update to its M&E policy, with an emphasis on further integrating the tenets of the framework into program design and strategy.⁵⁸⁹

The 2017 policy further synchronized State’s M&E practices with USAID’s. During the program design stage, bureaus were required to align programming to a strategy, create program goals that advanced this strategy, and articulate a specific theory of change describing how program activities would achieve strategic goals. Theories of change were to be captured in logic models that linked program inputs with desired outputs and outcomes. Bureaus were required to develop monitoring plans for their programs and projects that described how data would be collected and used to track progress against key performance indicators. All of these requirements were to be captured in program performance management plans. Cumulatively, these steps were the foundation for State’s more institutionalized M&E system.⁵⁹⁰

State's investment in improved M&E is laudable. Over time, further use and refinement of the Managing for Results framework could impose greater discipline on program design, implementation, and oversight. Nevertheless, throughout the reconstruction effort, State relied too heavily on output metrics to gauge success, struggled to effectively implement important aspects of its M&E policies, generally lacked an understanding of what impact its programs were having, and did not use M&E in ways that were conducive for decision-making.⁵⁹¹ In sum, State's M&E apparatus evolved and improved over time, but is still not fully achieving its purpose.

As with the chapter on USAID, this chapter is structured around two key aspects of M&E: finding the right metrics to assess progress, and acting on information provided by monitoring and evaluation. It begins with brief summaries of what State was attempting to achieve at the programmatic level, and is organized principally around core findings.

STATE RECONSTRUCTION PROGRAMMING IN AFGHANISTAN

In 2001, the United States entered Afghanistan with the goals of eliminating al-Qaeda, overthrowing the Taliban regime, and preventing Afghanistan from being a sanctuary for terrorists that could attack the U.S. homeland.⁵⁹² To accomplish those goals, the administration of President George W. Bush sought to strengthen the Afghan state.⁵⁹³ As DOD dealt with warfighting and training the Afghan security forces, and USAID addressed economic development, State focused its programming on several key areas: **rule of law**, counternarcotics, conventional weapons destruction, refugee assistance, counterterrorism, and women's empowerment. These early lines of effort would remain the primary avenues of State programming throughout reconstruction.⁵⁹⁴

The largest funding sources of State programming are presented in Table 4 below. Brief descriptions of each line of effort follow after the table.

TABLE 4

MAJOR STATE DEPARTMENT FUNDS		
Name of Fund	Appropriated Amount (millions)	Description
Economic Support Fund (ESF)	\$21,100.9	Intended to meet short- and long-term political, economic and security needs in Afghanistan. Funds both State and USAID projects.
International Narcotics Control & Law Enforcement (INCLE)	\$5,446.2	Funds projects and programs for advancing the rule of law and combating narcotics production and trafficking.
Migration and Refugee Assistance (MRA)	\$1,536.9	Funds programs that protect and support refugees, internally displaced persons, stateless persons and vulnerable migrants.
Non-Proliferation, Antiterrorism, Demining & Related (NADR)	\$881.3	Develops the Afghan government's capacity to address terrorist threats, protect international borders, and remove explosive remnants of war.

Note: The majority of funds obligated through ESF flow to USAID. The Afghanistan Infrastructure Fund (AIF) is not listed above because funds for AIF were appropriated through DOD's budget.

Source: CRS, "Security Assistance and Cooperation: Shared Responsibility of the Departments of State and Defense," R44444, May 26, 2016, p. 49; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, pp. 42, 44–46, 176.

Rule of law is the principle that persons, institutions, and entities, public and private, including the state itself, are accountable to laws that are publicly promulgated, equitably enforced, and independently adjudicated, and which are consistent with international human rights law.

SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 2.

Rule of Law: Programmatic Effects May Be Achieved Only If Political Will Materializes

State's Bureau of International Narcotics and Law Enforcement Affairs has worked to develop rule of law in Afghanistan since 2003.⁵⁹⁵ INL's most significant rule of law effort, the Justice Sector Support Program, was established in 2005 to provide training, mentoring, and advisory services to the Afghan justice system.⁵⁹⁶ The second-largest rule of law program, the Corrections System Support Program, began shortly after, providing advising, training, and infrastructure to create a "safe, secure, and humane" prison system.⁵⁹⁷

In September 2009, State issued its first rule of law strategy, which emphasized reducing corruption in the Afghan government.⁵⁹⁸ The strategy focused on preventing funds from going to the insurgency, boosting the legitimacy and viability of the Afghan state, and improving the oversight of contracting and development assistance.⁵⁹⁹ In 2013, the U.S. embassy in Kabul drafted a Rule of Law and Law Enforcement Framework to replace the 2009 strategy, toning down program expectations and goals in light of an uncertain post-2014 military presence. The framework suggested no specific measures of success.⁶⁰⁰ State's 2018 **integrated country strategy** had the goals of creating an "accountable government" and "[strengthening] law enforcement capacity."⁶⁰¹

Despite a significant commitment of resources, materially improving Afghan rule of law has proved very difficult.⁶⁰² So too was ascertaining the extent to which U.S. programs may have contributed to the effort: A 2015 SIGAR audit of U.S. rule of law programming costing more than \$1 billion concluded that every agency involved in the sector had problems with performance measurement and therefore could not fully determine the effectiveness of their programs.⁶⁰³ Despite significant spending, it was unclear what overall outcomes and impacts resulted.⁶⁰⁴

Since 2015, U.S. rule of law programming has continued to face substantial challenges. In 2018, the Department of Justice described the situation in Afghanistan as "consistent with a largely lawless, weak, and dysfunctional government."⁶⁰⁵ A 2020 assessment by **Freedom House** summarized the many obstacles to successfully promoting rule of law in Afghanistan. Although Afghanistan's constitution sets a foundation for rule of law, Freedom House said, "political rights and civil liberties are curtailed in practice by violence, corruption, patronage, and flawed electoral processes."⁶⁰⁶ Any eventual improvement may depend on factors outside the control of specific programs: A 2019 SIGAR audit found that transformative change of the justice system will remain elusive until the Afghan government develops sufficient political will to address resource shortfalls at anti-corruption institutions and the impunity of powerful individuals.⁶⁰⁷

Counternarcotics: Attempts to Curb the Opium Trade Appear to Have Failed

Since 2002, stemming illicit drug production in Afghanistan has been an important U.S. goal. The drug trade has provided funding to insurgent groups, fueled government corruption, and eroded state legitimacy.⁶⁰⁸ State is responsible for coordinating all international drug control programs implemented by the U.S. government, and INL

An **integrated country strategy** is a four-year strategic plan that articulates whole-of-government priorities in a given country and incorporates higher-level planning priorities.

State, 18 FAM 301.2 Department of State Strategic Planning, May 2019, pp. 1–2.

Freedom House is a nonprofit organization that researches democracy, political rights, and civil liberties. Its flagship publication, *Freedom in the World*, presents a global, comparative assessment of these themes.

Freedom House, "About Us," accessed September 16, 2020.



Participants listen at a criminal law workshop in Nangarhar Province. (USAID photo)

is the primary entity within State responsible for formulating and implementing international narcotics control policies.⁶⁰⁹

In 2003, INL began counternarcotics programming with small projects aimed at introducing marketable crops as alternatives to poppy.⁶¹⁰ In 2005, INL developed a five-pillar strategy to counter the effects of narcotic production and trafficking in Afghanistan: mobilizing government and public support, alternative development programs, eradication, interdiction, and building law-enforcement and judicial capacity.⁶¹¹ While these lines of effort have served as organizing categories for programming to the present, emphasis has shifted between them as conditions and decision makers changed.⁶¹² Over time, the focus on counternarcotics faded: The 2013 Civilian-Military Strategic Framework for Afghanistan included only a passing reference to the subject, and several counternarcotics programs quietly concluded.⁶¹³ Today, State's counternarcotics programming is guided by the 2018 Integrated Country Strategy, which does not emphasize counternarcotics or provide any metrics towards reducing the drug trade. Instead, its goals are "strengthening institutional border security and law enforcement mechanisms and bodies."⁶¹⁴

State has acknowledged it has not been fully successful in achieving counternarcotics objectives.⁶¹⁵ In a 2018 letter to the Congress, SIGAR stated that the entirety of programming dedicated to counternarcotics appeared to have failed.⁶¹⁶ A 2020 SIGAR special project reaffirmed this conclusion and added \$1.6 billion to the total amount of waste—meaning that nearly all of the \$8.9 billion allocated for counternarcotics programming since 2002 has been wasted, according to SIGAR's analysis.⁶¹⁷ According to the United Nations, the area under opium poppy cultivation in Afghanistan in 2020 was

among the four highest levels ever measured.⁶¹⁸ Although yearly changes in cultivation have been volatile due to a variety of factors—including insecurity, droughts, and floods—opium poppy cultivation has generally increased since 2010.⁶¹⁹

Refugee Assistance Faced an Uphill Struggle

Between 2002 and 2010, more than 5.6 million Afghans returned to Afghanistan, increasing its estimated population by more than 20 percent.⁶²⁰ To provide lifesaving assistance to refugee returnees and to offset the inability of the Afghan government and economy to provide services and livelihoods to this population, State's Bureau of Population, Refugees, and Migration (PRM) allocated funding to international organizations such as the UN High Commissioner for Refugees and the International Organization for Migration to meet such basic needs as water, shelter, health services, and education, as well as "gap" programming through various Afghan NGOs to provide critical assistance to vulnerable refugee populations.⁶²¹

The **UN High Commissioner for Refugees** is the United Nations body charged with protecting and assisting refugees. It provides shelter, blankets, and clean water, among other goods and services, to refugees.

UN High Commissioner for Refugees, "What We Do," n.d., accessed April 2020.

Refugees are persons who are outside their country of origin for reasons of feared persecution, conflict, generalized violence, or other circumstances that have seriously disturbed public order and, as a result, require international protection.

Migrants are persons who change their country of usual residence, regardless of the reason for migration or legal status. According to the UN, there is no formal legal definition of an international migrant.

UN, "Refugees and Migrants: Definitions," 2019; UN High Commissioner for Refugees, "Protecting Refugees: questions and answers," February, 2002.

In the first years after the initial defeat of the Taliban—from 2002 to 2005—**refugees** and **migrant** returnees were reintegrated into their local communities with few major difficulties. However, more recent returnees have had increasingly less land, shelter, livelihoods, and family to which to return.⁶²² The Afghan government developed the Land Allocation Scheme initiative in 2005–2006 to address the needs of landless returnees; however, this Afghan-led initiative has suffered from mismanagement and corruption and has been largely ineffectual in providing returnees with land capable of sustaining a basic livelihood.⁶²³ The UN High Commissioner for Refugees withdrew financial and technical support to the Afghan Land Allocation Scheme, and is now providing assistance directly to the returnees living at the sites. PRM used some of its grants program to mitigate the humanitarian situation at these sites.⁶²⁴

From 2007 to 2010, PRM awarded 57 grants to various NGOs, totaling \$51 million, to provide gap-filling assistance to meet needs, including skills training, basic education, shelter, water and sanitation, gender-based violence programs, income generation, health, and agriculture.⁶²⁵ From 2012 to 2014, PRM's two-year capacity-building program, which was carried out through an implementing partner, intended to develop capacity within Afghanistan's Ministry of Refugees and Repatriation.⁶²⁶ The program was hampered by an "extremely challenging" working relationship with the ministry under its prior leadership.⁶²⁷ While humanitarian funding is not intended to change the government, a lack of Afghan government capacity and will limited the sustainability of all refugee programming.⁶²⁸ As one PRM official said in 2013, refugees "are not a priority issue" for the Afghan government."⁶²⁹

PRM continues to work directly with implementing partners to assist refugees and returnees. However, challenges are growing: PRM has told SIGAR that the combined effects of COVID-19 and economic contraction have led to high numbers of returns of Afghan migrant laborers from Iran.⁶³⁰ According to the International Organization for Migration, the number of Afghan migrants returning from Iran in 2020—more than 850,000—was the highest ever recorded.⁶³¹



Mine clearance workers inspect a field. (State photo)

Conventional Weapons Destruction: Gains have been Overshadowed by the Ongoing Conflict

State's mine action and weapons removal programming is handled by the Bureau of Political-Military Affairs Office of Weapons Removal and Abatement, which uses contractors and NGOs for clearance operations, victim assistance, and education.⁶³² Current programs are primarily focused on clearance operations, with programming reduced in scale since the drawdown of coalition forces in 2014.⁶³³ Gains from State's conventional weapons destruction programming have generally been overshadowed by the ongoing conflict: Casualties from mines and unexploded ordinance increased from 36 per month in 2012 to 130 per month by 2019 (the last year for which data was available).⁶³⁴

FINDING THE RIGHT METRICS TO ASSESS PROGRESS: MEASURING THE MEASURABLE

Afghanistan is a complex environment where program-level progress is sometimes difficult to assess. The Standard Foreign Assistance Indicators—developed by State in 2006 in an effort to measure the aggregated effects of foreign assistance, and in use by 2008—provided a menu of metrics that could be used to track program progress.⁶³⁵ Bureaus and offices could also create custom indicators specific to individual programs.⁶³⁶ But whether the indicators were standardized or customized, they tended to focus on outputs and often defaulted to measuring that which could be measured most easily. Moreover, they were not always clearly connected to programmatic goals.

A close examination of an individual program's performance metrics illustrates the pattern. State's Legal Aid through Legal Education program aimed to improve justice services for Afghan women and vulnerable populations.⁶³⁷ State intended to

accomplish this goal by opening legal clinics, increasing the quality of legal education at Afghan universities, and increasing the number of trained lawyers in the justice system.⁶³⁸ According to the program's fiscal year 2018 performance management plan, performance metrics included at least one standard indicator—the “number of individuals/groups from low income or marginalized communities who received legal aid or victim's assistance with [U.S. government support]”—and numerous custom indicators, such as the “number of [legal] course packets published.”⁶³⁹ Although such indicators allowed for tracking of program activities, they did not necessarily correlate with an improved legal system that better met the needs of Afghan women and other vulnerable populations.

The program's performance management plan tied specific indicators to discrete program objectives. But a rise in the values of many indicators would not necessarily have been associated with achievement of those objectives. For instance, the number of “adjunct [legal] instructors trained,” was used to track progress toward the program objective of “improved capacity of adjunct instructors.”⁶⁴⁰ However, if the quality of training was poor, the number of instructors trained could increase without improving their capacity. Likewise, the indicator “number of [public legal awareness] campaigns on legal aid produced and aired” was used to track progress toward the program objective “public legal awareness needs identified and improved.”⁶⁴¹ But public legal awareness campaigns could be produced and aired without actually identifying and improving public legal awareness needs. Although a separate indicator, the “number of client referrals as a result of [public legal awareness] initiatives,” opened an avenue to track whether increased public legal awareness was being achieved, it lacked a target value, which reduced its utility.⁶⁴²

In other cases, metrics would have been useful to track achievement of program objectives, but were not actually tracked in a way that accurately reflected what the indicator claimed to capture. For example, the “number of partner organizations' capacity built” was associated with the program objective “enhanced capacity of legal advice providers.” However, the indicator value presented in the performance management plan was not a measure of capacity. Rather, it was simply the number “4,” referring to the number of partner organizations receiving technical assistance through the program.⁶⁴³ Although the previous year's performance management plan noted that periodic assessments would be used to track organizational capacity, the fact that the baseline and target values for the indicator (both the number 4) were identical, even as the assistance was still being provided, raised questions about the usefulness of the metric.⁶⁴⁴ This was not the only purported outcome indicator tracked via a numerical value that was actually just an output. The “quality of legal representation” was measured merely by the number of clients at legal aid clinics receiving U.S. government assistance.⁶⁴⁵ Meanwhile, one objective of the program was to “create sustainability” for the program's initiatives, but none of the indicators explicitly tracked sustainability.⁶⁴⁶

In sum, many of the indicators presented in the Legal Aid through Legal Education program's objectives were conflated with the act of implementation. Table 5 on

Anecdotes Can Supplement Raw Indicators, but Have Their Own Limitations

Anecdotes can add nuance and descriptive detail to complex, abstract issues. They may be particularly useful for policymakers in the Congress and for other decision makers who have little time to digest significant amounts of data. However, when anecdotes suffer from selective presentation of facts, they can distort reality.

For example, a December 2018 State briefing to a congressional delegation noted that the Afghan government had recovered almost \$500 million in stolen assets from Kabul Bank—a story that implied the Afghan government was moving aggressively to fight corruption.⁶⁴⁹ A closer look, however, told a different story. Asset recovery reports showed that between September 2017 and September 2018, recovered assets rose from \$447.35 million to \$451.65 million, an increase of less than 1 percent.⁶⁵⁰ Not only was the \$500 million reported in the congressional briefing a sizeable overstatement, it was a total, cumulative amount that hid the marginal year to year improvement. In fact, at that recovery rate, it would have taken more than 125 years to recover the rest of the stolen assets.⁶⁵¹ The \$500 million was an indicator of failure, not success.

Also missing from the picture was the Afghan government's limited ability to independently prosecute corruption. A June 2018 Department of Justice report summarized the situation as “consistent with a largely lawless, weak, dysfunctional government,” with many corruption cases languishing due to lack of political will.⁶⁵² Not only did the congressional briefing point fail to accurately frame the Kabul Bank case itself, it also painted an incomplete picture of the state of rule of law in Afghanistan.

pages 86–87 presents a critical analysis of all indicators listed in the Legal Aid through Legal Education program's fiscal year 2018 performance management plan. The central takeaway is that State's M&E system sometimes produced complicated tables and spreadsheets that looked impressive but did not meaningfully track program progress.

Of course, such indicator issues were not unique to the Legal Aid through Legal Education program. As was the case with counterpart agencies, metric selection within State's M&E system often risked creating **goal displacement**, as quantifiable measures substituted for the actual outcomes programs were intended to achieve. State's Justice Sector Support Program aimed to build rule of law in Afghanistan. SIGAR criticized the program for relying too heavily on short-term performance outputs such as the number of courses conducted and the total hours implementing-partner staff spent advising trainees, and for failing to look to long-term performance outcomes that could help determine the program's overall effects on the development of justice sector.⁶⁴⁷ Tracking those effects may have been challenging. But even after the program established outcome indicators related to, for example, increasing “efficiency and effectiveness of the justice and corrections systems of Afghanistan,” the program's quarterly reporting did not always provide information on these new indicators.⁶⁴⁸ Similarly, a 2014 State inspector general audit of INL's assistance to Afghanistan found that of the seven

Goal displacement is a scenario in which the original goals of a strategy, organization, or project become displaced by other goals that are not necessarily related.

TABLE 5

CRITICAL ANALYSIS OF PERFORMANCE INDICATORS FOR STATE'S LEGAL AID THROUGH LEGAL EDUCATION PROGRAM, AS PRESENTED IN THE PROGRAM'S FY 2018 PERFORMANCE MANAGEMENT PLAN

Program Objective	Performance Indicators	Target	Baseline (if available)	SIGAR Observations
Practical legal education provision improved for law students and sharia students of the targeted Universities	Percentage of law students and sharia students with improved practical legal skills			Although LALE's performance management plan lists four distinct indicators for this program objective, only two appear to have baselines and targets. The indicators lacking baseline and target values (listed second and third, respectively) do not include an explanation of how they are fully distinct from the first indicator. Additionally, the value of the indicator "number of law students and sharia students attending supplementary practical skills courses" can increase without necessarily contributing to the program objective.
	Increased knowledge and practical skills of law graduates	30–40%	5–10%	
	Quality rating of legal teaching, staff, research, curriculum, and student learning outcomes, disaggregated by university and continuing			
	Number of law students and sharia students attending supplementary practical skills courses	1260+/-	320	
Availability of quality teaching and learning materials improved	Number of course packets published			The indicator value can rise merely by publishing course packets, regardless of the quality of the material. Additionally, the indicator overlaps to some extent with the one above—the "number of law students and sharia students attending supplementary practical skills courses"—in that the course packets are to be published for the supplementary practical skills courses. Moreover, there is no explanation of how the second indicator will be tracked and the first indicator does not establish a target.
	Quality of legal and/or regulatory standards for legal education, disaggregated by university and continuing	TBD	500	
Improved capacity of adjunct instructors	Number of adjunct instructors trained	24	12	Providing training to instructors does not necessarily mean that their capacity to provide quality instruction will be improved.
Improved legal aid services to indigent and vulnerable population	Number of individuals/groups from low income or marginalized communities who received legal aid or victim's assistance with USG support	TBD	4160 clients	Additional, INL-specific indicators (the second and third on the left) are distinct from the first indicator, but appear to be grouped with it for tracking purposes. This grouping conflates the "quality of legal representation" with the number of people who receive legal aid. Moreover, targets were not established for any of the indicators.
	Availability of free legal assistance for indigent defendants			
	Quality of legal representation			
	Percentage of cases handled by fellows	TBD	8862 cases	
	Percentage of cases advised by students	TBD	300 cases	
	Recent graduates of students and fellows employed in jobs that contribute to the development of country's Legal System	TBD	TBD	
Enhanced capacity of legal advice providers in legal aid delivery and organizational development	Number of partners organizations' capacity built (HR & organization) (in the area of M&E, supervision, quality control and advocacy & public legal awareness campaigns)	4	4	The value of the indicator is not a measure of capacity. Rather, it is simply the number 4, which merely reflects the number of partner organizations receiving technical assistance through the program. Although the FY 2017 performance management plan noted that periodic assessments would be used to track organizational capacity, the fact that the baseline and target values for the indicator were identical even as the assistance was still ongoing raised questions about the usefulness of the metric.
Access to viable legal advice facilities improved	Number of clinics established	4	4	This was an important program output. However, as no performance indicator tracks the impact of the clinics (or plausible proxies), it is difficult to understand what effect the establishment of the clinics may have had on legal outcomes for Afghan women and other vulnerable populations.
Increased public legal awareness through advocacy campaigns, and create a dialog on legal education policy	Number of client referrals as a result of PLA initiatives	TBD	128	This appears to be a useful outcome metric. However, no target was established.

Continued on the following page

CRITICAL ANALYSIS OF PERFORMANCE INDICATORS FOR STATE'S LEGAL AID THROUGH LEGAL EDUCATION PROGRAM, AS PRESENTED IN THE PROGRAM'S FY 2018 PERFORMANCE MANAGEMENT PLAN (CONTINUED)

Program Objective	Performance Indicators	Target	Baseline (if available)	SIGAR Observations
Public legal awareness needs identified and improved through Public Legal Awareness (PLA) campaigns	Number of PLA campaigns on Legal aid produced & aired	TBD	Produced 12 radio & TV sport Aired 459 times	Conducting public awareness campaigns does not necessarily equate to identifying and improving public legal awareness needs and no target was established.
Multi-stage holistic advocacy initiative taken	A holistic advocacy strategy developed	1	0	It is unclear what a "multi-stage holistic advocacy initiative taken" means, as it is not defined in either the FY 2017 or FY 2018 performance management plan. It is therefore unclear whether the performance indicators adequately measure progress towards this program objective and only one of the three indicators includes a target.
	Number of conferences arranged	TBD	2	
	Number of sessions/meetings organized	TBD	36	

Source: State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," pp. 8-9; State, INL, "TAF LALE PMP Revised Final from Implementer," February 2, 2017; SIGAR analysis.

counternarcotics-related performance management plans it reviewed, none contained "accurate and targeted output, outcome, and impact methodologies or measures to track program achievements and results."⁶⁵³

Shifting the definition of "success" based on what was quantifiable was a key feature of State's counternarcotics programming, where the original goal of reducing Afghanistan's narcotics trade was for a time displaced by the goal of destroying poppy fields. According to a joint assessment by State and DOD's inspectors general, INL believed farmers would be deterred from planting poppy if the Afghan government could destroy a sufficiently high number of hectares under cultivation. Between 2004 and 2009, it operated a poppy eradication force that attempted to do just that. Contractors working on the project told SIGAR that even with perfect weather, security, and no vehicle breakdowns, they could have destroyed only approximately 70 percent of the targeted amount of poppy.⁶⁵⁴

Ultimately, the effort proved fruitless: Eradication had no lasting impact on the opium-poppy problem.⁶⁵⁵ The combination of goal displacement and tunnel vision regarding methods resulted in what Richard Holbrooke (then soon to be State's Special Representative for Afghanistan and Pakistan) called "the single most ineffective program in the history of American foreign policy"—nearly \$300 million dollars spent for no discernable result.⁶⁵⁶ The number of destroyed hectares under cultivation, and the amount of poppy replanted in the next season, are two different metrics; the former is easiest to measure, but the latter is more representative of the actual desired result. Contractors working on the project reported to SIGAR that INL's desire to quantitatively demonstrate progress prevented them from considering harder-to-measure alternatives, such as public information campaigns and alternative development programs.⁶⁵⁷



Afghan counternarcotics police destroy an opium crop in Badakhshan in 2012. (UN photo)

Smaller counternarcotics programs also substituted quantitative metrics that failed to measure actual program objectives. INL's interdiction initiative, for instance, focused on the operation and maintenance of facilities that housed the Afghan government's counternarcotics police and the U.S. Drug Enforcement Agency—but the program's performance management plan listed one planned outcome as “increase in percentage of investigations planned.”⁶⁵⁸ It was unclear how the operation and maintenance of facilities correlated with this outcome.⁶⁵⁹ In a 2014 audit, the State inspector general attributed INL's “largely unknown” impact on counternarcotics trends in part to inappropriate metrics presented in performance management plans.⁶⁶⁰

Compliance requirements and pressure to ensure sufficient oversight of funds created another set of challenges. The goal of State's Corrections System Support Program (CSSP) is to create “a safe, secure, humane, and transparent Afghan corrections system.”⁶⁶¹ The program is intended to support broader strategic goals. For example, its fiscal year 2017 performance management plan noted that CSSP was aligned with strategies that aimed to “help eliminate Taliban justice and defeat the insurgency” and “help increase the Afghan government's legitimacy.”⁶⁶² But at the programmatic level, partly in response to oversight scrutiny, INL spent significant resources ensuring that contractual requirements were met—monitoring whether blankets were delivered to Afghan prisons, or making sure that concrete mixed for Afghan prison construction did not contain too much sand.⁶⁶³ Remarking on the disconnect between what can be monitored and what was actually relevant to core objectives, one State official wondered rhetorically why, in a country with a deeply flawed justice system and prisons not built to modern standards, “we should spend money to see if someone is changing the lightbulbs?”⁶⁶⁴ Theoretically, contracts could

Data Collection Mechanisms, Data Interpretation and Use Remain Key Challenges

Relevant, accurate data is the foundation of M&E, but data collection has been a challenge for State. Some bureaus lack the resources to collect needed data, or fear that a negative evaluation based on unflattering data will result in a funding cut.⁶⁶⁸ Moreover, it is not always clear what data is relevant, nor is it always obvious what constitutes an appropriate amount. For example, to respond to the Congress' demand for tangible measures of reconstruction progress, INL contracted with Deloitte to develop an M&E system and framework for the bureau's use.⁶⁶⁹ But the system created by Deloitte was far too complicated, with "too many data points to make sense out of it," according to INL officials.⁶⁷⁰ Though Deloitte developed it based on "industry standards," INL determined that it did not actually require the volume of data produced by the system.⁶⁷¹ INL eventually awarded contracts that focused on data management, verification, and control in order to improve its data collection.⁶⁷²

Designing ways to triangulate data also proved challenging. State adopted USAID's multi-tiered monitoring framework, which is intended to help verify data by collecting it from multiple sources. Information collected through multi-tiered monitoring is organized into tiers based on how accurate it is believed to be. Within this hierarchy, U.S. government reporting is the most trusted form of monitoring. However, officials with State's Bureau of South and Central Asian Affairs noted that even U.S. government employee site visits can be

of limited value because "recipients and implementing partners are always going to put their best faces on for a site visit, which may not represent reality."⁶⁷³ Further, government personnel are not always well trained in monitoring techniques, which leads to dependence on contractors to provide program information. Such dependence can lead to reports of overly favorable results.⁶⁷⁴ Still, SIGAR's work shows that sometimes there is no substitute for physically "kicking the tires."⁶⁷⁵ In Afghanistan, this can be particularly challenging, given the mobility restrictions due to an unstable security environment.⁶⁷⁶

The next most trusted form of monitoring within the multi-tiered monitoring framework is third-party monitoring, which also has strengths and weaknesses. Strengths include the opportunity to monitor projects implemented in areas inaccessible to U.S. government personnel.⁶⁷⁷ However, a drawback is that the quality of reporting is not always high.⁶⁷⁸ Nevertheless, if carefully managed, third-party monitors can provide useful verification of basic project outputs, such as whether a building has been constructed.⁶⁷⁹

In essence, there is no silver bullet when it comes to data collection and interpretation. Every mechanism carries certain limitations. In the end, careful data triangulation is a key aspect of ensuring accuracy and precision in performance measurement.⁶⁸⁰

"perform" perfectly—that is, meet their contractually required deliverables—but in the words of one State official, "whether the taxpayer received value is another question."⁶⁶⁵

Selected Metrics Sometimes Did Not Fully Capture Broader Political Goals

Reconstruction programming was often used as a means of achieving political ends. Continuing work on the problem-plagued Kajaki Dam project, for instance, was intended in part as a signal that the U.S. was committed to Afghanistan.⁶⁶⁶ In essence, reconstruction is ultimately in many ways a political enterprise, and conducting activities to obtain political influence is a central aspect of State's mission. Yet, throughout reconstruction efforts in Afghanistan, State's M&E systems struggled to capture the centrality of political objectives.⁶⁶⁷



An Afghan engineer works in the Kajaki Dam facility. (USAID photo)

It is difficult, for example, to quantify progress towards the Corrections System Support Program's goal of a "safe, secure, humane, and transparent Afghan corrections system." Tracking the number of instances of training, the number of prisoners participating in prison industries, and the number of hours a program spent advising on gender issues allowed INL, in its own words, "to measure contractor performance and respond to frequent [requests for information] from . . . SIGAR, Congress, and other outside entities."⁶⁸¹ It is unclear, however, whether increases in any of these indicators actually correlate with a "safe, secure and humane" prison system.⁶⁸²

Further complicating measurement, it is not clear whether the existence of such a prison system is necessarily correlated with the strategic objectives the program was intended to support. For example, the Corrections System Support Program's fiscal year 2017 performance management plan stated that the program was aligned with the U.S. government's rule of law strategy for Afghanistan, one objective of which was to promote "a culture that values the rule of law above powerful interests."⁶⁸³ But a country can display strong rule of law characteristics even with a prison system that invites international criticism. Japan, for instance, maintains an almost perfect rule of law score from Freedom House, despite calls from Human Rights Watch to eliminate "hostage justice" that allows suspects to be detained for more than three weeks without indictment and that prevents them from having lawyers present during questioning from law enforcement.⁶⁸⁴ The point is that the relationship between programmatic goals and strategic objectives can be extremely complicated.

Similar challenges of relevance apply to measurement of State's Justice Sector Support Program. A central aspect of the program is its case management system, an online

database that tracks the status of criminal and civil cases in Afghanistan across all criminal justice institutions, from arrest through release from prison.⁶⁸⁵ According to INL officials, key indicators for the program include the number of Afghan government staff who have access to the case management system, who have been trained to use it, and who are uploading data into the system, as well as the number of system terminals that have power and internet.⁶⁸⁶ INL modified the program's contract to "enhance focus" on the database, even creating a new version of the system to hand off to the Afghan government.⁶⁸⁷ The Afghan government's progress in implementing the case management system is a highly visible metric that SIGAR has tracked closely.⁶⁸⁸

Theoretically, full adoption of the system may result in significant improvements in Afghanistan's justice sector. In May 2021, INL reported that the system held more than 800,000 criminal and civil cases and that Afghan law enforcement had conducted 80,000 criminal background checks using it.⁶⁸⁹ SIGAR welcomes the Afghan government's increasing use of the case management system. But, on their own technical solutions may not result in better rule of law. It is easy to come up with examples of countries where technology is both functional and widely used but the rule of law is arbitrarily applied. China is an authoritarian dictatorship without due process or an independent judiciary, and engages in arbitrary detention of political dissidents and ethnic minorities—it also employs a sophisticated technological apparatus to maintain control.⁶⁹⁰ As SIGAR underscored in a November 2019 audit of the Afghan government's anticorruption strategy that discussed the Afghan government's implementation of the case management system, among other topics, without the political will to fight corruption, transformative change in Afghanistan's rule of law will remain elusive.⁶⁹¹ In May 2021, INL emphasized the extent to which the Afghan government now relies on the case management system to help operate its justice system.⁶⁹² Corruption in Afghanistan remains pervasive.⁶⁹³

Another sector in which metrics did not fully capture political objectives involved the U.S. government's largest effort to advance Afghanistan's higher education system: its 14-year investment in the American University of Afghanistan (AUAF) in Kabul.⁶⁹⁴ State, DOD, and USAID have collectively invested at least \$167.3 million in the university through a combination of grants, cooperative agreements, and contracts.⁶⁹⁵ Considered to be Afghanistan's flagship university, the American University of Afghanistan was believed to be an important symbol of U.S. commitment.⁶⁹⁶

However, the metrics used to assess the performance of AUAF did not directly capture whether U.S. support to the university advanced this symbolic, ultimately political commitment. USAID tracked performance against indicators like "[the percentage] of undergraduate and graduate students who complete their degree programs each semester."⁶⁹⁷ Meanwhile, State tracked metrics such as the number of scholarships it provided to AUAF students. The number of scholarships provided to women helped increase the proportion of the student body that was female, which was another metric.⁶⁹⁸ Although such indicators may have been useful in assessing AUAF's educational performance and student demographics, they could not be used to answer



Ambassador Karl Eikenberry speaks at the American University of Afghanistan's first commencement ceremony in the spring of 2011. (USAID photo)

the more fundamental question of whether perceptions surrounding AUAF's political symbolism were justified.

Over time, serious deficiencies in AUAF's management and operations attracted significant attention from SIGAR and USAID's inspector general.⁶⁹⁹ As a result, oversight itself, and whether the university tightened its financial controls, essentially became the new metrics by which the performance of U.S. funding was assessed, particularly after SIGAR and USAID's inspector general found that the university could not account for \$63 million received from the U.S. government.⁷⁰⁰

But in the minds of U.S. officials, the motivation to fund the university remained far broader. As one USAID official put it, "Because the [university] was a high-visibility institution, its failure would equate to American failure in Afghanistan."⁷⁰¹ Although officials believed such failure would be detrimental to U.S. national security interests, the way the university's existence explicitly advanced those interests was not captured by the indicators used to assess the effectiveness of U.S. funding.⁷⁰² It may not have been unreasonable to assume that the collapse of a highly visible project would harm U.S. security and political goals, but that assumption was never formally assessed or tested.

In interviews conducted for this report, officials from State's Bureau of South and Central Asia Affairs openly acknowledged the tension between programmatic goals and political objectives. As several officials put it, while the size of a program's budget can signal its political importance, there is still a "quantum leap" between programs and contracts

Weighing the Relative Importance of U.S. Programs and Examining Macro Trends Can Help Determine Whether Political Aims Are Being Achieved

Assumptions around political symbolism can hold donors hostage. SIGAR's 2020 audit of State, USAID, and DOD assistance to the American University of Afghanistan found that the university had experienced significant financial, management, and staffing problems for over a decade. USAID officials told SIGAR that they believed they had less leverage to address these problems because AUAF's management was aware of the university's perceived symbolic importance.⁷⁰⁶

An exercise as simple as rank ordering U.S. assistance programs according to the degree to which they are believed to be effective, or otherwise instrumental to the reconstruction effort, could have provided perspective on the relative importance of different efforts.

Such a ranking could have helped stress-test assumptions about the political effects of programming, such as those attached to U.S. funding of the American University of Afghanistan. For example, if ranking revealed that security assistance or economic growth programs were deemed more fundamental to the survivability of the Afghan state than higher education

programs, the assumption that the university's failure would be synonymous with U.S. failure in Afghanistan could be considered less valid. Ranking could also reveal the inverse: Smaller programs attached to efforts with great political symbolism may actually be some of the most important and therefore worth preserving.

In March 2013, SIGAR asked the Secretaries of State and Defense, as well as the USAID Administrator, to provide a list of the 10 Afghanistan reconstruction projects or programs their departments deemed most successful and least successful.⁷⁰⁷ None of them did so. SIGAR's request remains unfulfilled.⁷⁰⁸

A second way of assessing whether political, or otherwise broad or abstract, objectives are being achieved is to examine the high-level, *prima facie* evidence. This is essentially how SIGAR concluded that the entirety of U.S. spending on certain programmatic areas was likely wasted. For example, pointing to the fact that even a substantial focus on counternarcotics appeared to have done very little to stem the production and export of illicit drugs, SIGAR concluded that U.S. counternarcotics efforts appeared to have failed.⁷⁰⁹

on one hand and strategic effects on the other. Consequently, they added, program performance management plans should include explicit measures of political objectives, where applicable, as "development impact is a consideration but it's not the only one."⁷⁰³

In the end, metrics that were purely quantitative or technical in nature often could not fully capture the transformational social and cultural changes sought by political programming.⁷⁰⁴ The limitations and frustrations of distilling such change into discrete indicators are reflected in the musings of one State official interviewed for this report, who expressed their opinion that no matter what improvements State and other U.S. agencies made to their contracting or M&E processes, most development programs in Afghanistan would still fall well short of their goals. The problem, the official said, using language that cannot be printed here, was "out there" beyond the walls of the U.S. embassy in Kabul, where U.S. development programming was pursuing transformative change in a country where such change was not possible.⁷⁰⁵

LACK OF PERFORMANCE METRICS HINDERED M&E OF RULE OF LAW ASSISTANCE



Parliamentarians gather to review prisoner sentences and prison conditions during an oversight trip to Herat Province. (USAID photo)

The impetus for State's rule of law programming was the widespread American and Afghan perception that the formal justice system was corrupt and difficult to access, while the informal system frequently provided outcomes that did not meet international human rights standards.⁷¹⁰ The U.S. and Afghan governments believed that strengthening rule of law would build the legitimacy of the Afghan government and foster sustainable development.⁷¹¹ A 2015 SIGAR audit showed that State, DOD, and USAID could not fully determine the effectiveness of U.S. programs that aimed to advance these objectives.⁷¹² The programs cost more than \$1 billion.⁷¹³

State issued its first interagency rule of law strategy for Afghanistan in 2009. The strategy focused on reducing the "culture of impunity" in the justice sector by reducing corruption.⁷¹⁴ It defined a scope for rule of law activities, emphasizing expanding access to the formal justice sector, reforming detention policies and the Afghan corrections system, providing "security and space" for the informal justice sector, and growing leadership capacity in the Afghan government and civil society.⁷¹⁵

The strategy listed measures of effectiveness, such as levels of public confidence in the Afghan justice system as shown through polls and the number of cases heard by the formal justice system, against which progress could be monitored.⁷¹⁶ However, despite the existence of the performance measures, no measuring and reporting was done against them. Even if data had been collected to measure performance, the strategy's lack of baselines and targets would have precluded complete assessments of progress.⁷¹⁷ U.S. embassy officials described the 2009 strategy as "overambitious" and said it did not reflect the operating environment or available resources.⁷¹⁸

Lack of measurement in rule of law assistance was also a problem at the programmatic level. In 2015, SIGAR noted that State's inconsistent M&E of the Justice Sector Support Program led to problems assessing the program's actual impact.⁷¹⁹ Indicators focused on outputs instead of outcomes and lacked baseline data. Based on a cumulative assessment of M&E data generated by the program, SIGAR concluded in 2015 that INL could not determine the extent to which it had contributed to the development of rule of law and the sustainability of the Afghan justice sector.⁷²⁰

LEARNING AND ACTING ON INFORMATION

M&E Information Was Not Always Systematically Used—or Useful

The purpose of State’s Managing for Results framework was establishing “a clear line of sight from what the department wants to achieve, as documented in its strategic plans, to how the department intends to achieve it through key programs and projects.”⁷²¹

However, performance against strategic objectives was not always tracked. Although State’s Bureau of South and Central Asia Affairs was required by policy to assess progress against the goals of State’s 2015–2018 Joint Regional Strategy, a 2018 inspection of the bureau by State’s inspector general found that it lacked a system to do so. In interviews with the agency’s inspector general, bureau employees said they had not monitored progress toward regional strategy goals since the strategy’s promulgation in 2014.⁷²² Similarly, as SIGAR reported in a 2015 audit, no measuring or reporting was conducted against 27 performance indicators that could have helped determine progress towards meeting U.S. rule of law objectives in Afghanistan.⁷²³

Infrequent use of key M&E processes may have also hampered strategic-level learning. A critical aspect of strategic-level learning is the use of evaluations to examine the performance and outcomes of programs and projects.⁷²⁴ State’s current M&E policy requires bureaus to “consider evaluation findings to make decisions about policies, strategies, priorities, and delivery of services.”⁷²⁵ The overall number of evaluations across all State programs increased from “negligible” in 2012 to 180 in 2017, and the quality of evaluations seems to have improved over time.⁷²⁶ But as a 2018 meta-evaluation of State’s foreign assistance evaluation efforts stated, bureaus generally used evaluations in a manner that was “more sporadic than systemic.”⁷²⁷

This finding broadly conformed to interviews conducted for this report. In response to SIGAR inquiries on whether evaluations resulted in changes to programming, State said it does not track whether midterm evaluations lead to contract or grant modifications.⁷²⁸ Similarly, INL contractors interviewed by SIGAR could not think of any contract modifications that were made based on information gleaned either from evaluation reports or from other M&E activities.⁷²⁹

One reason for this may be that evaluations are primarily performed only on expensive flagship programs, due to the perception that the costs of evaluations sometimes outweigh the benefits.⁷³⁰ For example, U.S. Embassy Kabul’s Public Affairs Section, whose portfolio generally includes smaller programs, said in May 2019 that it had not conducted an impact evaluation in the last five years—adding, “There are no meaningful examples of course correction resulting from evaluation.”⁷³¹

But even where evaluations of State’s reconstruction programming were conducted, the quality was not always strong. According to State’s M&E policy, evaluations are performed for the purpose of “improving effectiveness and informing decisions about current and future programs, projects and processes.”⁷³² In order to do so, evaluations

must adhere to standards of usefulness and methodological rigor.⁷³³ However, there were cases where evaluations did not seem to fully live up to these standards.

For instance, after nine and a half years and more than \$250 million spent, a State-commissioned evaluation of the Corrections System Support Program was completed in July 2015.⁷³⁴ The evaluation was completed at the same time that a SIGAR audit criticized INL's decision to extend the program despite INL's inability to determine whether it was achieving its overall objectives.⁷³⁵ That key question was unanswered by the evaluation. The program's goal was to build "a safe, secure, and humane prison system."⁷³⁶ But in response to the evaluation question, "Have CSSP interventions improved conditions for and treatment of inmates as well as the fairness and responsiveness of the overall justice system?" the evaluators deflected. After stating that "a total of 1,557 hours of advising were conducted between 2012 and 2014 on the topics of human rights and humane treatment," they noted that "only 61 percent . . . of inmates agree that inmates are treated humanely."⁷³⁷ This could have understated the finding, given that prison guards were present during the interviews, meaning that "obtaining truthful responses was of particular concern . . . due to fear of repercussions if anything negative was communicated."⁷³⁸ In the end, the evaluators did not clearly answer the question of whether the program's interventions, including human rights training, had improved conditions for inmates.⁷³⁹ Yet, they recommended more training.⁷⁴⁰

A second critical shortcoming of the evaluation was "the absence of baseline data or a control group"—the latter referring to facilities that did not receive support from the program.⁷⁴¹ Consequently, the effects of the program on outcomes of interest could not be fully ascertained.⁷⁴² There were also confounding results. Evaluators hypothesized "that if CSSP's activities had an impact on the performance of each facility, then we should see a positive correlation between the amount of resources a facility received (in terms of dollars, training courses, mentoring hours, or infrastructure projects) and the current performance level of each facility."⁷⁴³ However, according to the evaluators, "regression analyses testing the relationship between the amount of training in a given facility and perceptions of safety, security, humaneness, management, and sustainability were mixed."⁷⁴⁴ Increased training in correctional facilities was on one hand associated with "higher perceptions of sustainability, but lower perceptions of safety and humane treatment."⁷⁴⁵ Additionally, the evaluators stated, "similar to tests of the relationship between training hours and impacts, the relationships between advising hours and perceptions of humane treatment and security are negative."⁷⁴⁶ Evaluators noted additional unexpected regression results: "Similar to the training and advising data, when regression analyses were conducted, there were significant and negative relationships between infrastructure dollars per facility and facility manager perceptions of humane treatment, management, and sustainability."⁷⁴⁷ Less than 50 percent of correctional officers, trainers, and facility managers believed their facilities could operate without support from the program.⁷⁴⁸

Evaluators provided several possible explanations for the "unanticipated" results. For example, with respect to the relationship between training and perceived impact, they



A 2014 SIGAR inspection showed that conditions in these maximum security cells at Pul-e Charkhi Prison, a facility supported by CSSP, were overcrowded and poor. (SIGAR photo)

suggested that the training could have been successful but that managers were unable to implement necessary changes at their facilities because of their lack of ability to do so.⁷⁴⁹ Another plausible explanation was that the program had provided more training and advising in those corrections facilities that needed them the most.⁷⁵⁰ But the fact that many aspects of the analysis presented in the evaluation rejected the evaluators' initial hypothesis raised important questions about the program's effectiveness.

Despite this, in many cases the evaluation's recommendations seemed only to suggest reconfiguring program inputs, such as "[including] questions about the feasibility of implementing practices learned in training courses," continuing to work with corrections officials "to develop internal capacity to repair and maintain facility infrastructure," and continuing "efforts to centralize the classification and case management system."⁷⁵¹ The highlight following this section details analytical flaws in two additional, Afghanistan-specific evaluations. Table 6 at the end of the following highlight summarizes our analysis of key flaws of all State evaluations examined for this report.⁷⁵²

Shortcomings in evaluations of State's reconstruction programming reflected broader trends. The aforementioned 2018 meta-evaluation found that, across the department, "the number of fair or poor reports exceeded the number of good reports by a substantial margin."⁷⁵³ The meta-evaluation noted that common problems included inactionable recommendations, failure to connect findings to bureau strategy, and poor evaluation design. As a result, as of 2018, support for evaluations within State varied widely, and evaluation findings and recommendations were only used "to some extent."⁷⁵⁴

EVALUATIONS OF STATE'S RECONSTRUCTION PROGRAMMING SUFFERED FROM ANALYTICAL FLAWS



A farmer assisted by the Community-Based Agriculture and Rural Development project prepares his tomato harvest for sale at a local market. (UN photo)

Lost in Equivocation: An Evaluation of a Justice-sector Training Program Praised a Project's Design without Convincingly Demonstrating the Project Was Likely to Have Impact

The evaluation of State's Corrections System Support Program was not the only one that suffered from analytical shortcomings. A 2014 midterm evaluation of the Justice Training Transition Program exhibited similar flaws. The \$47.7 million, 30-month training program aimed to improve the legal knowledge and skills of Afghan justice-sector officials and, separately, to eventually transition the training provided under the program to the Afghan government.⁷⁵⁵ The program's overall goal was to "increase the confidence of Afghan citizens in the justice sector."⁷⁵⁶

Some of the evaluation's conclusions and recommendations were puzzling. On one hand, the evaluators concluded that the training program was designed in a way that adequately addressed the needs of the Afghan justice sector, with "needs" sweepingly defined as the "formation of [a] competent rule of law system needed for achieving peace and security, poverty reduction, economic growth, and the protection of fundamental rights in Afghanistan."⁷⁵⁷ On the other, the evaluators noted that the program's hypotheses around desired effects "just [tackled] the short and medium term results that [showed] direct program outputs; meanwhile, they [did] not describe the full path of expected changes on the individual and institutional levels."⁷⁵⁸

This raised questions about whether the program was in fact adequately designed. Specifically, the program's logic model (which described the relationship between program activities and outputs and presumed outcomes and impact) had omitted key conditions like "political commitment to improve [the] justice sector system" and "political will to fight corruption"

that were necessary for achievement of desired results.⁷⁵⁹ “Further consideration of these *risks* and *assumptions* could have strengthened the initial program design,” the evaluators wrote while adding, “The *activities* as initially planned for are generally relevant for the fulfillment of the stated objectives and results.”⁷⁶⁰ Translating these confusing equivocations: the training was relevant for addressing certain needs of the Afghan justice sector, but could ultimately end up being ineffective if other “essential” conditions, like the presence of political will to enact important reforms and fight corruption, were not met.⁷⁶¹

The evaluators seemed deeply uncertain about whether such political will would materialize. Citing data indicating that “confidence among Afghans in the justice sector has increased since the fall of the Taliban, but may have plateaued in recent years,” the evaluators noted, “this is clearly worrying.”⁷⁶² The evaluators then seemed to attempt to write away from this observation by showing that “attitudes are in fact more nuanced.”⁷⁶³ However, after running through a litany of indicators from an annual national survey—some of which were tangential to the justice sector, like levels of fear when encountering international military forces and voting in national elections—the evaluators circled back to negative recent trends: “The perception that corruption is a major problem in Afghanistan as a whole was at its highest point in 2013 since 2006.”⁷⁶⁴

In the end, despite pointing out significant omissions in the training program’s logic model and never establishing that the program was likely to impact the end goal of restoring public confidence in the Afghan justice system, the evaluators recommended that the program be continued and extended by six months.⁷⁶⁵ The certainty that underpinned this recommendation was difficult to understand in light of the evaluator observations described above—but it was also a quintessential example of how M&E practices created the risk of doing the wrong thing perfectly.

An Evaluation of an Alternative Livelihood Project Mostly Sidestepped the High-risk Nature of Counternarcotics Programming in Afghanistan

At other times, analytical problems were less about equivocation and more about missing the forest for the trees. An evaluation of a \$24 million counternarcotics program funded by State presented interesting granular analysis but struck an optimistic tone despite the extraordinarily high-risk nature of counternarcotics programming in Afghanistan.⁷⁶⁶

The purpose of the Community-Based Agriculture and Rural Development–West project was to reduce the cultivation of opium poppy by encouraging farmers in Farah and Badghis Provinces to grow alternative high-value crops, such as apples, citrus fruits and pistachios.⁷⁶⁷ A key assumption underlying the program’s theory of change was that farmers preferred to plant the most profitable crop. The evaluation therefore presented 16 pages of market analysis, price calculations, and graphics that supposedly demonstrated the potential long-term viability of selling these alternative crops.⁷⁶⁸ The evaluation concluded, “it is likely that high-value crops could make an important contribution to an overall counternarcotics and development strategy,” although it noted that the full impact of the program would remain unknown until 2028.⁷⁶⁹

A June 2018 SIGAR lessons learned report on U.S. counternarcotics came to the opposite conclusion: Past experience combined with then-current security challenges gave little reason to assume that the Community-Based Agriculture and Rural Development program, whatever its development impact, would have lasting effects on poppy cultivation.⁷⁷⁰ Specifically, the SIGAR report found that many Afghan farmers were unable to make a long-term investment in crops

substituted for opium poppy because they lacked the income or credit access necessary to cover household needs while substitution crops matured.⁷⁷¹

Perhaps the evaluation of the Community-Based Agriculture and Rural Development–West project occurred too contemporaneously with SIGAR’s lessons-learned report to benefit from SIGAR’s findings. But skepticism surrounding efforts to reduce opium poppy cultivation has been a recurring theme of reconstruction and extends beyond SIGAR. In 2004, the Government Accountability Office reported that “efforts to counter [deteriorating security and increasing opium cultivation] have had little success.”⁷⁷² In 2006, the GAO stated, “The worsening security situation and the lack of Afghan capacity are tremendous challenges to the success of U.S. counternarcotics programs in Afghanistan.”⁷⁷³ In 2010, GAO reported that progress in counternarcotics was “difficult to fully assess” and “in addition, challenges, such as lack of security, political will, and Afghan government capacity affect progress in all program areas.”⁷⁷⁴

These deeply entrenched challenges persist. They have been the primary contributors to the exponential rise in opium poppy cultivation and drug production since the onset of reconstruction.⁷⁷⁵ They are largely beyond the control of counternarcotics programs and imply serious limitations to the U.S. capacity to bring about large-scale, lasting reductions in poppy cultivation and drug production.⁷⁷⁶ In 2018, opium poppy cultivation was at its second-highest level on record, despite the combination of a severe drought and lower prices resulting from record-high cultivation in 2017.⁷⁷⁷

It soon became clear that the Community-Based Agriculture and Rural Development–West project would not be as successful as hoped. The evaluation stated, “This is a generally well-designed, well-run livelihoods project operating in very difficult conditions.”⁷⁷⁸ However, as SIGAR reported in January 2020, 40 percent of saplings in Farah and 62 percent of saplings in Badghis died.⁷⁷⁹ The project implementer, the United Nations Development Programme, attributed low sapling survival rates to the severity of recent floods, drought, and conflict in areas selected for intervention, and described the effects of such challenges as “unforeseen.”⁷⁸⁰ But in a country prone to bouts of drought and flooding, and which also suffers from persistent insecurity, it was difficult to understand how all such challenges were completely unforeseeable.⁷⁸¹ It also appeared that the UN Development Programme was unprepared to fully address potential project risks by providing timely support to beneficiaries, due to the remoteness of project locations, limited accessibility in times of flood, general insecurity, and weak management and coordination at local, district, and provincial levels.⁷⁸² In other words, it seemed the project had been bedeviled by at least two of the same challenges that stymied many other U.S. efforts in the sector: insecurity and weak Afghan government capacity.

Based on the low survivability rates of the saplings and the limited ability of the UN Development Programme to expeditiously respond, it appeared that the evaluator may have overstated the degree to which the program was in fact “generally well-designed” and “well-run.”⁷⁸³ Ultimately, the significant loss of crops, viewed in the broader context of the high-risk nature of U.S. counternarcotics programs in Afghanistan, rendered irrelevant the complex market analysis and price calculations performed by the evaluator.⁷⁸⁴

Table 6 on page 26 summarizes key flaws in all State evaluations examined for this report, including those described in this highlight.

Periodic Program Reviews Did Not Always Adequately Assess Progress towards Objectives

Evaluations were not the only mechanism used by State to determine how programs were progressing. As State's M&E guidance evolved, it began to explicitly encourage "intentionally incorporating regular reviews of progress" that were intended, among other purposes, to help program teams "test theory of change hypotheses" and "identify and monitor assumptions and context."⁷⁸⁵

Theoretically, program performance management plans provided the basis for such reviews. However, State programs initiated in the earlier years of reconstruction generally lacked such plans. For example, the Corrections System Support Program began in 2006, but State did not issue a strategy to develop the Afghan justice system until 2009.⁷⁸⁶ INL first produced a performance management plan for the Corrections System Support Program three years after the production of State's Afghan justice system strategy and six years after the program was initiated.⁷⁸⁷ The plan represented INL's first attempt to begin formally evaluating the effectiveness of the program.⁷⁸⁸ According to State's inspector general, before then, INL did not use program data to measure program effectiveness.⁷⁸⁹ The performance management plan sought to link outcomes, outputs, and activities with program goals.⁷⁹⁰

Shortcomings in CSSP's Performance Management Plan

State policy dictates that every program be informed by a logic model that "articulates how and why the program or project is expected to contribute to achieving the program/project goals and objectives."⁷⁹⁵ Logic models are a key component of performance management plans, the foundation of program-level M&E. While logic models are defined as "a rigorous methodology," they do not always present as rigorous.⁷⁹⁶ For example, while the 2018 performance management plan for State's Corrections System Support Program had a logic model, there was no clear explanation of the supposed causal linkages between program outcomes, objectives, and the ultimate goal of developing a "safe, secure, humane, and transparent corrections system."⁷⁹⁷ It is not clear whether program outcomes, even if achieved, would meaningfully contribute to the overall goal, given that achievement of the goal may depend in large part on the political will of the Afghan government to tackle pervasive issues like corruption.⁷⁹⁸

The performance management plan provided a framework and benchmarks against which effectiveness could be assessed during periodic reviews.⁷⁹¹ However, while the first performance management plan was produced in October 2012, INL did not begin conducting periodic reviews of the program before 2014.⁷⁹² From then on, INL held reviews every six months that theoretically allowed INL to "course-correct in a timely manner."⁷⁹³ Course correction was premised on the idea that such reviews were a key way of continually assessing program effectiveness.⁷⁹⁴ But in practice, these periodic reviews sometimes seemed perfunctory.

TABLE 6

ILLUSTRATIVE CONTRADICTIONS OR LOGICAL FLAWS IN STATE EVALUATIONS

Project/Program	Evaluation Year	Project/Program Purpose	Statement One
Justice Training Transition Program (JTTP) Midterm Evaluation	2014	Increase Afghan citizens' trust in the formal state criminal justice sector by increasing capacity through training and mentorship of Afghan legal professionals (such as judges and prosecutors)	The JTTP's "formulation is adequate to addressing the existing needs of the Afghan justice sector, (i.e. formation of the competent rule of law system needed for achieving peace and security, poverty reduction, economic growth, and the protection of fundamental rights in Afghanistan)" (p. 23). "Assumptions on the outcome level ... do not reflect all necessary conditions ... for achievement of expected results, in particular political commitment to improve [the] justice sector system and create favorable conditions for fair administration of justice, political will to fight corruption and allocate sufficient resources ... for ongoing professional development at all were not considered" (p. 24). The evaluation states that, according to Asia Foundation survey data, "There was ... an increased willingness to report on crimes and violence, from 52 percent in 2012 to 64 percent in 2013." The evaluators use this to suggest, as part of a broader answer to the question "Is public confidence in the justice system in Afghanistan improving?" that even though "trust in the role of the [Afghan] State ... has not increased much since" 2007, "attitudes are in fact more nuanced" (pp. 46-47).
Corrections System Support Program (CSSP) Final Evaluation	2015	Develop a safe, secure, humane and transparent corrections system that meets international standards and Afghan cultural requirements by training and mentoring Afghan corrections officials and implementing small-scale infrastructure projects	"Regression analyses testing the relationship between the amount of training in a given facility and perceptions of safety, security, humaneness, management, and sustainability were mixed; training is associated with higher perceptions of sustainability, but lower perceptions of safety and humane treatment." The evaluators noted, "negative relationships between [the] amount of training and perceived impacts is unanticipated" (pp. 2, 36). In response to the evaluation question, "Have CSSP interventions improved conditions for and treatment of inmates as well as the fairness and responsiveness of the overall justice system?" the evaluation stated, "A total of 1,557 hours of advising were conducted between 2012 and 2014 on the topics of human rights and humane treatment, and 618 staff were trained ... on humane treatment. Nearly all ... correctional officers and facility managers perceive that inmates are treated humanely" (p. 3).
Justice Training Transition Program (JTTP) Final Evaluation	2016	Increase Afghan citizens' trust in the formal state criminal justice sector by increasing capacity through training and mentorship of Afghan legal professionals (such as judges and prosecutors)	"Although there is no comprehensive system in place for monitoring the performance of the Afghan criminal justice system, there is at least <i>prime facie</i> [sic] evidence that it is improving and that programs such as the JTTP are having a positive impact. Annual surveys by the Asia Foundation show a rising willingness of people to report on crimes and violence, and seek redress through the state system" (p. 40).

Statement Two	SIGAR Analysis
"The program logic model just tackles the short and medium term results that show direct program outputs; meanwhile, they do not describe the full path of expected changes on the individual and institutional levels" (p. 24).	If JTTP's logic model accounted for only short and medium-term results, it may not have actually addressed the sweeping list of justice-sector needs described in statement one. Indeed, it is difficult to see how any individual program could possibly address every need listed.
"The activities as initially planned for are generally relevant for the fulfillment of the stated objectives and results" (p. 24).	The evaluators claimed that JTTP activities were relevant to stated objectives and results even though programmatic assumptions at the outcome level did not reflect all necessary conditions for achieving desired effects. Activities may not be meaningfully relevant if programmatic assumptions do not address key conditions necessary for success.
"The perception that corruption is a major problem in Afghanistan as a whole was at its highest point in 2013 since 2006. Although people's willingness to report incidents of crime or violence had increased significantly over the previous year, only half of those surveyed (51 per cent) had some level of confidence that perpetrators of violence or crime would be punished; which was a clear decline on previous years. Confidence in public institutions, organizations, and officials also reached an eight-year low, with a particularly sharp fall for Afghanistan's parliament. Overall, Afghans considered the local informal justice system to be more fair, trustworthy, effective, efficient, and in line with local norms and values than the formal state courts" (p. 47).	Claims of nuance are seemingly undermined by the broader note of pessimism in statement two. Attitudes can be nuanced even as an overall impression or takeaway is simultaneously very clear. Ultimately though, the evaluators did not provide an explicit answer to the question "Is public confidence in the justice system in Afghanistan improving?," despite a six-paragraph, nearly two-page single-spaced response.
"INL/CSSP should develop and implement additional training on human rights" (p. 40).	The evaluation found an inverse relationship between the amount of training provided and perceptions of safety and humane treatment. To their credit, they provided plausible explanations for the surprising result, including the possibility that "the training was successful in communicating to managers the standards that are required for a facility to be considered safe and humane by international standards, but managers were not able to implement the necessary changes at their facility (perhaps due to ability levels, funding, space, etc.)." In the end, however, the evaluators recommended more training even though they had not credibly established that the training was effective.
"However, in response to close-ended questions, only 61% ... of inmates agree that inmates are treated humanely; 14% ... report having been physically abused by guards, 13% ... report that guards use restraints as a punishment, and 53% ... report that guards accept money in exchange for favors." The evaluation then provided additional data on alleged mistreatment of inmates. (pp. 3–4)	Evaluators did not provide a clear answer to the basic question, "Have CSSP interventions improved conditions for and treatment of inmates as well as the fairness and responsiveness of the overall justice system?" Instead, they presented conflicting evidence of effectiveness without providing an informed judgement of whether the evidence, considered as a whole, pointed towards effectiveness or ineffectiveness. Additionally, the first part of the response to the question provided data on training that, in relation to the substance of the question, seemed to be mostly a non sequitur. Finally, the response did not even attempt to answer the second part of the question regarding whether CSSP had improved the responsiveness of the overall justice system. Based on the evaluation as a whole, a more accurate response to the question might have been "We don't know."
Statement one does not appear to accurately characterize Asia Foundation survey data. See SIGAR analysis for more.	Based on the source to which statement one cites, <i>prima facie</i> claims of success made by the evaluators are not completely accurate. The evaluators cite to the 2015 Asia Foundation survey. That survey states, "Nearly two-thirds (62.0%) of Afghans who say they experienced violence or crime say they reported the incident to an authority or justice dispute mechanism outside of their family." JTTP was initiated in January 2013. The 2013 Asia Foundation survey states that respondents' willingness to report incidents of crime or violence stood at 64%, indicating that the 2015 data point actually represents a slight decrease in reporting. Additionally, the evidence supporting the assertion of <i>prima facie</i> improvement appears to have been cherry-picked amid more relevant, but less optimistic, data. For example, Asia Foundation surveys included a question with a Likert scale assessment of the statement "The state courts are fair and can be trusted," which could be interpreted as a direct reflection of Afghan citizens' trust in the formal justice system. From 2011 to 2015, "agree" responses ranged from 59% to 69% with little discernable pattern, hitting the highest point in 2014, and then dropping back to 60% the following year. The 60% data point presented in the 2015 survey fell within the margin of error of the 59% reported in 2011.

Continued on the following page

ILLUSTRATIVE CONTRADICTIONS OR LOGICAL FLAWS IN STATE EVALUATIONS (CONTINUED)

Project/Program	Evaluation Year	Project/Program Purpose	Statement One
Community-Based Agriculture and Rural Development–West (CBARD–West) Midterm Evaluation	2019	Reduce opium cultivation by providing farmers with the ability to grow alternative, high-value crops	“It is likely that high-value crops could make an important contribution to an overall counternarcotics and development strategy” (p. 9).
			“This is a generally well-designed, well-run livelihoods project operating in very difficult conditions” (p. 5).

Source: Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 2–4, 10, 36, 40; Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, pp. 16, 23–24, 44, 46–48; Conor Foley and Orsolya Szekely, “External Final Evaluation of Afghanistan Justice Training Transition Program (JTTP),” March 2016, pp. vii, 40; Steve Goss, “Mid-Term Evaluation of the Community-Based Agriculture and Rural Development–West Project,” prepared for UNDP Afghanistan and INL, February 2019, pp. 5, 9; SIGAR, *Counternarcotics: Lessons Learned from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. xii, 126;

Some annual reviews, for instance, made claims of effectiveness that were unsubstantiated. The 2014 annual program review said that the Corrections System Support Program’s efforts to improve the management capacity of the General Directorate for Prisons and Detention Centers—the body that ran Afghanistan’s prisons at the time — “contributed to stability and confidence in the Afghan government.”⁷⁹⁹ But the only evidence the review provided for this claim was that the project “now has a full slate of senior advisors resident in Kabul and working daily with [General Directorate for Prisons and Detention Centers] leadership” and the directorate had “formed an assessment team to provide headquarters staff with audit reports from facilities around the country.”⁸⁰⁰ Later in the narrative, the review again sought to substantiate the assertion that the program contributed to stability and popular confidence—this time, by simply restating the program’s theory of change:

Combined, these [program activities] should help improve the capacity of [the General Directorate for Prisons and Detention Centers] and the management of [the center’s] facilities, and ultimately bolster the confidence of the Afghan people in their corrections system. And with better correctional facilities, there will likely be fewer opportunities for radicalization within prisons, and more economic opportunities for those released from prison, thus reducing some of the drivers of the insurgency.⁸⁰¹

Similarly, the Corrections System Support Program’s 2015 annual review equates “contribution to stability and confidence” with program activities, including advising that the Afghan government transfer “national security threat prisoners” to different prisons.⁸⁰² While such a step could theoretically contribute to stability, that assertion is left unsubstantiated by further evidence, with the result that program inputs and outputs are again conflated with impact.⁸⁰³ After 2015, program reviews dropped any specific mention of how the program contributed to stability and confidence.⁸⁰⁴

Such problems reflected an ongoing conflation of activities and objectives. As State’s inspector general observed in a September 2013 audit, one of the program’s objectives

Statement Two	SIGAR Analysis
SIGAR has reported that past experience combined with security challenges give little reason to assume that the Community-Based Agriculture and Rural Development program, whatever its development impact, will have lasting effects on poppy cultivation. Fundamental challenges like lack of security, a poor economy, weak governing institutions, and the failures of the wider reconstruction effort pose serious limitations to the U.S. capacity to effect material, long-term reductions in opium-poppy cultivation.	The evaluation avoids a deeper analysis of underlying programmatic assumptions. Given the significant, fundamental obstacles to success described in statement two, such an analysis may have revealed that the probability of achieving desired programmatic outcomes and impacts was low.
In January 2020, SIGAR reported that, according to the UN Development Programme (CBARD-West's implementer), 40% of saplings in Farah and 62% of saplings in Badghis died. UNDP stated that low sapling survival rates were due to the severity of recent floods, drought, and conflict in areas selected for intervention, which were too much for local communities to cope with.	Droughts, flooding and armed conflict are common occurrences in Afghanistan. Evidently, the UN Development Programme recognized the need for, but did not develop, a formal management plan for responding to natural disasters and armed conflict. Whatever informal plan existed at the time of the sapling deaths "was not successfully implemented," according to the UNDP's semi-annual update. All of this raised questions about whether CBARD-West was in fact well designed and run.

Source (continued): SIGAR, *Quarterly Report to the United States Congress*, January 2020, p. 139; State, INL, "Quarterly Reporting Template for US/INL Funded Projects: Community-Based Agriculture and Rural Development-West (CBARD-West)," November 3, 2019, pp. 2, 17; The Asia Foundation, "Afghanistan in 2011: A Survey of the Afghan People," 2011, pp. 149, 179; The Asia Foundation, "Afghanistan in 2012: A Survey of the Afghan People," 2012, p. 153; The Asia Foundation, "Afghanistan in 2013: A Survey of the Afghan People," 2013, p. 210; The Asia Foundation, "Afghanistan in 2014: A Survey of the Afghan People," 2014, p. 200; The Asia Foundation, "Afghanistan in 2015: A Survey of the Afghan People," 2015, p. 204.

was to "provide technical training and mentoring of corrections staff and prison leadership."⁸⁰⁵ However, that objective was self-fulfilling in that it did not "define an intermediate outcome or include measurable factors."⁸⁰⁶ Rather, the inspector general noted, "the objective simply describe[d] an activity that the program should undertake as part of its mission."⁸⁰⁷ Responding to a draft of this report, INL stated that it has to balance competing requirements of determining program impact and monitoring contract performance, and is required to track objectives of this nature in order to ensure that taxpayer money is being spent in an appropriate and approved manner.⁸⁰⁸ But it remained unclear why the provision of training and mentoring had been considered an objective unto itself.

Annual reviews were also intended to assess the Corrections System Support Program's interaction with U.S. strategic goals. However, the reviews generally conflated *alignment with strategy* with *contribution to strategy*. For instance, the 2014 annual review asserted that the Corrections System Support Program contributed to strategic goals simply because it provided assistance to Afghanistan's prison directorate.⁸⁰⁹ This was functionally equivalent to claiming that the program advanced strategic goals merely because it existed. Although the 2015 annual review provided a longer narrative explanation that described how the corrections program was aligned with the objectives of recently elected Afghan President Ashraf Ghani, it also claimed to be contributing to strategic goals in part by merely "exploring providing funding for several initiatives related to reducing prisoner populations in Afghanistan through alternatives to incarceration."⁸¹⁰ It was difficult to imagine a scenario in which the act of "exploring funding" could make any meaningful contribution to strategy.⁸¹¹ Later program reviews simply stated the objectives of various strategies to which the program was presumed to contribute, without explaining why the program was likely to materially affect those objectives.⁸¹² This was despite the reviews' inclusion of an explicit prompt to "describe how this program contributes to larger [U.S. government] . . . strategic goals."⁸¹³



A guard walks a corridor lined with plastic sheeting in Pul-e Charkhi prison in 2014. (SIGAR photo)

By simply listing relevant strategic goals and assuming that the Corrections System Support Program advanced them, the annual reviews skirted serious analysis of whether the program was likely to have strategic impact. The third-party evaluation of the program described previously could have been a way to assess the program's contributions. However, the evaluation equivocated in its response to a key question regarding whether the program had improved both the corrections system and overall justice system.⁸¹⁴ Claims of impact relative to strategic targets were therefore never comprehensively assessed.⁸¹⁵

Aspects of M&E Became “Work-Makers”

More M&E meant more work. One finding of the aforementioned 2018 meta-evaluation was that M&E had become so complex that some bureaus could not comply with its requirements even though they wanted to.⁸¹⁶ INL staff interviewed by SIGAR described the litany of regular reports they received—a combination of monitoring documents and contract deliverables. Those included weekly contractor reports, each consisting of six or seven mini-reports on specific areas such as the case management system, infrastructure, and gender; monthly and quarterly reports describing activities performed by the contractor during that period; spot reports on emerging program issues; serious incident reports; and third-party monitoring reports. Some INL officials referred to these documents as “work-maker reports,” implying that certain M&E and oversight requirements were sometimes too onerous to be meaningfully absorbed.⁸¹⁷ INL's rational response was to modify the nature of the reporting so that it could be better used.⁸¹⁸ Similarly, Bureau of Population, Refugees, and Migration officials said that, although third-party monitoring reports were useful because U.S. government personnel were unable to visit program sites, they significantly increased workload.⁸¹⁹

Similar challenges have cropped up in other bureaus. According to a State Office of Inspector General audit report, a Bureau of Diplomatic Security contracting officer

Differing Bureaucratic Timelines May Prevent State from Making Necessary Changes to Programs

State M&E policy requires bureaus to align programs with higher-level strategies.⁸²⁰ To fulfill this requirement, all programs must maintain a performance management plan detailing this alignment and the theory of change that undergirds it.⁸²¹ A “logic model” within the performance management plan translates the theory of change into measurable program inputs, activities, outputs, and outcomes.⁸²² To measure progress towards a desired outcome, measurable program outputs are contrasted against a baseline from which progress can be monitored and evaluated.

While efficient in theory, in practice creating an effective performance management plan is difficult, due to differing timelines for budgeting, strategic planning, program implementation, and M&E. While strategies can change quickly, rigid regulations on how program funds can be spent prevent significant deviation from a program’s original purpose.⁸²³

The rigidity of rules established by the Congress mean that it takes about a year to revise or terminate a contract. This results in the unsatisfying choice of simply not renewing the contract, or having a contracting officer’s representative informally work with an organization to correct any issues. As a result, bureaus usually move on to new awards instead of trying to perfect ongoing ones.⁸²⁴ Moreover, because the long period of time needed to complete an evaluation is not always aligned with decision-making time frames, bureaus do not consistently incorporate evaluation findings into budgets and strategic planning.⁸²⁵

The rigidity of contracts poses obstacles to acting on midterm evaluations and periodic program reviews: If it takes too much time and money to make significant, midcourse changes to programs, the processes designed to guide those changes may be less effective.⁸²⁶

replaced written reports with a weekly conference call—something he was authorized to do, as long as the phone conferences were documented and the contract was adjusted to reflect a reduction in work. However, in this case the contracting officer deviated from oversight requirements by making no effort to document the phone conferences, nor did he adjust the total value of the contract to reflect any reduction in work. As a result, no records existed to ensure the transfer of institutional knowledge, and State ended up paying for work it did not receive.⁸²⁷

Even when information gleaned from program monitoring was eventually absorbed, it was not always consistently acted upon. There were times when problems identified by monitoring reports remained unresolved for years. For example, according to the same State inspector general audit, the Bureau of Diplomatic Security failed to use monitoring information to improve program outcomes for training classes intended to increase the ability of Afghan law enforcement officers to combat terrorism. An October 2013 after-action report of one training class noted that course handouts were in English, which none of the students could read. A March 2014 report from the same course reported that the handouts had been translated, but the translation was poor. A February 2015 after-action report described the handouts as outdated, poorly translated and in need

of a rewrite. By July 2016, rather than being improved, the handouts had reverted to a language that numerous students still could not read.⁸²⁸

State's Contract Oversight Apparatus Was Understaffed

An additional challenge was finding enough contracting officers and contracting officer's representatives. As the Commission on Wartime Contracting noted in August 2011, those responsible for operational acquisition sat far down in State's organizational hierarchy, which was a "telling marker of the status of acquisition at the State Department."⁸²⁹ Institutional perceptions within State that contracting was not a core function meant that understaffing of essential contracting functions was a persistent challenge.⁸³⁰

State's 2010 Quadrennial Diplomacy and Development Review, then the guiding document for the agency's efforts globally, noted that stagnant staffing levels combined with an expansion in contracting had "resulted in reliance on fewer, large awards that cover a broad range of activities, with less oversight."⁸³¹ Yet changing institutional habits proved difficult. Despite the Commission on Wartime Contracting's recommendation that State create a contingency contracting office, State viewed that step as "inefficient and unnecessary."⁸³²

Understaffing remains a problem. According to INL officials, their bureau needs at least three and ideally four full-time contracting officer's representatives in Afghanistan for effective contract oversight, but staffing is a "leaky glass" due to the need for rotations, clearances, and rest and relaxation.⁸³³ In September 2017, one permanent INL contracting officer's representative was responsible for monitoring both the Justice Sector Support Program and the Corrections System Support Program, cumulatively worth hundreds of millions dollars.⁸³⁴ By November 2017, that person rotated out and INL had no permanent contracting officer representatives in country. The remaining two temporary staff members faced an overwhelming workload.⁸³⁵ As one official put it, "Most [contracting officer's representatives] are just trying to get through the day. . . . They function more like a boxer trying to get through a round."⁸³⁶ Even though many contracting officer's representatives faithfully worked to perform their duties, sometimes corners were inevitably cut.⁸³⁷ Insufficient staffing introduced potential opportunity costs: With overworked personnel struggling to manage multiple contracts, it was sometimes difficult to perform adequate M&E. In December 2019, INL had five personnel at the U.S. Embassy in Kabul, each handling three or four portfolios consisting of numerous contracts. Two INL officials told SIGAR that they were able to do little to assess actual outcomes, in part because it was challenging to manage the various contracts under these circumstances.⁸³⁸

Attempts to ease this problem were not always effective. In one instance, an attempt by the U.S. embassy in Kabul to hire Afghan nationals to assist contracting officer's representatives resulted in the Afghan nationals performing duties legally executable only by U.S. government staff and in a large number of reports being misfiled.⁸³⁹ In another instance, State attempted to outsource oversight of its antiterrorism assistance program to contractors, resulting in third-party contractors being the only oversight authority, a violation of State policy.⁸⁴⁰

CHAPTER SUMMARY AND KEY FINDINGS

State's M&E systems evolved over time. As the first decade of reconstruction came and went, State steadily formalized and elaborated its M&E systems and processes, which became similar to those of USAID. Across State, performance management plans became required components of every program, evaluations became more widespread, and the Managing for Results framework sought to better connect strategy to budgets and programs. State also diversified the mechanisms by which it performed oversight, hiring third-party monitors as security deteriorated and as officials found it increasingly difficult to visit program sites. This evolution eventually resulted in a complex monitoring and evaluation apparatus that, in theory, can better track program effectiveness.

Yet, as with USAID, State's M&E efforts faced many challenges. Metrics were generally concentrated on outputs rather than outcomes and in some cases did not capture progress against the implicit political objectives that sometimes drove programming. Increasing complexity relegated many aspects of M&E to the realm of "work-makers"—a problem exacerbated by the perennial problem of overworked, understaffed contract oversight personnel. Most fundamentally, key components of State's M&E system, including evaluations, were not always used to periodically retest the theories and assumptions underlying programming. Many of these themes reverberate in the next chapter, which explores DOD's efforts to monitor and evaluate its capacity-building and counterinsurgency efforts in Afghanistan.

Key Findings

- State M&E efforts became more systematized over time.
- That evolution resulted in greater M&E complexity, sometimes creating more work than could be meaningfully performed and more information than could be used.
- Tracking program activities and measuring outputs was more common than outcome and impact assessment.
- There were times when metrics shifted conceptions of "success" towards whatever could be easily counted or measured.
- Some State programs had implicit political objectives that were not accounted for in formal M&E.
- Information produced by M&E was only as useful as the accuracy of assumptions about what data mattered, including assumptions about whether outputs meaningfully correlated with expected outcomes and impacts.
- Evaluations of State's reconstruction programming often demonstrated material analytical flaws, undermining their usefulness.
- M&E systems were not always employed as intended. In some cases, effects were claimed but not substantiated.
- Bureaucratic, legal, and technical obstacles often prevented State from consistently using information provided by M&E to modify programs or contracts.



CHAPTER 4

DOD

US Air Force photo

DOD'S APPROACH TO M&E HAS EVOLVED OVER TIME

DOD's primary mission and capabilities vary significantly from those of State and USAID, leading to different decision-making and contracting practices. One researcher interviewed for this report noted that before 2016, "DOD had no experience with a formal M&E process or [M&E] activities."⁸⁴¹ While this may be true in a technical sense, for a number of years DOD has slowly but consistently developed increasingly formal and sophisticated M&E practices and policies for non-warfighting activities outside of the Afghan context, including security cooperation and humanitarian assistance. Most recently, DOD issued an Assessment, Monitoring, and Evaluation policy for security cooperation in 2017.⁸⁴²

DOD's AM&E policy has several key features that may eventually prove helpful in assessing the effects of its security cooperation efforts. For example, the policy mandates the development of an "initiative design document"—essentially a performance management plan—for "all significant security cooperation initiatives."⁸⁴³ The document is required to describe how a security cooperation initiative links to broader goals or objectives, and to include indicators to measure outputs and outcomes.⁸⁴⁴ It must also include a "comprehensive performance management section" that incorporates a logic framework that visually depicts the process through which activities are assumed to contribute to objectives.⁸⁴⁵ Another key component is the

policy's emphasis on independent and rigorous strategic-level evaluations to measure the effectiveness of significant security cooperation initiatives in meeting expected outcomes.⁸⁴⁶ These are all sound standards. However, DOD's AM&E policy, which has only recently been applied in Afghanistan, is not central to the story of how DOD employed M&E for most of the reconstruction effort.⁸⁴⁷

Many of the M&E systems and processes that DOD did apply sounded good on paper; problems arose in implementation. Broadly speaking, DOD's reconstruction efforts and the M&E-like systems used to assess their efficacy fall into the following categories:

- **Efforts that sought to address challenges that were so complex and abstract that measurement, though attempted, proved extraordinarily difficult**

For example, since 2005, DOD has attempted to evaluate the warfighting and administrative capacity of the Afghan National Defense and Security Forces. Due to the innate difficulties of measuring the capacity of Afghan security force units—including the challenge of forecasting their future ability to operate independently of U.S. support—DOD's assessment methodology has changed at least four times.⁸⁴⁸ As a result, these assessments have been plagued by questions about reliability and consistency.⁸⁴⁹

- **Programs that had some reasonable (though frequently output-oriented) performance metrics and that were sometimes extensively evaluated, but whose actual impact was uncertain**

For example, the Commander's Emergency Response Program supported DOD's counterinsurgency strategy.⁸⁵⁰ Under this program, U.S. military commanders could rapidly fund and implement development projects ranging in category from transportation to agriculture.⁸⁵¹ The program's underlying theory was that development could help quell an insurgency—but research attempting to measure the program's effects on levels of violence has returned inconclusive results.⁸⁵² Another permutation was the Task Force for Business and Stability Operations. Although the task force undertook an economic impact assessment of task force projects, that assessment did not examine whether the task force accomplished its core mission: "to promote stability in Afghanistan."⁸⁵³

- **Programs that had some reasonable (though frequently output-oriented) metrics but whose impact has not been determined**

In April 2011, the Congress created the Afghanistan Infrastructure Fund, which was intended to support U.S. counterinsurgency goals by building large-scale water, power, and transportation projects.⁸⁵⁴ Although the progress of construction was monitored, DOD did not assess whether these projects actually achieved their counterinsurgency objectives.⁸⁵⁵ Instead, DOD at one point said such an assessment was an "intuitive process," and essentially cited counterinsurgency theory itself as evidence of the infrastructure program's future outcomes and impact.⁸⁵⁶

- **Programs for which few performance metrics were identified and whose impact was not directly assessed**

For example, since 2010, DOD has awarded at least four contracts, worth at least \$1.62 billion combined, to DynCorp International to provide contracted advisors at the Afghan ministries of defense and interior.⁸⁵⁷ However, performance work statements for the contracts did not always clearly describe the work DynCorp was to accomplish in a way that could be credibly measured and assessed.⁸⁵⁸

Comprehensively assessing DOD's efforts to monitor and evaluate its reconstruction activities proved more challenging than examining State's and USAID's. Partly, this is because DOD did not have an overarching, uniformly applied M&E system—meaning there was no single set of comprehensive standards we could apply to every program. Additionally, while DOD is one of many agencies operating in Afghanistan, the scale of its reconstruction activities—including contracted programs—is far broader than State and USAID's. As of March 31, 2021, \$88.3 billion had been allocated for security-related reconstruction, compared to \$36.0 billion for governance and development.⁸⁵⁹

We examined the challenges DOD faced in monitoring and evaluating its programming through illustrative 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 (see Table 7 below). We did this by focusing on high-dollar-value programs within these areas and by discussing measurement systems developed by DOD to assess overall progress in building the Afghan National Defense and Security Forces' capacity.

The last part of the chapter explores DOD's overarching M&E-like process: operation assessment. Assessment is doctrinally defined as:

A continuous process that measures the overall effectiveness of employing . . . capabilities during military operations. It involves monitoring and evaluating the current situation and progress toward mission completion. Assessments can help determine whether a particular activity contributes to progress with respect to a set of standards or desired objective or end state.⁸⁶⁰

In other words, operation assessments seek to periodically determine the aggregated effects of all inputs—including those related to DOD contracts and grants—on overall

TABLE 7

DOD RECONSTRUCTION PROGRAMMING IN AFGHANISTAN

Name of Fund	Cumulative Appropriations (millions)	Description	Line of Effort
Afghanistan Security Forces Fund (ASFF)	\$82,899.8	Intended to build the capability of Afghanistan's military and police forces	Reconstructing the ANDSF
Commander's Emergency Response Program (CERP)	\$3,711.0	Provided field commanders with funding to implement urgent reconstruction projects to reduce violence and improve the legitimacy of the Afghan government	Counterinsurgency through economic development
Afghanistan Infrastructure Fund (AIF) ^a	\$988.5	Funded large-scale infrastructure projects in Afghanistan	Counterinsurgency through economic development
Task Force for Business and Stability Operations (TFBSO)	\$822.9	Funded economic development projects that focused on the private sector and aimed to reduce violence, enhance stability, and promote economic normalcy in Afghanistan	Counterinsurgency through economic development

^a AIF funded both DOD and USAID projects. Of the total, 67% of obligated funds (\$682.8 million) went to DOD. No additional funding has been appropriated since FY 2014. Remaining projects involve only power-infrastructure.

Note: DOD's \$3.3 billion Drug Interdiction and Counter-Drug Activities Fund (DOD CN) is excluded from this table. As of March 1, 2018, DOD CN was paying for contracted logistics support for Mi-17 and PC-12 aircraft used by Afghanistan's Special Mission Wing. It was also funding base security and base operations and maintenance contracts, including dining facility and ablution services. These kinds of contracts do not fall within the scope of this report.

Source: SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, pp. 168, 174; SIGAR, *Stabilization: Lessons From the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May, 2018, pp. ix, 93; SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January, 2018, p. 1; SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, p. 49.

military campaign objectives. For a variety of reasons, operations assessment was often a fraught process in Afghanistan. Common issues included the tendencies for assessments to be overly optimistic, to collect too much data, and to misuse quantitative rating schemes.⁸⁶¹ We discuss these and other challenges DOD faced in determining the overall effects of reconstruction inputs. The chapter concludes with a brief summary of key points.

RECONSTRUCTING THE ANDSF: MEASURING OPERATIONAL EFFECTIVENESS PROVED DIFFICULT

Since 2002, DOD's attempts to reconstruct, build, and sustain combat operations of the Afghan National Defense and Security Forces (ANDSF) have been a key component of U.S. efforts to stabilize Afghanistan. As of March 31, 2021, security-related reconstruction funds appropriated to DOD comprised approximately 61 percent of total reconstruction appropriations.⁸⁶²

But measuring the actual progress of rebuilding the ANDSF proved extraordinarily challenging. Rating systems designed to capture ANDSF operational effectiveness have shifted over time, reflecting persistent questions about what data mattered, how to balance quantitative and qualitative information, and, more fundamentally, how exactly to measure capability in the first place. The first-order question of how to accurately project how the ANDSF would perform against an adversary in the absence of direct U.S. combat enabler support—which will end no later than September 2021—remains difficult to answer and the systems designed to measure that capability have been criticized for being inconsistent.⁸⁶³ This section provides a brief history of U.S. efforts to build the ANDSF, and explores the evolution of the rating systems that sought to capture ANDSF capacity and performance. It ends with a discussion of contracts aimed at building ANDSF capacity through ministerial advising at the ministries of defense and interior.

A Brief History of Reconstructing the ANDSF

In 2002, the United States and its coalition partners concluded that developing a professional Afghan national security force could mitigate the need for an expansion of international forces in Afghanistan.⁸⁶⁴ By May 2002, the United States had begun training a new Afghan National Army. The training program quickly expanded from developing small infantry units to larger military formations.⁸⁶⁵ In 2004, a deteriorating security environment convinced U.S. policymakers that the original plan—in which Germany took the lead on developing Afghanistan's police force while the United States focused its efforts on Afghanistan's military—was not working. In 2005 the United States assumed the lead on both military and policing assistance.⁸⁶⁶

To accomplish this new combined mission, in 2006 the United States created the Combined Security Transition Command – Afghanistan (CSTC-A), responsible for training, advising, assisting, and equipping the entire ANDSF, military and police forces alike.⁸⁶⁷ With insecurity mounting, the United States rapidly expanded the ANDSF on a



An American soldier instructs Afghan National Police officers in the proper use of a safety shield.
(U.S. Army photo)

condensed training and development timeline.⁸⁶⁸ As part of this expansion, the United States began transitioning the Afghan National Army from a light-infantry army to a combined-arms service with army, air force, and special forces elements. The train, advise, and assist programs for these specialized forces resulted in fairly proficient Afghan Special Forces and a promising Afghan Air Force.⁸⁶⁹ To supplement the strength of security forces, the United States also attempted to create auxiliary police units, many of which were functionally equivalent to local militias. Some such units were reported to have engaged in human rights abuses, drug trafficking, and other corrupt activities, detracting from security.⁸⁷⁰

In 2009, with the ANDSF struggling to secure the country in the face of a resurgent Taliban, President Barack Obama authorized a surge of U.S. combat forces. At the same time, President Obama announced a withdrawal date for U.S. combat forces and directed that responsibility for security be transferred to the ANDSF beginning in mid-2011.⁸⁷¹ Conforming to the president's plan, the U.S. military sought to rapidly improve security while continuing to develop the struggling ANDSF. In pursuit of quick security gains, the U.S. military substituted its capacity for that of the Afghans by filling key gaps in ANDSF capability, such as medical evacuation and close air support. With Afghan units modeling their fighting after the United States, they became increasingly dependent on U.S.-provided advanced weapons systems, vehicles, and other equipment—which, because of widespread illiteracy in its ranks, the ANDSF could not reliably use and maintain on its own.⁸⁷² Chronic understaffing of the train, advise and assist effort added to the difficulties of building the ANDSF, partially because NATO allies never fully lived up to commitments to provide much needed trainers.⁸⁷³



Resolute Support advisors from the United States and other coalition countries remotely meet with their Afghan Ministry of Defense counterparts. (Resolute Support photo)

Not surprisingly, as the United States and NATO neared the conclusion of a phased drawdown in 2014 and transitioned to training and advising at the regional and institutional level, the ANDSF struggled to succeed.⁸⁷⁴ It was not until 2015 that the United States and NATO prioritized security sector governance and defense institution building over improving the fighting capabilities of the force. Up to then, developing ministerial capability was primarily focused on initiatives that aimed to more directly bolster combat effectiveness. But with the U.S. military withdrawing, concern shifted to the issue of the long-term viability of the ANDSF. With that in mind, in January 2015 DOD prioritized its train, advise, and assist focus to higher levels of command (the Afghan National Army corps level, the Afghan National Police zone level, and the ministries of defense and interior). As a result, the number of U.S. touchpoints with field-based ANDSF units has declined significantly, prompting increased reliance on Afghan-provided data to understand unit-level performance and capability gaps.⁸⁷⁵

Over time, rising insecurity further restricted the U.S. advising effort. Civilian advisors who were once able to drive themselves to various government offices around Kabul began to transit to the ministries of defense and interior in convoys, or even by helicopter. The U.S. military, meanwhile, was generally confined to large bases.⁸⁷⁶

Recent events have only exacerbated the challenges. The confluence of significant developments is daunting: a stalled intra-Afghan peace process, ongoing conflict, the COVID-19 pandemic, Afghanistan's resulting economic contraction, and, in April 2021, President Joseph R. Biden's decision to withdraw U.S. troops from Afghanistan by September 11, 2021.⁸⁷⁷ U.S. defense contractors will withdraw along with U.S. troops, raising concerns about the sustainability of Afghan security force components that are especially reliant on contractors to maintain vital equipment.⁸⁷⁸

The next section discusses DOD's efforts to monitor and evaluate the effectiveness of the ANDSF over the last 16 years.

M&E of Reconstructing the ANDSF: Shifting Attempts to Measure the Abstract

In July 2005, the United States started implementing assessment mechanisms to evaluate the warfighting and administrative capacity of the ANDSF. These mechanisms functioned as an M&E tool to track ANDSF performance, identify gaps, and tailor advisory efforts. Since 2005, the ANDSF assessment methodology has changed at least four times, reflecting the inherent challenges of attempting to measure capacity, as well as persistent questions about the nature of data required to do so.⁸⁷⁹

The first mechanism was the Capability Milestone rating system, which was replaced by the Commander's Unit Assessment Tool in April 2010.⁸⁸⁰ In July 2013, the latter was replaced by the Regional ANDSF Assessment Report.⁸⁸¹ In January 2015, that changed to the Monthly ANDSF Assessment report.⁸⁸² Individual advisor evaluations and reports have been built into the Advisor Network system since January 2015, with a recent modification in April 2020.⁸⁸³

These systems were fundamentally similar in their use of rating scales: All of them graded Afghan units along comparable spectrums of dependence on coalition support. However, they emphasized different inputs, analyzed different levels of command, varied in their thresholds for achieving a given score, and in many cases used different words to describe individual rating levels.⁸⁸⁴ Moreover, the top rating levels between systems differed in their focus on sustainability. Some systems included only a secondary focus on sustainability, while others explicitly prioritized it as a necessary condition for achieving top ratings.⁸⁸⁵ To some degree, these changes and differences reflected the inherent challenges in accurately measuring true ANDSF capabilities, and the struggle to find the right balance between qualitative and quantitative data in the assessments.⁸⁸⁶ Still, because they seemed to change so frequently, the systems have also been criticized for being inconsistent.⁸⁸⁷ Each of these assessment mechanisms is described below, followed by a comment on the role of third-party independent assessments.

Capability Milestones: Missing the Intangible

The Capability Milestone ratings scale divided ANDSF units into four groups, based on the amount of international community support they required to conduct primary operational missions.⁸⁸⁸ It focused on the *kandak* (battalion) level, and emphasized quantitative inputs that, in theory, reflected levels of readiness, such as staffing levels and equipment.⁸⁸⁹ That failed to measure more intangible readiness factors, such as leadership, corruption, and motivation—all factors that could affect a unit's ability to put its staffing and equipment to use during actual warfighting. This led some commanders to conclude that the mechanism did not adequately measure operational effectiveness.⁸⁹⁰ Data reliability and consistency also posed a major challenge: there was a shortage of U.S. personnel to verify reported data, and even basic statistics such as the number of evaluated units in a given time period varied significantly between published DOD reports.⁸⁹¹



U.S. Marines discuss operations with Afghan police officers in Helmand in late 2012.
(U.S. Marine Corps photo)

These challenges contributed to perceived weaknesses in the Capability Milestone system. First, there were large capability discrepancies between units with the same rating.⁸⁹² Second, ratings proved unreliable at measuring an ANDSF unit's ability to sustain a given capability level over time, since many units regressed at least one rating even during relatively short observation periods.⁸⁹³

Commander's Unit Assessment Tool: Balancing Quantitative and Qualitative Data

After the International Security Assistance Force Joint Command assumed responsibility for assessing ANDSF forces during its creation in 2009, it replaced the Capability Milestone system with the Commander's Unit Assessment Tool in April 2010.⁸⁹⁴ The new mechanism also focused on the kandak level, but sought to address weaknesses in the previous rating system by including both quantitative and qualitative components. Quantitative data included force strength numbers and equipping levels; qualitative data included subjective assessments of leadership quality and competency, as well as unit morale.⁸⁹⁵ Although this balance added more detail, it also introduced greater complexity. This required additional training and guidance on rating definitions—and since many ANDSF advisors did not receive this training, the ratings were inconsistent and possibly inaccurate.⁸⁹⁶ Further complicating the matter, the rating system itself changed from April to July to September of 2010. Inconsistent ratings and continued data reliability challenges meant that the Commander's Unit Assessment Tool still did not fully capture the ANDSF's capabilities.⁸⁹⁷

Regional ANDSF Status Report: Addressing Inconsistency

In July 2013, the International Security Assistance Force (ISAF) Joint Command replaced the Commander's Unit Assessment Tool with the Regional ANDSF Status Report.⁸⁹⁸ The status report attempted to shift from assessing theoretical indicators

of readiness to measuring ANDSF units' ability to use their staffing, equipping, and training to successfully engage in combat.⁸⁹⁹ It also focused on assessing a small number of brigade and regional level units with greater consistency every month, which in theory allowed ISAF to focus on units of critical importance.⁹⁰⁰ To address the issue of inconsistent ratings under the previous mechanism, the ISAF Joint Command also issued more detailed instructions on the data that advisor units should include in the comments supporting their ratings, and started conducting quality control checks of the provided narratives.⁹⁰¹ In mid-2014, the ISAF Joint Command determined that future ANDSF capability assessments would be classified, reducing the transparency of the status report.⁹⁰²

Monthly ANDSF Assessment Report: Increased Focus on Higher Levels of Command

The transition from the Regional ANDSF Status Report to the Monthly ANDSF Assessment Report took place in January 2015, along with the transition from the ISAF combat mission to the Resolute Support train, advise, and assist mission.⁹⁰³ The decreased coalition presence meant that the monthly report focused only on the headquarters level of command instead of the operations level; it was not intended to be used to assess the entire ANDSF.⁹⁰⁴ In addition to the monthly report, U.S. Forces – Afghanistan adopted a broad assessment framework with eight essential functions meant to assess whether components of Afghanistan's security forces were prepared to function without coalition advising and support.⁹⁰⁵ The primary criticism of the monthly report was that its lack of coverage below the ANDSF headquarters level precluded a complete evaluation of the ANDSF's operational effectiveness.⁹⁰⁶

Advisor Network: A Further Attempt to Capture Qualitative Information

Since 2015, largely qualitative advisor assessments of Afghan counterpart officers and units have been built into the Advisor Network, an electronic software system that helps advisors track engagements with and assess the performance of the ANDSF.⁹⁰⁷ Advisors use the network to log meetings with Afghan counterparts and provide ranked assessments of various criteria based on their observations.⁹⁰⁸ According to CSTC-A, average assessment scores can help advisors ascertain progress toward key goals, indicate which strategies are working, and enable more effective use of resources.⁹⁰⁹ While the network is not a perfect analogue for the other assessment mechanisms described here, it serves similar purposes in determining where advisory efforts should focus and enabling strategic decision making.⁹¹⁰

The original evaluation system suffered from a subjective narrative format and a tendency to elicit more information than was necessary or relevant. In an effort to produce more rigorous and less arbitrary assessments, the new Advisor Network assessment mechanism from April 2020 employs a **Likert scale** to generate quantifiable performance data that can create historical trend lines.⁹¹¹ Although, as of March 2020, CSTC-A had not yet definitively shown how the new mechanism would control for such factors as **social desirability bias**, it views the new system as a significant upgrade.⁹¹²

A **Likert scale** is a rating scale commonly used in surveys to measure people's perceptions, opinions, or attitudes, often with categories of response that can be coded numerically such as strongly agree, agree, neutral, disagree, and strongly disagree.

Social desirability bias is the tendency "for participants to respond in a way that they believe will be viewed favorably by others."

Britannica, Likert-Scale, accessed July 5, 2020, p. 1; ScienceDirect, "Social Desirability Bias," accessed July 5, 2020, p. 1.

The Military's Perspectives on ANDSF Capability Have Sometimes Been Called into Question

Large discrepancies in various reports of ground progress generated some frustration for the Congress in the early 2010s.⁹¹³ On one hand, official statements of progress often implied successes in a variety of areas. For example, in March 2011 testimony to the Congress, General David Petraeus—then the commander of the International Security Assistance Force—claimed that “investments in leader development, literacy, and institutions have yielded significant dividends” for the ANDSF, that Afghan forces were taking on significant combat roles against the Taliban, and that Afghan Local Police units were increasingly limiting the Taliban’s ability to intimidate communities.⁹¹⁴ Other reports indicated “the absence of success on virtually every level.”⁹¹⁵ In a 2012 *Armed Forces Journal* article, Lt. Col. Daniel Davis, who spent a year in Afghanistan speaking with U.S. troops and their Afghan counterparts, wrote that his observations “bore no resemblance to rosy official statements by U.S. military leaders about conditions on the ground.”⁹¹⁶ Among other evidence supporting this conclusion, Davis cited a conversation with a local official who assessed that the ANDSF “are definitely not capable,” and recalled another province where “the Afghan policemen rarely leave the cover of the checkpoints” and the Taliban “literally run free.”⁹¹⁷



An Afghan Air Force finance technician attends a training class in the use of Microsoft Excel in 2018. (DOD photo)



Afghans assigned to the 203rd Thunder Corps lead and execute a clearance operation in southeast Afghanistan, accompanied by U.S. advisors in 2019. (U.S. Army photo)

Independent Assessments Were Also Used to Ascertain the ANDSF's Operational Effectiveness

Through the 2013 National Defense Authorization Act, the Congress directed DOD to conduct an independent assessment of what would be needed to create an ANDSF capable of providing security and protection from terrorism in Afghanistan.⁹¹⁸ The Office of the Under Secretary of Defense for Policy contracted with CNA, an Arlington, Virginia-based federally funded research and development corporation, to conduct this independent assessment. CNA released its report in January 2014. At the request of the Office of the Under Secretary of Defense for Policy, CNA published a second report in 2015.⁹¹⁹ Both reports focused on aggregate measures of ANDSF readiness, and did not replace U.S. military assessments that aimed to measure the capacity of individual units.⁹²⁰

The assessments added value to DOD's more routine attempts to measure ANDSF capacity by evaluating not only the current capabilities of the force, but also the future security environment. Projections of future security, combined with an analysis of the ANDSF's present capability, led to recommendations about how the force might be restructured and better employed to combat the insurgency. For example, CNA's 2014 assessment recommended that elements of the ANDSF be redistributed to higher-threat areas and that some combat battalions be replaced with more logistics and support units.⁹²¹ Although both assessments calculated ideal force authorization levels that were close to the existing force authorization of 382,000, they recommended adjustments to the force structure that would result in fewer combat battalions but more logistics and support battalions.⁹²²

Some aspects of the assessments proved prescient. While the 2015 fall of Kunduz City (located in a northern Afghan province of the same name) surprised some observers, CNA's 2014 assessment had warned that "the Taliban insurgency [would] become a greater threat to Afghanistan's stability in the 2015–2018 time frame than it [was at the time of the assessment]."⁹²³ Other observations seemed self-evident: the 2015 assessment concluded that "the performance of the ANDSF so far this fighting season has been decidedly mixed, and that the force continues to suffer from a number of significant capability gaps."⁹²⁴

Overall, independent assessments are best thought of as a potentially useful complement to, rather than a replacement for, DOD's internal mechanisms. Comparisons between internal and external assessments can help identify irregularities and discrepancies.⁹²⁵

The Afghanistan Compact

The Afghanistan Compact was launched in August 2017 by the U.S. and Afghan governments.⁹³² It was a voluntary, non-binding agreement designed to track the Afghan government's performance in meeting key reform benchmarks in various sectors, including security.⁹³³ Although DOD has said that the Compact "was not a method to assess progress," the use of benchmarks provides the basic framework for a quasi-assessment mechanism.⁹³⁴ One weakness of the Compact was its unmanageable number of security-related benchmarks: It outlined 257 security-related milestones, compared to 163 benchmarks for the other three areas combined.⁹³⁵ The number of security-related benchmarks eventually increased to over 1,300, a number that cannot be feasibly evaluated or meaningfully completed.⁹³⁶ Because using these benchmarks did not seem to improve the day-to-day workings of Afghan security institutions, CTSC-A began experimenting with other frameworks in 2019.⁹³⁷ According to DOD, the Compact is no longer in use and is best thought of as a list of relatively minor tasks.⁹³⁸

M&E of Contracted Advisors: Tracking Inputs without Measuring Direct Outcomes

At the ministerial level, U.S. advising efforts focused on improving the administrative and technical capabilities of Afghanistan's ministries of defense and interior. Advisors fell into a variety of categories.⁹²⁶ Some were uniformed military personnel, some were civilian employees of DOD with specialized skills or experience, and a significant number were defense contractors.⁹²⁷ As of July 2017, contractors made up approximately 67 percent of the total 577 advisors at the ministries.⁹²⁸ Recent examples of their areas of focus include long-term budgeting, helping ANDSF leadership strengthen its ability to identify future requirements, and helping to improve internal control mechanisms to guard against corruption and prevent waste, fraud, and abuse.⁹²⁹ Since 2010, DOD's contractor for ministerial advising has been DynCorp International, which has provided personnel with expertise in such areas as finance and procurement.⁹³⁰ The cumulative value of DOD's advising contracts with DynCorp is at least \$1.62 billion.⁹³¹

Tracking how effectively these contractors were contributing to DOD's ministerial capacity-building objectives has been a challenge for DOD. There were times when it seemed to be out of compliance with Federal Acquisition Regulation standards requiring the use of meaningful metrics that can be used to judge contractor performance. The FAR requires that agencies "enable assessment of work performance against measurable performance standards" to the maximum extent practicable, and requires agencies to identify a "method of assessing contractor performance against performance standards."⁹³⁹ These requirements are in place to ensure the quality, timeliness, and quantity of services provided (Appendix B provides further information on the FAR).⁹⁴⁰

In October 2018, SIGAR reported that DOD was unable to track DynCorp's performance. The contracts' performance work statements—key documents that, according to regulation, must describe the work contractors are supposed to perform in terms of measurable outcomes—provided four broad requirements for DynCorp to follow. But they did not include the kinds of measurable performance standards required by the FAR.⁹⁴¹ Instead, the performance work statements merely required DynCorp to "provide advisors and mentors" and "provide advice and mentoring."⁹⁴² A DOD quality assurance representative who monitored the day-to-day performance of the contractor told SIGAR that the contracts did not specify measurements of progress.⁹⁴³

Human intelligence is any intelligence derived from information collected and provided by human sources.

DOD, "DOD Dictionary of Military and Associated Terms," January 2020, p. 98.

The Legacy and Afghanistan Source Operations Management Programs: Lack of Performance Metrics Made It Difficult to Determine Return on Investment

The DynCorp train, advise, and assist contracts were not the only outsourced capacity-building efforts that lacked clear metrics to determine progress towards outcomes and to ascertain impact. DOD's intelligence development contracts had the same problems.⁹⁴⁴

From 2010 to 2016, DOD administered the \$457.7 million Legacy and Afghanistan Source Operations Management (ASOM) programs. The programs aimed to develop the ANDSF's **human intelligence** capabilities through training and mentoring.⁹⁴⁵ Legacy was initially structured as a package of four research and development contracts that were awarded in 2010. (Legacy's successor, ASOM, was awarded as a single contract in 2013.) Research and development contracts are unique in that they are generally "directed toward objectives for which the work or methods cannot be precisely described in advance," according to the FAR.⁹⁴⁶ Consequently, DOD's use of the contracting vehicle did not require the development of adequate performance metrics or clearly define outcomes, making Legacy's effectiveness difficult to ascertain.⁹⁴⁷

The last contract awarded for the Legacy program ended in 2013.⁹⁴⁸ The intelligence capacity-building program was continued under ASOM through the award of a services contract. Unlike research and development contracts, services contracts are required by the FAR to include measurable performance standards.⁹⁴⁹ However, SIGAR found that the ASOM contract still lacked clear performance indicators.⁹⁵⁰ Under ASOM, the primary indicator by which success was measured was the successful transition of intelligence sites to the Afghan government.⁹⁵¹ Although this may have been a useful metric in theory, it was not objective: The contractor itself was responsible for assessing each intelligence training site's readiness to transition to the Afghan government.⁹⁵²

Ultimately, because of a lack of performance metrics for the Legacy contracts and a reliance on contractor-provided data for the ASOM program, it was difficult to gauge the government's return on its \$457.7 million investment.⁹⁵³

In September 2018, DOD produced a draft performance work statement for a follow-on advising contract that incorporated two general performance requirements. These aimed to measure the contractor's performance by assessing the extent to which CSTC-A was satisfied with the efforts of the contractor and how well the contractor adhered to important intermediate goals.⁹⁵⁴ Still, the lack of performance metrics in prior contracts made it difficult for DOD and others to assess the effectiveness of the advising effort over time.

COUNTERINSURGENCY THROUGH ECONOMIC DEVELOPMENT: WAS THE UNDERLYING THEORY VALID?

Although economic development is often considered a civilian activity, the military spent considerable resources on stabilization programming in Afghanistan, with the intent of reducing violent resistance and swaying popular support towards coalition forces and the Afghan government.⁹⁵⁵ DOD's economic development programs reflected the theory, promulgated in the U.S. Army's December 2006 *field manual on counterinsurgency*, that "political, social, and economic programs are usually more valuable than conventional military operations in addressing the root causes of conflict and undermining an insurgency."⁹⁵⁶ Counterinsurgency doctrine was embraced by some military leaders and civilian analysts as "the authoritative playbook for success," in the words of Karl Eikenberry, U.S. ambassador to Afghanistan from 2009 to 2011 and a critic of the counterinsurgency strategy there.⁹⁵⁷ Indeed, in the words of one historian, it was the "holy writ."⁹⁵⁸

But some observers cast doubt on aspects of counterinsurgency theory.⁹⁵⁹ An early academic review of the new doctrine questioned both its historical and theoretical underpinnings, and criticized its lack of grounding in social science research.⁹⁶⁰ Scrutiny of the theory increased as the war continued, often focusing on the extent to which development aid could be employed to achieve counterinsurgency effects. As a June 2011 staff report from the Senate Committee on Foreign Relations bluntly stated:

Our stabilization strategy assumes that short-term aid promotes stability in counterinsurgency (COIN) operations and "wins hearts and minds" by improving security, enhancing the legitimacy and reach of the central government, and drawing support away from the Taliban. It presumes that the international community and the Afghan government have shared objectives when it comes to promoting longer-term development, good governance, and the rule of law. These assumptions may not be correct.⁹⁶¹

The Senate report ultimately concluded, "Given the conflicting research on the effects of aid in promoting stability, more analysis is needed before we continue investing a significant amount of our aid in conflict zones."⁹⁶² Among the report's sources was a conference readout from *Wilton Park* summarizing the views of leading experts on the ways in which development aid could (or could not) be used to achieve counterinsurgency objectives.⁹⁶³ The readout acknowledged that development aid could have some tactical benefits, such as helping to build relationships between coalition forces and local communities or gathering intelligence. However, in line with the broader Senate committee report, it too concluded there was "little evidence"

Intended to fill a doctrinal gap, the U.S. Army's December 2006 *field manual on counterinsurgency* (also known as FM 3-24) sought to provide U.S. Army soldiers and U.S. Marines fighting in Afghanistan and Iraq with principles and guidelines for counterinsurgency operations. Before its publication, it had been 20 years since the Army had issued field manuals devoted exclusively to counterinsurgency, and 25 years since the Marine Corps had done so.

U.S. Army, "FM 3-24: Counterinsurgency," December 2006, electronic pp. 1-2.

Wilton Park is a UK-based organization that convenes discussions on topics of strategic interest, including defense and security issues.

Wilton Park, "About Us," n.d., accessed August 11, 2020.



Members of Khost Provincial Reconstruction Team meet with engineers and beneficiaries to discuss an upcoming infrastructure project on May 2, 2010. (DOD photo)

that these kinds of “relatively short-term transactional relationships” produced “more strategic-level effects of populations being won over to the government as a result of development aid.”⁹⁶⁴ The central conclusion of the Wilton Park conference readout was striking for its implications: “Given its centrality to current COIN doctrine and strategy, there is still a surprisingly weak evidence base for the effectiveness of aid in promoting stabilization and security objectives.”⁹⁶⁵

Despite the questions they raised about the idea that development aid could produce counterinsurgency effects, aspects of both the 2011 Senate report and the 2010 Wilton Park conference report reflected an evolving conversation around the effectiveness of one particularly prominent stabilization program: the Commander’s Emergency Response Program. Both reports pointed to a study examining the effects of CERP in post-surge Iraq that concluded CERP was effective in reducing violence.⁹⁶⁶ But exactly what drove the finding was unclear. As one academic observed in 2017, “A subsequent study by the same authors qualifies that result in important ways. [The authors] find that the violence dampening effect actually occurs only in interaction with increased troop presence.”⁹⁶⁷

Was CERP effective? If so, under what circumstances? These were key questions for M&E that were complicated by both the prevailing faith in counterinsurgency doctrine and metrics that were conflated with the act of implementation.⁹⁶⁸ For many military and civilian officials, the answer was self-evident: CERP worked. Writing in his personal capacity in the February 2004 edition of the *Army Lawyer*, then-Deputy Legal Counsel for the Joint Chiefs of Staff Mark Martins said, “The program’s early success demonstrates that relatively small amounts of money spent locally and intelligently by commanders can yield great benefits.”⁹⁶⁹ Testifying before the Senate in 2008, then-Secretary of Defense Robert Gates called CERP the “single most effective program to enable commanders to address local populations’ needs and get potential insurgents in Iraq and Afghanistan off

the streets and into jobs.”⁹⁷⁰ In concurrent testimony, then-Chairman of the Joint Chiefs of Staff Admiral Michael Mullen asserted that “CERP has proven in most cases more valuable and perhaps more rapid than bullets or bombs in the fight against extremism.”⁹⁷¹ In the introduction to the March 2012 standard operating procedures for CERP in Afghanistan—which outlined how CERP projects should be monitored and evaluated—then-U.S. Ambassador to Afghanistan Ryan Crocker and then-ISAF Commander General John Allen stated that CERP was “vital to the [counterinsurgency] strategy.”⁹⁷² But the end result was \$2.3 billion spent in exchange for unknown effects.⁹⁷³

The Commander’s Emergency Response Program: A Brief History

Congress authorized the Commander’s Emergency Response Program in Afghanistan in November 2003. Initially created for Iraq, it enabled military commanders to “respond to urgent humanitarian relief and reconstruction requirements within their areas of responsibility,” and to improve security by creating jobs and “winning hearts and minds.”⁹⁷⁴ (The highlight on the following page discusses CERP’s purpose, which was not always clearly articulated.) Although funds could only be spent on certain categories of projects, U.S. military commanders had wide latitude to identify and select projects based on local conditions.⁹⁷⁵ Overall, from 2004 to 2020, CERP supported projects in all 34 of Afghanistan’s provinces, disbursing a total of \$2.3 billion.⁹⁷⁶

When the program arrived in Afghanistan in 2004, CERP disbursements were \$35 million. At their peak in 2009, coinciding with an increase in troop numbers and operations, disbursements were \$500 million.⁹⁷⁷ CERP activity from 2004 to 2009 was concentrated in the eastern and southeastern provinces, where funds were used for a wide range of projects, including transportation, education, agriculture, water, and sanitation.⁹⁷⁸ With the U.S. troop surge that began at the end of that period and lasted until 2012, and the adoption of a modified counterinsurgency strategy, the program increasingly focused on improving government legitimacy and displacing Taliban influence—or as an ISAF Joint Command document put it, to “protect the population, neutralize insurgent networks, and promote effective governance that is acceptable to the Afghan people.”⁹⁷⁹ After taking responsibility for the program in 2009, United States Forces – Afghanistan encouraged military commanders to implement CERP projects after clearing operations as a way of demonstrating the benefits that would accompany Afghan government presence. The hope was the government would come to be seen as legitimate and capable, and the Taliban would not be welcome to return.⁹⁸⁰

After 2009, the focus of CERP spending began to focus on agricultural projects and other attempts to improve local economic conditions. While overall disbursements declined as the surge took shape, the number of initiated projects increased. Small, short-term projects of less than \$5,000 accounted for 95 percent of all CERP projects at the height of the U.S. and international troop surge in 2011.⁹⁸¹ Although the program was a key component of the surge, declining disbursements and the shift towards smaller projects were attributable in part to new restrictions on projects over \$500,000 and the creation of the Afghanistan Infrastructure Fund (discussed further later in this chapter).⁹⁸² With

WHAT WAS CERP'S PURPOSE?

CERP's purpose was not always clearly articulated. The Congress appropriated funds for the Commander's Emergency Response Program to enable military commanders "to respond to urgent humanitarian relief and reconstruction requirements within their areas of responsibility."⁹⁸³ But exactly what responding to those requirements was meant to accomplish was not self-evident. The terms "humanitarian relief" and "reconstruction" do not necessarily imply security or stability-related objectives.⁹⁸⁴ Yet, the aforementioned 2004 article authored by Mark Martins (expressing his personal views while serving as the deputy legal counsel in the Office of the Chairman of the Joint Chiefs of Staff) described CERP as "a program by which field commanders . . . can fund initiatives to win hearts and minds [and] hunt enemies" (in addition to "[promoting] the growth of local institutions").⁹⁸⁵ In Senate testimony in March 2004, then Under Secretary of Defense (Comptroller) Dov Zakheim stated that the purpose of CERP was "not only [to] help people . . . but also gain their support in defeating terrorists" because "where we establish a relationship in the community . . . the information starts to flow."⁹⁸⁶ In a subsequent 2005 article, Martins describes an interaction between Joint Staff members and Senate staffers in which military officers "related CERP to a stabilization tool no less essential to victory than the world's finest military equipment."⁹⁸⁷

Indeed, there was a tension between how the initial appropriations language described the purpose of CERP and how and why CERP funds were actually employed. Limited restrictions on CERP spending meant that commanders were free to use CERP according to their understanding of what the funding was intended for. In an interview with senior congressional staffers conducted for a RAND impact assessment, "several interviewees emphasized that there never was a congressional 'intent' for CERP, as the program was not developed by Congress but was simply a DOD request."⁹⁸⁸

According to the RAND assessment, "The pervasive view in Congress in support of CERP, at least initially in Iraq, was that it was a 'force protection' measure to keep

U.S. troops safe."⁹⁸⁹ As the Special Inspector General for Iraq Reconstruction described in a report exploring perceptions of CERP, "Many commanders noted that, regardless of project type, their ultimate goal in the use of CERP was to reduce violence and, hence, casualties."⁹⁹⁰ A majority of commanders surveyed considered the level of violence to be the most important indicator of a CERP project's success.⁹⁹¹

One view is that force protection is a distinct concept from counterinsurgency.⁹⁹² But some of the primary pathways through which force protection was to be attained are consistent with counterinsurgency theory. For example, the RAND assessment noted that one pathway was "job creation, based on the 'theory that if people are working, they won't rent themselves out to the insurgency.'" Another was by "winning 'hearts and minds.'"⁹⁹³ Even before the promulgation of counterinsurgency doctrine, U.S. soldiers in Iraq were employing CERP, under such allowable categories as humanitarian aid and battle damage payments, to create goodwill and deprive the enemy of support, as well as to gain intelligence.⁹⁹⁴ The "Sons of Iraq" militias, considered a major counterinsurgency success, were funded with more than \$300 million of CERP money—under the rubric of "temporary contract guards for critical infrastructure" in order to comply with spending guidelines.⁹⁹⁵

Overall, CERP is best viewed as a tool for implementing the aspects of counterinsurgency theory implicit in the title of the program's guiding document for much of the war: "Money as a Weapon System."⁹⁹⁶ As a 2009 Center for Army Lessons Learned handbook described, by "motivating antigovernment forces to cease lethal and nonlethal operations, by creating and providing jobs along with other forms of assistance to the indigenous population, and by restoring or creating vital infrastructure," commanders were ultimately "defeating [counterinsurgency] targets without creating collateral damage."⁹⁹⁷ As the RAND assessment noted, "Most CERP implementers in Afghanistan nested the use of CERP within an overall counterinsurgency strategy, which involved contesting the Taliban by improving security, governance, and development."⁹⁹⁸

the surge concentrated in the south, program spending was redistributed to the Taliban heartland in Helmand and Kandahar.⁹⁹⁹

As U.S. military forces began to transition security responsibilities to the ANDSF in 2012, the number of new CERP projects began to decline. With the withdrawal of U.S. forces from the south and east between 2012 and 2014, the program's focus pivoted toward Kabul and the neighboring provinces of Wardak, Parwan, and Logar. By 2014, those areas accounted for approximately 90 percent of CERP spending. By December 2014, Operation Enduring Freedom in Afghanistan had concluded. Between fiscal years 2015 and 2020, Congress appropriated \$37.5 million for CERP and DOD disbursed approximately \$18 million.¹⁰⁰⁰

Monitoring and Evaluation of CERP: M&E Was Emphasized in Guidance but Rarely Sufficiently Performed

Guidance on the use of CERP funds was codified initially in DOD's 2005 Financial Management Regulation, which outlined which types of projects were prohibited (such as entertainment and reward programs) and which were allowed (such as agriculture and electricity projects).¹⁰⁰¹ The 2005 regulation divided responsibility for CERP program oversight among several DOD entities. While the Under Secretary of Defense (Comptroller) was responsible for broad program policies and procedures, as well as reports to the Congress, the Secretary of the Army was responsible for developing detailed procedures to ensure that commanders spent funds in accordance with applicable laws and regulations. The commander of U.S. Central Command (responsible for military operations in central Asia and the Middle East), meanwhile, was responsible for requesting and allocating funds across subordinate commands. The authority for selecting and implementing projects was delegated to commanders at the brigade level and below.¹⁰⁰²

Standard operating procedures for CERP evolved as funding for the program grew. In May 2009, U.S. Forces – Afghanistan published the first of numerous permutations of “Money as a Weapon System – Afghanistan,” which provided guidance on CERP processes. The initial version of the document made performance metrics a requirement for CERP projects over \$50,000. There were also higher approval thresholds for more expensive projects: for example, while CERP projects valued at \$200,000 or less could be implemented directly with the approval of brigade commanders, projects greater than \$2 million had to be approved by the commander of U.S. forces in Afghanistan. If the Afghan government was “not willing to fund operating costs or maintain the investment,” the handbook directed users to “not fund the project.” Initial guidance for commanders emphasized the importance of assessing impact and required quarterly “narratives” that explained how their CERP projects were advancing U.S. strategy. Almost immediately, CERP guidance began to tighten: The December 2009 update to “Money as a Weapon System – Afghanistan” required measures of effectiveness for all CERP projects, regardless of dollar value. Still, there was little direction on how to actually measure a project's impact after implementation. Instead, the handbook listed



Using CERP funding, Afghan men work on rebuilding market spaces in Helmand Province in late 2009. (U.S. Marine Corps photo)

required considerations to justify initiating a project, such as sustainability and the projected impact on the local population.¹⁰⁰³

The title of “Money as a Weapon System – Afghanistan” mirrored what had become a well-known counterinsurgency aphorism: that “some of the best weapons for counterinsurgents do not shoot.” This tenet was based on the belief, engrained in the counterinsurgency field manual, that “counterinsurgents often achieve the most meaningful success in garnering public support and legitimacy for [a host-nation] government with activities that do not involve killing insurgents.”¹⁰⁰⁴ As the field manual summarized, “money is ammunition.”¹⁰⁰⁵ Paralleling the basic assumptions of the field manual, aspects of “Money as a Weapon System – Afghanistan” seemed to reflect the proposition that positive impacts could be achieved merely through the act of implementation: According to the May 2009 version of the standard operating procedures, “CERP [provided] a quick and effective method to institute an immediate positive impact on the Afghan people.”¹⁰⁰⁶ The causal pathway through which counterinsurgency goals would be achieved was embedded in CERP guidance’s straightforward project selection criteria: “(1) execute quickly; (2) employ many Afghans; (3) benefit the Afghan people; and (4) be highly visible to the local populace.”¹⁰⁰⁷ According to this guidance, an ideal project would meet all four of these criteria in order to increase support for the Afghan government (and coalition forces) and decrease support for the Taliban.

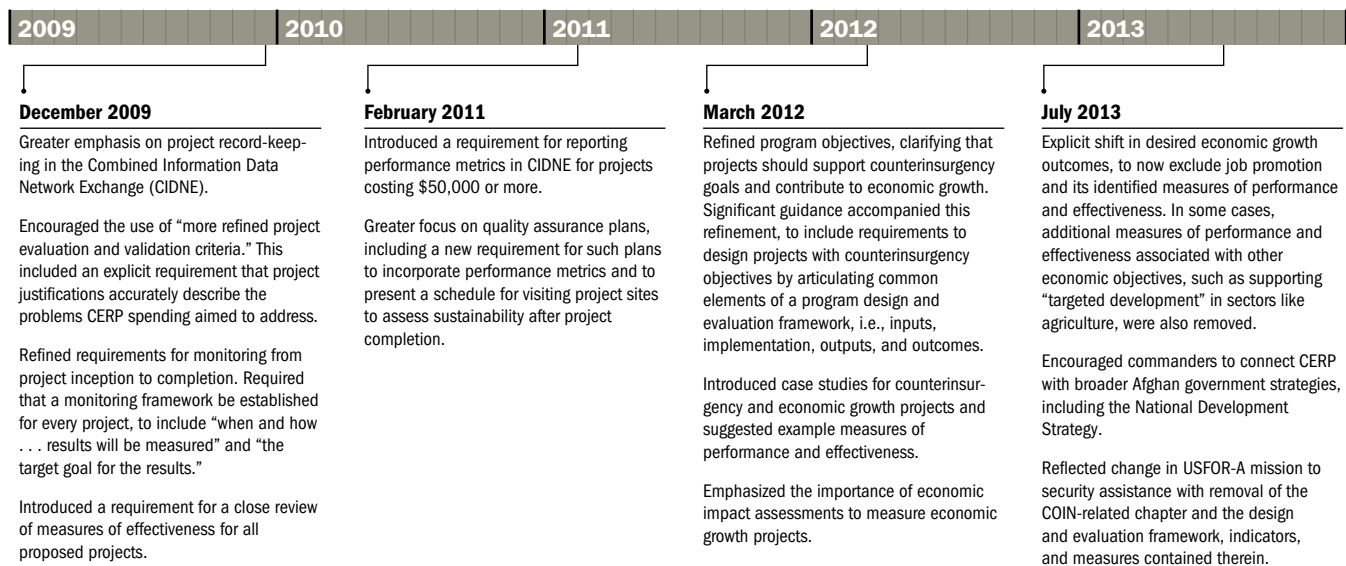
In 2009, in response to congressional scrutiny of CERP, the U.S. Central Command instructed forces in Afghanistan to use more refined project evaluation criteria for CERP in order to “preserve the program as a key non-kinetic [counterinsurgency] tool.” By August 2010, ISAF was developing program guidance that emphasized closer evaluation

of the effects of expenditures. Building on the metaphor that defined the program, the guidance emphasized that “CERP is a weapon and will be treated as one,” adding that personnel “[would] train and practice in its use.”¹⁰⁰⁸

CERP standard operating procedures continued to evolve in response to outside criticism (Figure 7 presents key M&E-related changes to those procedures over time). In November 2010, the U.S. Army Audit Agency concluded that “[U.S. Forces – Afghanistan] and its predecessors hadn’t established a correlation to prove whether CERP funding affected [counterinsurgency] operations. Although CERP had many humanitarian and developmental benefits, we questioned the effectiveness of CERP as a [counterinsurgency] tool due to the absence of a program baseline and performance measures.”¹⁰⁰⁹ Consequently, U.S. Army auditors called for the development of CERP baselines before completing projects to enable units to clearly compare conditions before and after a project was implemented.¹⁰¹⁰ Shortly thereafter, a February 2011 update to “Money as a Weapon System – Afghanistan” emphasized the need to evaluate the impacts of CERP projects on their area of operations.¹⁰¹¹ The March 2012 version of “Money as a Weapon System – Afghanistan” placed additional emphasis on collecting data and measuring performance and effectiveness, including changes that emphasized measuring CERP projects for counterinsurgency effects.¹⁰¹² As with other programs, the overall trend in CERP guidance was an increasing focus on monitoring and evaluating for outcomes and impact.

FIGURE 7

EVOLUTION OF MONEY AS A WEAPON SYSTEM – AFGHANISTAN: KEY CHANGES TO CERP M&E REQUIREMENTS AFTER MAY 2009*



* The initial standard operating procedures for CERP were released in May 2009. This table presents key, M&E-related changes to the standard operating procedures over time.

Source: U.S. Department of the Army, “Money as a Weapon System – Afghanistan (MAAWS-A),” U.S. Forces – Afghanistan Publication 1-06, December 2009, pp. 2, 18, 92–94, 114; U.S. Department of the Army, “Money as a Weapon System – Afghanistan (MAAWS-A),” U.S. Forces – Afghanistan Publication 1-06, February 2011, pp. 94, 96; U.S. Department of the Army, “Money as a Weapon System – Afghanistan (MAAWS-A),” U.S. Forces – Afghanistan Publication 1-06, March 2012, pp. 119–121, 123–124, 132–135; U.S. Department of the Army, “Money as a Weapon System – Afghanistan (MAAWS-A),” U.S. Forces – Afghanistan Publication 1-06, July 2013, pp. 3, 12, 124, 128–129.

Still, many metrics that were actually used for CERP were intended to prevent fraud and ensure compliance with contract terms instead of to understand its impacts on counterinsurgency goals.¹⁰¹³ Compliance was certainly a core aspect of CERP project management: CERP implementers interviewed for one RAND report described “onerous bureaucracy and paperwork that was required for project planning, approval, funding, and assessment.”¹⁰¹⁴ The implementers viewed such administrative requirements as posing opportunity costs, claiming that they “distracted [the implementers] from other responsibilities” and “hampered CERP’s ability to achieve tactical and strategic effects,” in the words of the RAND report.¹⁰¹⁵ But in another sense, it was the sheer difficulty of assessing impact that militated most aggressively against selecting and using meaningful indicators of success, leaving compliance itself as something of a default metric. Impact assessment was usually delegated far down chains of command to field-based personnel who faced substantial conceptual and practical barriers to assessing CERP’s effects. As one stabilization operations planner based in eastern Afghanistan told SIGAR:

I wanted to develop metrics for measuring impact for each project, but it was so hard to know what was causing security in any particular area to improve or worsen, and we could only get about 10 percent of the data we wanted to evaluate anything. So we decided it was too hard to focus on impact at the project level. The only metrics we ever developed were anchored in financial management: making sure people got paid and making sure things were built.¹⁰¹⁶

In other words, by default, measures of success were often conflated with sound implementation, rather than determined by actual results.¹⁰¹⁷ Metrics such as the number of times CERP implementers left their bases or the number of contacts they had with Afghan officials were not necessarily compliance-based, but were similarly linked directly to implementation.¹⁰¹⁸ In some cases, the metrics recommended by CERP guidance to assess a project’s effects were so synonymous with implementation that they equated paying out on a contract with effectiveness. For example, the March 2012 version of “Money as a Weapon System – Afghanistan” contained a case study of a food production and distribution project in which an indirect outcome of the project was calculated by applying a Keynesian spending multiplier (a multiplier used in economics to estimate the overall effect of government spending) of 1.5 to the value of a \$35,000 grant provided to the business. According to this formula, the result was a one-time boost to the local economy of \$52,500 (1.5 times \$35,000).¹⁰¹⁹ While this was not the only metric suggested by the case study (others included the number of new jobs created and increases in sales for the business that received the grant), it is still a classic example of the ways in which impact was sometimes assumed to follow naturally from implementation.¹⁰²⁰

In part, the view of CERP as a tool of fiscal stimulus emerging in later iterations of “Money as a Weapon System – Afghanistan” reflected concerns about the impending transition of authority to the Afghan government. As the March 2012 version of the standard operating procedures stated, updates to CERP procedures were not only intended as guidance on improving the program’s counterinsurgency effects but as a way to mitigate the impact of reduced U.S. spending, amid concerns that it might create “an economic downturn that [threatened] to reverse some of [the United States’]

hard-won achievements.”¹⁰²¹ Still, under the simplified logic of the Keynesian multiplier, any project at all could have been “effective” on some level merely by injecting cash into the economy.¹⁰²² Self-justifying metrics reflected the interaction of pressure to generate results on short timelines, an institutional drive within the U.S. military to “do something” even if conditions were not ideal, and the presumption that “hearts and minds” counterinsurgency theory was fundamentally valid.¹⁰²³

The result of this confluence of factors was that spending was often treated as a success unto itself. In the words of former ISAF commander General David Petraeus, “What drove spending was the need to solidify gains as quickly as we could, knowing that we had a tight drawdown time line. . . . And we wound up spending faster than we would have if we felt we had forces longer than we did.”¹⁰²⁴ Meanwhile, despite the lack of impact evaluations even during the height of CERP implementation, many military and civilian personnel at every level continued to insist that the program was effective in advancing counterinsurgency goals.¹⁰²⁵ Reflecting the general consensus, in a 2010 U.S. Senate hearing considering his nomination to be the commander of U.S. Central Command, General James Mattis described CERP as one of several programs “that allow us to checkmate our enemies.”¹⁰²⁶

CERP stands out as an example of what happens when impact measurement is demanded—and even, on paper, prioritized—but there is either little attempt, or insufficient capacity, to do it. Although CERP guidance explicitly encouraged impact assessment—“If we cannot determine how to measure the effectiveness of a project, then we should not do it,” the March 2012 edition of “Money as a Weapon System – Afghanistan” said—in practice, it rarely happened.¹⁰²⁷ The difficulty of actually coming up with a meaningful way to measure how much CERP projects were reducing support for the Taliban meant that impact, in the words of one former senior military official, “was always the last 10 yards that we couldn’t run.”¹⁰²⁸ One senior civil affairs officer told SIGAR that his division staff “would regularly tell COMISAF [Commander, ISAF], ‘CERP is a terrible development tool, but it’s a great stabilization tool,’ but we never knew if it was true.”¹⁰²⁹ Seventeen years and \$2.3 billion later, that question still cannot be definitively answered.¹⁰³⁰

After the Fact, There Were Numerous Attempts to Measure CERP’s Impact

An irony of CERP is that, while the program’s effectiveness was underexamined while it was in wide use in Afghanistan from 2008 to 2011, it came under exhaustive academic scrutiny afterwards. In a 2017 systematic review of 19 academic studies on the relationship between development aid and violence, six focused on assessing the impact of CERP.¹⁰³¹

The program was conducive to quantitative study because the U.S. military had generated a substantial amount of data that could be used to evaluate the program’s effects—including “significant activity” data from the U.S. military documenting insurgent attacks and location-specific reconstruction spending data.¹⁰³² The studies generally used statistical analysis to assess the effects of CERP spending on levels of violence. Violence was treated as a proxy for security conditions and, more broadly,



A U.S. Army officer speaks with contractors about the progress of a school construction project funded through CERP in Balkh Province in 2012. (DOD photo)

for counterinsurgency effects; lower levels of violence would, in theory, reflect successful counterinsurgency interventions.¹⁰³³

On the whole, the research returns inconclusive and sometimes conflicting results, including at least two studies that could not identify any effects.¹⁰³⁴ Still, some of the conclusions that emerged from the research point to *specific circumstances* in which CERP *might* have had some positive impact. These include the possibility that smaller CERP projects valued at less than \$50,000 may more effectively reduce violence than larger ones—though only when security force presence was high.¹⁰³⁵

The academic literature’s verdict on CERP is, on balance, consistent with a broader body of research and the 2017 systematic review’s conclusion that “the evidence for a violence-dampening effect of aid in conflict zones is not strong.”¹⁰³⁶ In CERP’s case, estimated positive effects are qualified: Positive impacts may have been possible only through the interaction of troop strength with development spending.¹⁰³⁷

“Investing in the Fight”: DOD Funds Its Own Impact Assessment of CERP

In 2012, DOD sponsored its own impact assessment of CERP in Afghanistan.¹⁰³⁸ Carried out by the RAND Corporation, the study (titled “Investing in the Fight”) sought to add to the existing literature on CERP by using both qualitative and quantitative methods to explore the program’s effects on tactical operations in Afghanistan from 2010 through 2013 (a period that overlapped with both the U.S. troop surge and implementation of a modified counterinsurgency strategy).¹⁰³⁹

To understand the experience of CERP implementers—including how they chose to use the program to achieve tactical objectives—RAND researchers conducted structured interviews with 197 military officers and noncommissioned officers in conventional Army and Marine Corps units, as well as in Special Operations Force units (which include elite elements such as the Navy Seals that engaged in village-level stability operations).¹⁰⁴⁰ Next,

with geospatial data identifying where CERP projects were implemented, the researchers used quantitative analysis to try to understand CERP's effect on a variety of indicators encompassing both Afghan "population- and coalition-focused outcomes."¹⁰⁴¹

Population-focused outcomes included measures of population movements, agricultural activity, and economic activity.¹⁰⁴² (In this context, population movement refers to migration within Afghanistan's borders, and was selected as an outcome indicator on the assumption that improved security and economic conditions cause internal migration).¹⁰⁴³ Coalition-focused outcomes included the volume of intelligence about enemy activity, coalition-force freedom of movement (that is, the ease with which coalition forces could conduct vehicular patrols in their areas of operation), and enemy attacks on coalition forces, captured via "significant activity" (SIGACT) data.¹⁰⁴⁴ The

CERP Impact Assessments Had to Grapple with Data Quality Issues

Well-documented data quality issues were a significant challenge for researchers attempting to determine CERP's impact. CERP data in Afghanistan were often spotty, with gaps and duplicate entries.¹⁰⁴⁵ In 2009, SIGAR reported that more than half of CERP project files it reviewed from the 2007–2009 period were incomplete.¹⁰⁴⁶ According to DOD, the quality of CERP data improved over time.¹⁰⁴⁷

Measures of so-called "significant activity" data—referred to as SIGACT data—used by researchers to measure levels of violence also suffer from some quality problems.¹⁰⁴⁸ For example, from conversations with former brigade and battalion staff officers, several researchers hypothesized that the "proportion of true incidents recorded as SIGACTs drops as the intensity of violence rises." That is, in areas where violence is high, units may under-report the number of incidents because violence is so regular that some incidents go unrecorded.¹⁰⁴⁹ As one researcher observed, "if it is difficult to conclusively disprove the accuracy of the SIGACT data, it is also difficult to prove their accuracy."¹⁰⁵⁰

To mitigate the potential for spurious conclusions based on bad data, researchers used a variety of methods, including removing individual records with missing data or questionable coding, excluding periods of time or districts from their analysis, or otherwise cleaning data to ensure only unique records remained.¹⁰⁵¹ These approaches may have addressed, at least to some extent, discrepancies and omissions in SIGACT reporting and CERP project documentation practices.¹⁰⁵²

In some cases, the presence of potential measurement errors—specifically in SIGACT data—that were not possible for researchers to correct may have strengthened results even as they biased them. For instance, undercounting errors, if present and consistent as hypothesized, would have introduced a conservative bias.¹⁰⁵³ As an example, several researchers found that an additional dollar of CERP spending in Iraq over a half-year period from 2007–2008 resulted in 1.59 fewer violent incidents per 100,000 residents.¹⁰⁵⁴ Were a conservative bias in fact present, and the analysis accurate, the unbiased estimate of the reduction in violent incidents would have been above 1.59 per 100,000 residents. In other words, in this case, a conservative bias would have strengthened, not weakened, the conclusion that CERP reduced violence. (The researchers were cautious to note that the relationship between CERP and violence is not necessarily causal.)¹⁰⁵⁵

The presence of unmeasured and unknown variables makes it difficult to determine the true impact of CERP projects.¹⁰⁵⁶ More consequentially, it is not possible to know fully whether underlying data-quality issues were adequately addressed by researchers—or whether they could have been adequately addressed at all. This adds another layer of uncertainty to CERP impact assessment and underscores the importance of sound data.

study compared areas where CERP projects were implemented with similar control areas, identified using statistical matching data.¹⁰⁵⁷

Some results confirmed previous findings on specific circumstances in which CERP *may* have been effective; others raised questions about, or otherwise supplemented, the literature on CERP. The study included a wider array of dependent variables (other than SIGACTs or other indicators of violence) in its quantitative analysis, which provided a way to measure broader potential counterinsurgency outcomes besides violence.¹⁰⁵⁸

Two conclusions generally supported the findings of other research:¹⁰⁵⁹

- Small (costing less than \$5,000) and medium-sized projects (costing \$5,000 to \$50,000) increased short-term intelligence collection and coalition freedom of movement.
- Small- and medium-sized projects increased enemy engagements in the year of implementation, but were associated with reductions in enemy attacks in the following year.¹⁰⁶⁰

The RAND study added to the previous literature on CERP by finding that CERP projects were associated with long-term increases in local economic activity and immigration.¹⁰⁶¹ At the same time, the RAND study raised questions about whether quantitative analysis could draw meaningful conclusions about the impact of CERP at all. CERP activity, it said, appeared to “be acting as a proxy for contemporaneous counterinsurgency activities”—so instead of capturing independent CERP activity, statistical analysis “may instead be capturing overall counterinsurgency behavior.”¹⁰⁶² The study concluded that “quantitative analyses cannot credibly identify the impact of CERP independently of overall military efforts.”¹⁰⁶³

The qualitative aspects of RAND’s assessment, however, were informative. “Softer” outcomes, such as building local rapport and enhancing local governance, “were more important to implementers than completing projects,” the study said, suggesting that output measures like construction milestones could not fully capture CERP’s perceived effects.¹⁰⁶⁴ Still, the impact assessment could not completely bridge the gap between outcomes that were important to CERP implementers and more observable outcome indicators, such as SIGACTs and economic activity: The analysis “found no consistent evidence that perceived CERP project outcomes [as communicated by interviewees] are correlated with measurable outcomes.”¹⁰⁶⁵

All of this underscored the extraordinary difficulty of assessing impact even for what was arguably the most-studied reconstruction program in Afghanistan. It also highlighted lingering questions about the “complex causal pathway” between CERP inputs, such as the type of project or the amount of dollars spent, and counterinsurgency outcomes.¹⁰⁶⁶ Finally, the study’s design (which focused on tactical operations) left unresolved the question of whether CERP projects contributed to strategic U.S. goals.¹⁰⁶⁷ Yet despite its limitations, the DOD-funded impact assessment of CERP was generally useful. If nothing else, it communicated the perspectives of nearly 200 (albeit not

randomly selected) CERP implementers and suggested dependent variables other than violence that might be used to assess outcomes.¹⁰⁶⁸ Table 8 explains some of the key strengths and weaknesses of RAND’s analysis.

TABLE 8

STRENGTHS AND WEAKNESSES OF DOD’S IMPACT EVALUATION OF CERP
Strengths
The RAND evaluation layered a significant qualitative component onto its quantitative analysis.
Although other impact evaluations of CERP generally used violence as the dependent variable (for instance, insurgent attacks against coalition forces), RAND’s quantitative analysis evaluated CERP’s effects on other dependent variables. Data included proxies for both additional coalition-focused outcomes, such as freedom of movement, and proxies for Afghan-population-focused outcomes, such as population movements and agricultural activity.
In addition to a literature review, the evaluation described the perspectives on CERP of the Congress, senior military leaders, senior civilian leaders, and the oversight and policy communities.
The evaluation included structured interviews of 197 CERP implementers and project designers. These interviews provided additional insights on the kinds of outcomes implementers hoped to achieve and generated qualitative data on 407 individual projects. Interviewees were asked to describe up to three projects: a success, a failure, and an “interesting” project. Success and failure were defined by the interviewees to allow for open-ended questioning that captured “a diverse range of experiences with CERP.”
To some extent, the evaluation explored the strategic basis for CERP through interviews with senior military and civilian leaders.
Weaknesses
Quantitative data functioned as proxies for outcomes of interest rather than the outcomes themselves. For example, the study used nighttime ambient light data to estimate changes in economic activity. According to the study, the use of nighttime ambient light data “requires assumptions about the relationship between nighttime light production and economic activity; in rural areas of Afghanistan, this relationship is likely to be tenuous.”
Quantitative analysis was limited to a subset of CERP projects that may not have been fully representative. Specifically, the dataset was limited to the 2010–2013 period and was further restricted to projects for which precise location data could be identified (constituting 77% of CERP projects and 55% of CERP obligations for the 2010–2013 period). There are further limitations that may limit the quantitative findings. For example, project timeframes are marked only to the fiscal year and locations are marked to single grid-reference points, making more diffuse impacts difficult to measure.
Quantitative analysis was insufficient to establish a clear cause-and-effect relationship between CERP projects and outcomes. That is, other variables could have accounted for observed effects. As the study said, “quantitative analyses cannot credibly identify the impact of CERP independently of overall military efforts.”
While the number of interviewees was high, the sample of interviewees was nonrandom, which limited the applicability of the study’s findings.
Ultimately, the evaluation left open the question of whether CERP contributed to strategic goals, recommending that DOD conduct a separate assessment to answer it.

Source: 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, pp. xi–xii, xiv, 7, 12–13, 45–49, 94–95, 202, 213.

The Afghanistan Infrastructure Fund Emerges from CERP

CERP was not the only DOD program that linked economic development with counterinsurgency objectives. As the reconstruction effort continued, new funding authorities were required to finance more ambitious projects. In 2010, the U.S. Senate Appropriations Committee began expressing concern that CERP was becoming “a means for senior leaders to circumvent proper budgeting for large-scale developmental programs” and was being used to fund DOD efforts that were outside the program’s original scope.¹⁰⁶⁹ These concerns led the committee to place a \$20 million cap on CERP projects.¹⁰⁷⁰ In response, DOD and State jointly requested that the Congress establish the Afghanistan Infrastructure Fund in order to finance the kinds of larger projects that were no longer allowed under CERP.¹⁰⁷¹

The Afghan Infrastructure Fund was explicitly linked to the U.S. counterinsurgency effort (a discussion at the end of this section examines the question of whether AIF should also be assessed against economic development outcomes, independent of the counterinsurgency effects it aimed to achieve).¹⁰⁷² Unlike CERP, however, AIF projects were to be formulated and agreed to by both DOD and State.¹⁰⁷³ AIF was created to fund power, water, transportation and other infrastructure projects, but in practice approximately 82 percent of AIF funds ended up financing power-sector projects.¹⁰⁷⁴ The Congress appropriated all AIF funds to DOD; 25 percent of those funds were eventually transferred to USAID as the joint effort evolved.¹⁰⁷⁵ DOD's AIF projects were implemented through military contracts managed by the U.S. Army Corps of Engineers.¹⁰⁷⁶

In both their scale and intent, AIF projects were ambitious. For example, DOD invested \$456.7 million to rehabilitate and expand Afghanistan's North East Power System and its counterpart, the South East Power System. These projects generally involved the construction of electrical transmission lines and substations.¹⁰⁷⁷ By increasing the government's ability to provide essential services and stimulate long-term economic growth with improved access to electricity, the logic went, the projects would increase the population's confidence in the Afghan government and reduce support for the Taliban.¹⁰⁷⁸

M&E of AIF: Differing Views on When Outcomes and Impact Could Be Assessed

However, DOD's monitoring and evaluation of AIF projects did not capture whether this goal was met. While the U.S. Army Corps of Engineers had a monitoring system that ensured that construction was completed to standard and hired qualified local personnel to monitor its construction projects, DOD (along with State and USAID, which also received AIF funds) did not determine whether AIF projects achieved their counterinsurgency objectives.¹⁰⁷⁹ In addition, as SIGAR has reported elsewhere, many AIF projects experienced significant delays.¹⁰⁸⁰ In 2017, for example, SIGAR reported that three fiscal year 2011 power-sector projects were nearly five years behind schedule.¹⁰⁸¹

DOD's AIF projects were eventually completed. (In contrast, three major, AIF-funded power-infrastructure projects implemented by USAID remain unfinished, including one that was nearly seven and a half years behind schedule as of April 2021.)¹⁰⁸² But delays raised questions about the timeline on which the projects would be able to achieve counterinsurgency effects.¹⁰⁸³ As early as 2012, it was clear that the scale of most projects, in addition to acquisition and funding delays, meant that the full contribution of AIF to the U.S. counterinsurgency strategy might not be realized for years.¹⁰⁸⁴ Because many AIF projects are interdependent, many Afghans may not experience the benefits of increased access to electricity until other portions of the power grid are complete. (For example, USAID is responsible for constructing a 470-kilometer transmission line that will eventually link the country's North East Power System with its South East Power System, allowing electricity imported from Central Asia to reach southern Afghanistan.)¹⁰⁸⁵

Equally worrying, though, was the lack of performance metrics needed to assess whether AIF was contributing to strategic counterinsurgency goals (or at least to intermediate outputs or outcomes along the way to those goals). In its 2017 audit of fiscal year 2011 AIF projects, SIGAR reported that no agency involved in the projects—

including DOD—developed the metrics necessary to determine whether such higher-order objectives were being achieved. Such metrics were not developed even though the 2011 “U.S. Integrated Civilian-Military Campaign Plan” (the guiding strategic document at the time AIF projects were first initiated) specifically cited indicators that could have been used, such as the number of districts in which the Afghan government reported delivering essential services like electricity and the percentage of Afghans in a given area who approved of the government. In the course of SIGAR’s 2017 audit, DOD, State, and USAID stated that they were not required to assess whether the projects were achieving their counterinsurgency objectives.¹⁰⁸⁶ However, the absence of measurement made it difficult to determine whether the projects were in fact achieving effects.

The question of how, and under what time frame, AIF was supposed to contribute to counterinsurgency goals was another key issue—one that had direct implications for measurement. In response to scrutiny from SIGAR, DOD initially took the view that SIGAR had misunderstood counterinsurgency doctrine and that some counterinsurgency effects would immediately accrue:

[SIGAR’s July 2012 audit report examining FY 2011 AIF projects] incorrectly assumes that counterinsurgency (COIN) benefits begin after a project is completed. There are immediate COIN benefits when the local population learns of a project, participates in project planning and approval, and sees construction begin. Progress, even in the early stages of a project, sustains positive expectations and encourages local commitment to future progress. Additionally, AIF projects will employ local civilians and contractors, adding to the immediate COIN benefits by generating economic activity.¹⁰⁸⁷

Responding to SIGAR’s 2012 audit, DOD did not concur with a recommendation to clearly indicate the amount of time that infrastructure projects will take to achieve counterinsurgency benefits.¹⁰⁸⁸ But later, responding to SIGAR’s 2017 audit, DOD emphasized that impact could only be measured after projects were complete: “Although projects *can have* positive benefits prior to completion, such as the hiring of local workers, the primary impact of increasing electrical capacity cannot be assessed until the projects are complete and electricity is available” [emphasis added].¹⁰⁸⁹

In another exchange, the logic of counterinsurgency was offered as a reason why effects did not need to be assessed.¹⁰⁹⁰ In May 2015, U.S. Forces-Afghanistan officials told SIGAR that determining whether a project achieves counterinsurgency objectives was an “intuitive process.”¹⁰⁹¹ They went on to describe the example of the (partially) AIF-funded Kandahar Bridging Solution, which provided fuel and technical support for diesel power plants in Kandahar City as a “bridge” until other power sources were available.¹⁰⁹² The officials said that, in this case, counterinsurgency benefits would naturally accrue because the Bridging Solution provides reliable power, which leads to the development of infrastructure, which stabilizes the local economy, thereby contributing to the improvement of Afghanistan.¹⁰⁹³

These explanations appeared to merely rehash the original logic underlying the Bridging Solution and seemed to reflect a belief that positive counterinsurgency effects were inevitable. In a memorandum endorsing funding for the project, then-CENTCOM

Commander General David Petraeus stated, “Sufficient electrical supply is a key component in building the population’s faith in their government, increasing the average Afghan’s economic prosperity, and dissuading the population from joining the Taliban. Approving this project will serve to win the hearts and minds of the Afghan people.”¹⁰⁹⁴ In other words, the “intuitive process” through which counterinsurgency effects could be measured and the project’s underlying logic were essentially one and the same. Such circular logic argued against measuring results, and failed to take other possibilities into account—for instance, a scenario in which increased access to electricity does not actually persuade Afghans to support a government riddled by corruption.¹⁰⁹⁵

Stress-Testing Aspects of AIF’s Claimed Intermediate COIN Impact

The causal chains that could conceivably produce positive counterinsurgency outcomes, though plausible, were highly complex. There was an intuitive appeal to the theory that providing jobs, essential services like electricity, and development aid more broadly could win over disaffected groups in “a war among the people.”¹⁰⁹⁶ But efforts that scrutinized some of the causal mechanisms of that theory painted a more mixed picture.¹⁰⁹⁷ One example of that was an examination of the assumption that creating employment opportunities unambiguously results in counterinsurgency benefits.

In a study published in 2011 (“Do Working Men Rebel?”), researchers from Stanford, Princeton, and the University of California San Diego tested the theory that gainfully employed young men were less likely to join insurgencies. The study, which examined the relationship between employment and levels of violence in Iraq, Afghanistan, and the Philippines, “emphatically [rejected] a positive correlation between unemployment and attacks against government and allied forces” and found “no significant relationship between unemployment and the rate of insurgent attacks that kill civilians.”¹⁰⁹⁸ The authors offered possible explanations for this conclusion that were consistent with counterinsurgency practice—for example, the hypothesis that in areas of higher unemployment, it may be cheaper for counterinsurgents to “buy” intelligence from locals that can then be used to combat insurgents.¹⁰⁹⁹ However, one key implication of the study was that “aid and development efforts that seek to enhance political stability through short-term job creation programs may well be misguided.”¹¹⁰⁰

A second study carried out by Mercy Corps and similarly framed as a question (“Does Youth Employment Build Stability?”) sought to “test the theories of change that link unemployment, poverty and economic deprivation to support for political violence, terrorism and insurgency.”¹¹⁰¹ Published in 2015, the study examined the impact on stability (as measured by surveys that sought to gauge respondents’ propensities towards political violence) of a Mercy Corps program that aimed to increase youth employment in Helmand Province through vocational and technical training.¹¹⁰² The study concluded that, while the program “was highly successful in impacting economic outcomes amongst participating youth in one of the most violent provinces of Afghanistan,” it had “nearly zero impact on political outcomes hypothesized to be linked to youth propensity towards political violence.”¹¹⁰³

THE TASK FORCE FOR BUSINESS AND STABILITY OPERATIONS: SOME IMPACT ASSESSMENTS DID NOT ACTUALLY EXAMINE IMPACT

The Commander's Emergency Response Program and the Afghanistan Infrastructure Fund were not the only programs that sought to achieve counterinsurgency effects through economic development. In 2011, the Congress authorized DOD's Task Force for Business and Stability Operations to fund economic development projects aimed at reducing violence, enhancing stability, and promoting economic normalcy in Afghanistan (the task force began operating in Afghanistan in 2010).¹¹⁰⁴ By the time its work had concluded, the task force had obligated at least \$675 million in support of those objectives.¹¹⁰⁵ Unlike other economic development entities, the task force was led by private sector experts rather than development professionals.¹¹⁰⁶ A former task force official described it as a "bridge to the private sector" or a "private equity firm" that promoted investment in key sectors like cashmere and natural gas.¹¹⁰⁷

Results were mixed.¹¹⁰⁸ Communication between task-force contracting officer's representatives and external contracting officers was not always strong, which inhibited the ability of the latter to solve problems when issues with contracts arose.¹¹⁰⁹ Moreover, lack of coordination with other U.S. agencies, such as USAID, led to conflicting or repetitive project implementation. For example, a \$39.6 million natural gas pipeline project that USAID had previously recommended against was ultimately executed by the task force without USAID's knowledge.¹¹¹⁰ One senior task force official stated that the task force viewed coordination with U.S. embassy officials as a "courtesy."¹¹¹¹ Lack of coordination consequently increased the risk that task force projects would not be sustained.¹¹¹² Neither USAID nor State expressed interest in taking over any of the task force's incomplete projects, citing their leadership's disinterest in doing so. In their opinion, the projects were considered to be liabilities because of safety concerns, lack of sustainable design, and other problems.¹¹¹³

Oversight and coordination were not the only challenges: From the outset, the task force lacked a standardized M&E approach.¹¹¹⁴ Its intended mission, objectives, and strategy were not clearly articulated until more than two and a half years after it began operating in Afghanistan.¹¹¹⁵ Although the task force did eventually articulate objectives such as creating jobs and increasing exports, poor record keeping and a lack of reliable data made it difficult to determine whether those objectives were achieved.¹¹¹⁶

Particularly in its early years, the task force relied on impressionistic metrics of effectiveness. Those metrics included "symbolic big-name deals" and instances where the task force got "things done that nobody else could," in the words of one State official.¹¹¹⁷ Although such metrics were useful for assessing intermediate progress, they did not measure longer-term effects. That analysis was left to an economic impact assessment commissioned by the



A restaurant supported by the Task Force for Business Stability Operations and the Assistance in Building Afghanistan by Developing Enterprises program in Herat, 2016. (SIGAR photo)

Task Force for Business Stability Operations, which was completed as the task force was closing down.¹¹¹⁸ The assessment was unusual in that it sought to quantify *potential future* impact instead of achievements to date.¹¹¹⁹ Although the assessment projected that the task force's projects would eventually make substantial contributions to Afghanistan's gross domestic product, some of its core assumptions proved to be overoptimistic.¹¹²⁰ For example, the task-force's upstream mineral development projects were projected to contribute \$8.1 billion to GDP in 2025.¹¹²¹ That figure was based on the assumption that exploration of task force-supported mining concessions would begin in 2016 and 2017. However, the Afghan government did not provide final approval for three large-scale contracts advanced by the task force until the fall of 2018—and two of the three contracts were canceled by the government a year later.¹¹²²

The impact assessment's overly optimistic economic assumptions were not unique.¹¹²³ A bigger problem with the assessment was that it did not examine whether the task force accomplished its core mission: "to promote stability in Afghanistan."¹¹²⁴ Economic outcomes may have been important, but they were also one causal step removed from that final goal. In the end, the \$3.4 million impact assessment never addressed the basic question: whether the task force accomplished what it originally set out to do.¹¹²⁵

Still, the study left open a small door for counterinsurgency theory. Although, in aggregate, the economic outcomes to which the Mercy Corps program contributed suggested that “young people’s current economic circumstances are not a major driver of propensity towards political violence within the context of southern Afghanistan,” one particular outcome the program affected—respondents’ positive perceptions of their future economic prospects—“was found to be significantly related to a lesser acceptance of the use of violence.”¹¹²⁶ As with the previously described findings specific to CERP, it seemed that economic development initiatives could potentially contribute to stability, but not in unqualified ways and only along certain pathways. Such studies underscored the importance of thoroughly assessing all claims of impact and the criticality of using performance measurement to mitigate the risk of doing the wrong thing perfectly.

The Afghanistan Infrastructure Fund Underscores Key Questions Surrounding Measurement

SIGAR’s experience in overseeing the Afghanistan Infrastructure Fund highlights two critical questions for M&E: What effects should be measured, and when?

Answering these questions requires a clear understanding of a program or project’s purpose—what it is ultimately supposed to achieve. In AIF’s case, the language that various stakeholders used to describe AIF’s intended effects was not always consistent. Congress authorized the AIF “with the expectation that these funds will be used for critical large-scale infrastructure projects . . . that are critical to counterinsurgency objectives.”¹¹²⁷ A 2017 SIGAR audit criticized DOD, State, and USAID for not assessing whether six AIF projects initiated in fiscal year 2011 had achieved those objectives.¹¹²⁸ It remains SIGAR’s view that AIF-funded projects should have been measured against the counterinsurgency effects they were intended to produce.

But, responding to a draft of that audit, DOD pointed to the language of the November 2010 correspondence from the Secretaries of State and Defense requesting that the Congress create the AIF.¹¹²⁹ The request stated that AIF funds would be used to support “projects critical to counterinsurgency objectives and *economic development*” [emphasis added].¹¹³⁰ One interpretation of the language of the request was that AIF projects were intended to achieve economic development objectives *in addition to* counterinsurgency objectives and, by implication, that economic development objectives were *distinct* from counterinsurgency objectives. However, DOD’s budget requests for fiscal years 2012, 2013, and 2014 stated that AIF funding would support counterinsurgency objectives.¹¹³¹

The debate about whether AIF was intended to achieve both counterinsurgency and economic development objectives, or ultimately only counterinsurgency objectives, was further complicated by the Russian doll of counterinsurgency theory. For example, AIF-funded power-infrastructure projects implemented in southern Afghanistan were intended to promote economic development in Kandahar and Helmand Provinces. Promoting economic development by increasing the Afghan government’s ability to provide basic services, DOD theorized, would help counter the Taliban insurgency.¹¹³² Thus, even if the objectives were ultimately counterinsurgency-related, they were to be achieved *through* economic development. If development was a means to an end, metrics of development

could be used to assess whether counterinsurgency campaigns were achieving success: The December 2006 version of the U.S. Army's counterinsurgency field manual listed "availability of electricity" as a broad indicator of progress.¹¹³³

Reflecting the tension between the concepts of economic development in service of counterinsurgency and economic development for its own sake, the congressional record shows that different stakeholders articulated the purpose of AIF in different ways. Testifying before the Senate Armed Services Committee in March 2011, General James Mattis, then-commander of U.S. Central Command, stated, "The new Afghanistan Infrastructure Program enables us . . . to undertake high-priority infrastructure projects to address critical needs for Afghan security, governance, and development. The Afghanistan Infrastructure Fund will be the vehicle for the Defense Department's contribution to this integrated program."¹¹³⁴ Fourteen days later, in testimony before the same committee, then-International Security Assistance Force Commander General David Petraeus said that the "larger projects" made possible by the AIF were "central to the conduct of a counterinsurgency campaign," and added, "these are not economic development, they're not economic assistance or something, these are projects that directly enable the success of our troopers on the ground."¹¹³⁵

The distinction between counterinsurgency and economic development carried implications for the timing of when effects could be expected. Responding to SIGAR's 2017 audit, DOD took the view that the impact of three AIF-funded power-infrastructure projects could not be measured until the projects were complete.¹¹³⁶ SIGAR agrees that, if power-infrastructure projects are intended *primarily* to achieve long-term development effects, and if projects are not significantly delayed or at risk of not being sustained upon completion, it may be more appropriate to assess impact until after the projects are fully operational, as electricity cannot flow through unconstructed transmission and distribution lines.¹¹³⁷ According to the World Bank, lack of access to electricity constitutes a crucial barrier to progress on a wide range of development-related goals, including poverty reduction, education, health, livelihoods, and food security.¹¹³⁸ Assuming the Bank is correct, those goals would not be meaningfully advanced by power-infrastructure projects before access to electricity actually improves.

However, if the purpose of projects is to achieve counterinsurgency objectives, counterinsurgency theory suggests that effects can be measured much sooner—during implementation or even before construction begins. In 2015, DOD stated that "Afghan officials used the . . . initiation of [AIF] projects, including ribbon-cutting ceremonies, the hiring of local workers, and initial construction, to demonstrate that [the Afghan government] can and will provide essential services that insurgents cannot."¹¹³⁹ When effects are expected sooner, instruments such as public perception surveys can be administered earlier in project timelines in order to ascertain whether counterinsurgency effects are being achieved. As DOD has told SIGAR, where success depends on the logic of counterinsurgency theory, effects are not fully contingent on project completion.¹¹⁴⁰ An important lesson of AIF is that adequate M&E depends in many ways on the clear and explicit articulation of what a program or project intends to accomplish—and when.

In the end, the Afghanistan Infrastructure Fund's contribution to counterinsurgency objectives, though somewhat plausible, has not been confirmed. However, there is still some hope: In response to SIGAR's 2017 AIF audit, USAID stated that it planned to conduct a "focused survey about access to electricity, as well as economic and social metrics" in order to better understand the impact of specific power-sector projects undertaken by USAID, DOD and State.¹¹⁴¹ According to USAID, the agency completed a baseline survey that would be part of a quasi-experimental impact evaluation measuring the impact of changes in electricity across seven dimensions, including the economy, health, education, and security.¹¹⁴² USAID plans to initiate a midline survey in August 2021.¹¹⁴³ The completion of the baseline survey, as well as the planned midline survey, are encouraging. Eventually, the impact assessment could yield important information about the effects of electricity access on a variety of outcome indicators. However, the reality remains that agencies made little effort to assess intermediate counterinsurgency effects. Moreover, the extent to which USAID's quasi-experimental evaluation will thoroughly examine longer-term counterinsurgency impacts is unclear.¹¹⁴⁴ Finally, whether deteriorating security will affect future data collection remains to be seen.¹¹⁴⁵

OPERATION ASSESSMENT: DOD'S OVERARCHING M&E ANALOGUE

In military doctrine, operation assessment is generally defined as process by which a commander and his or her staff determine what actions and resources are needed to achieve a goal, and measure progress toward it (Appendix B provides further discussion of assessment in doctrine).¹¹⁴⁶ Operation assessment is meant to be a continuous process of identifying, adjusting, and refining the means of accomplishing mission objectives and is therefore analogous to M&E.¹¹⁴⁷ However, it differs from M&E as practiced by other agencies due to its focus on measuring the holistic progress of military campaigns—a collection of linked operations, programs, and projects—towards their objective, rather than focusing on a specific program or project.

Assessment as practiced by DOD is also more flexible than M&E as practiced by State and USAID. Since commanders are ultimately responsible for the success of their command's activities, the methodology and timing of assessments are tailored to each commander's goals and decision-making style.¹¹⁴⁸ Not surprisingly, assessment processes and styles vary widely. As a 2012 NATO study noted, some assessment paradigms took a quantitative approach, whereas others adopted an approach based on commander's estimates and the judgment of subject matter experts. Whatever form the assessments took, however, the emphasis placed on the practice has consistently increased through DOD's reconstruction experience.¹¹⁴⁹ When it came to "winning hearts and minds" through counterinsurgency, operation assessment was put to use to monitor and evaluate the aggregate effects of both the military's reconstruction and warfighting activities.

The most prominent examples of operation assessment in Afghanistan were carried out by generals. It was common for commanders to conduct formal assessments when they assumed command or during key points in their tenure. For example, General Stanley McChrystal carried out an initial assessment that he delivered to then-Secretary of

Defense Robert Gates shortly after he assumed command of ISAF.¹¹⁵⁰ His successor, General David Petraeus (who took command of ISAF on July 4, 2010) released an “assessment of the situation in Afghanistan as we enter 2011” on January 25, 2011 in order to provide context for coalition military operations in the new year.¹¹⁵¹ But assessments are also conducted within subordinate commands. For example, the commander of an infantry company consisting of 160 soldiers or marines would assess against tactical goals within the company’s relatively limited area of operations.

Vague Guidance, Overoptimism, and the Complexity of Counterinsurgency Limited the Utility of Assessment in Afghanistan

Counterinsurgency operations in Afghanistan and Iraq prompted significant debate among practitioners and researchers about the utility of assessment, particularly in complex contingency environments.¹¹⁵² Because one of the basic tenets of DOD operation assessment doctrine is “commander centrality”—General McChrystal’s initial assessment, for example, was informed in part by his “core beliefs”—changes to campaign plans as commanders cycled in and out of Afghanistan shifted the focus, process, and metrics of assessment.¹¹⁵³ The frequency of these shifts and the inherent difficulties of measuring progress in a complicated environment like Afghanistan raised inevitable questions about what to measure: As one RAND report noted, “progress in [counterinsurgency] tends to be incremental and vague rather than dramatic and obvious.”¹¹⁵⁴ Others levied more blunt

The Afghan First Initiative: Sometimes Metrics Were Straightforward but Difficult to Reliably Track

The Afghan First Initiative was an economic program designed to support U.S. counterinsurgency objectives by creating jobs for Afghans.¹¹⁵⁵ As codified in the 2008 National Defense Authorization Act, the initiative limited bidding on certain contracts to Afghan firms. Two years later, the Afghan First Initiative was expanded when the commander of U.S. Forces in Afghanistan and the U.S. ambassador to Afghanistan issued guidance to ISAF and American civilian agencies urging the award of more contracts directly to Afghan companies and encouraging adherence to overall counterinsurgency goals when contracting on behalf of the U.S. government.¹¹⁵⁶

However, tracking the initiative’s progress toward actually creating jobs proved difficult. While DOD collected employment data from the initiative’s contract awards, program guidance lacked a standard definition of “employment” and did not require systematic collection of employment data to compare with the initiative’s goals.¹¹⁵⁷ In an attempt to better measure the impact of programs such as the Afghan First Initiative, ISAF developed a scorecard in September 2010 which reported data

on several relevant areas, including Afghan employment figures. However, its utility was limited by ISAF’s inability to verify the data.¹¹⁵⁸

In a 2012 audit, SIGAR determined that the lack of a standard data collection methodology and an inability to verify data made it difficult for DOD to effectively evaluate the Afghan First Initiative’s impact on employment.¹¹⁵⁹ The inability to measure the program’s effects ultimately limited the decision-making power of the program’s M&E and prevented clear determination of whether the program could have been improved, or if the concept could be useful in future contingency environments. However, oversight of the initiative was reasonably effective, in that most Afghan companies met contract requirements.¹¹⁶⁰

Left unaddressed by M&E was the fundamental assumption behind the Afghan First Initiative: that creating jobs affected counterinsurgency outcomes (see pp. 139 and 142). Even if complete and reliable employment data had been collected, the program’s impact still may have been unclear.



General Stanley McChrystal listens to a briefing before a patrol in Marjah, Helmand Province in 2010. (U.S. Navy photo)

critiques. While acknowledging that operation assessment “should be a vital part of any unconventional conflict, especially counterinsurgency,” one researcher simultaneously asserted that operation assessments, at least as then practiced, “fail.”¹¹⁶¹

Two articles that appeared side by side in the autumn 2011 edition of the *Naval War College Review* summarized much of this criticism. Among the issues identified were:¹¹⁶²

- Vague guidance that insufficiently described how to do operation assessments and inconsistencies about the balance of quantitative and qualitative information and the level of rigor required;¹¹⁶³
- Lack of practitioner training, leaving staff officers assigned to assessment billets to interpret vague doctrine on their own “or to hunt for assessment products created by others that they [could] copy;”¹¹⁶⁴
- Overuse of quantitative metrics to assess progress even in the “absence of a credible numbers-based theory of counterinsurgency.” Related issues included the mistaken assumption that quantitative metrics were inherently objective and the use of “junk arithmetic” that misapplied quantitative values to qualitative information;¹¹⁶⁵
- Collecting data on the chance that the information would be useful in the future, whether or not the data was actually relevant;¹¹⁶⁶
- Lack of transparency, including the use of color-coded assessments that masked nuance (see pp. 152–153 for application of this critique to Resolute Support’s district control maps);¹¹⁶⁷
- Over-optimism and an “institutional drive to produce ‘good news stories’” and the “imperative to show progress in time to serve the ends of various political timetables.”¹¹⁶⁸

Of all the criticisms, the last was arguably the most serious, and arguably the hardest to prevent: As one author put it, there is a “natural desire for good news to pass on up the chain of command.”¹¹⁶⁹ In the words of one former senior military official:

As intelligence makes its way up higher, it gets consolidated and really watered down; it gets politicized. It gets politicked because once policymakers get their hands on it, and frankly, once operational commanders get their hands on it, they put their twist to it. . . . Operational commanders, State Department policymakers, and Department of Defense policymakers are going to be inherently rosy in their assessments. They will be unaccepting of hard-hitting intelligence.¹¹⁷⁰

As commanders cycled in and out of Afghanistan, a consistent narrative emerged—one summed up by former British diplomat and politician Rory Stewart in the 2011 book, *Can Intervention Work?*: “Each new general in Afghanistan from 2002 to 2011 suggested that the situation he had inherited was dismal; implied that this was because his predecessor had had the wrong resources or strategy; and asserted that he now had the resources, strategy and leadership to deliver a decisive year.”¹¹⁷¹

The case study that follows explores this and another criticism generally levied against operation assessments in Afghanistan: the challenge of selecting the right metrics against which progress could be assessed.

“Each new general in Afghanistan from 2002 to 2011 suggested that the situation he had inherited was dismal; implied that this was because his predecessor had had the wrong resources or strategy; and asserted that he now had the resources, strategy and leadership to deliver a decisive year.”

—Rory Stewart

The McChrystal Assessment: Under-resourcing Could Lose the War

In 2009, debate about what to do about Afghanistan dominated foreign policy discussions within the new administration of President Barack Obama. President Obama had expressed support for rooting out al-Qaeda in Afghanistan as early as 2002 when he was still a young state senator. Both during his presidential campaign several years later and after he was elected president, he vowed to recommit the U.S. to what he called the “just war.”¹¹⁷² The result was a series of high-level assessments that aimed to chart the path forward in Afghanistan.¹¹⁷³ Amidst the debate, advocates of counterinsurgency attempted to strengthen their case by encouraging new ISAF commander General Stanley McChrystal to conduct his own on-the-ground assessment.¹¹⁷⁴

The final product became one of the most well-known documents of the war, both for its criticisms of Afghan government corruption and because it was leaked to the press in an apparent attempt to influence imminent decisions about possible troop increases.¹¹⁷⁵ Informed in large part by McChrystal’s “personal experience and core beliefs,” including

a predisposition towards counterinsurgency, the assessment sought to answer three key questions:¹¹⁷⁶

- Can ISAF achieve the mission?
- If so, how should ISAF go about achieving the mission?
- What is required to achieve the mission?

McChrystal painted a dire picture. Warning that ISAF suffered from a “culture of poverty,” the assessment asserted that under-resourcing U.S. and international efforts to stabilize Afghanistan “could lose [the war].”¹¹⁷⁷ Afghans, the assessment said, were “frustrated and weary after eight years without evidence of the progress they anticipated.”¹¹⁷⁸ Such frustration was inflamed in part by “widespread corruption and abuse of power by various officials,” giving “Afghans little reason to support their government.”¹¹⁷⁹ Contracting practices were part of the problem, the assessment added, as in many cases contracts had merely “[enriched] powerbrokers, corrupt officials, or international contractors” without much benefit to ordinary Afghans.¹¹⁸⁰

Exacerbating these challenges was ISAF’s inadequate execution of counterinsurgency, the assessment said, with ISAF’s current approach suffering from a lack of unity in both command and effort, while focusing too little on protecting Afghans.¹¹⁸¹ The solution, in essence, was a “comprehensive counterinsurgency campaign” that would prioritize accountable Afghan governance, focus resources on geographic areas deemed most critical, and increase the size of the ANDSF, all while regaining initiative from the Taliban.¹¹⁸² It was a new strategy that called for more resources.¹¹⁸³

The assessment’s style and conclusions reflected aspects of the criticism levied against operation assessments over the course of the U.S.-led intervention. In the following pages, we explore two specific themes.

“The big picture question is . . . about what are we doing here [in Afghanistan]? What did we get for this \$1 trillion effort? Was it worth \$1 trillion? These conversations are only happening in private.”

—Former senior U.S. official

Overoptimism Resulted from False Assumptions

First, while the assessment was blunt and critical in some ways, it was overly optimistic in others. Following what became a recognizable pattern, McChrystal asserted that with additional resources and a new strategy “success is achievable.”¹¹⁸⁴ Yet the scale of potential success envisioned in the assessment remained elusive, even as the new strategy was implemented and as additional resources poured into Afghanistan. Several former officials attributed this to unexamined assumptions. Writing in *Foreign Affairs* four years after the assessment was produced, Karl Eikenberry, a retired three-star general who was serving in a civilian capacity as the U.S. ambassador to Afghanistan



President Obama meets with U.S. Ambassador Eikenberry and General McChrystal at Bagram Airfield on March 28, 2010. (White House photo)

at the time the assessment was formulated and as the counterinsurgency strategy took shape, said the assessment relied on three critical assumptions:¹¹⁸⁵

- That protecting the Afghan population was possible;
- That increased foreign aid would increase the Afghan government's ability to function;
- And that counterinsurgency would be conducted in cooperation with a supportive Afghan government.

In the words of Eikenberry, these assumptions proved to be “spectacularly incorrect,” rendering the counterinsurgency campaign “incoherent and difficult to prosecute.”¹¹⁸⁶ As one former senior U.S. official who served on McChrystal's assessment team put it, military leaders “had faith that what happened in Iraq would happen in Afghanistan, but that never materialized.”¹¹⁸⁷ During a book tour nearly a decade after the assessment was finalized, McChrystal reportedly described a conversation he had with Secretary of State Michael R. Pompeo during which the Secretary asked advice on the way forward in Afghanistan. According to reporting, in response, McChrystal told Pompeo, “I'm not sure what the right answer [is]. My best suggestion is to keep a limited number of forces there and just kind of muddle along and see what we can do.”¹¹⁸⁸ But just what the U.S. could accomplish in the face of extraordinary obstacles to success was not always clear. In the words of the aforementioned senior U.S. official, “The big picture question is . . . about what are we doing here [in Afghanistan]? What did we get for this \$1 trillion effort? Was it worth \$1 trillion? These conversations are only happening in private.”¹¹⁸⁹

Struggling to Find the Right Metrics to Measure Progress

Second, the assessment directly reflected the challenge of selecting the right metrics by which progress could be assessed. When McChrystal arrived in Afghanistan in



U.S. Marines conduct a security patrol in Helmand Province in August of 2009. (U.S. Marine Corps photo)

June 2009, he asked his staff to brief him on the current state of the war.¹¹⁹⁰ In the telling of author and *Washington Post* reporter Rajiv Chandrasekaran:

They barraged him with “metrics” on PowerPoint slides—the frequency of Taliban attacks and the proportion of them the military deemed effective; the number of improvised explosive devices that had detonated and the few that troops had been lucky enough to discover before they blasted; the number and location of Afghan security forces across the country; and an evaluation of the Afghan government’s effectiveness in each province.¹¹⁹¹

Bluntly, in the words of Chandrasekaran, the conclusion was “obvious: The Americans and their NATO allies were losing.”¹¹⁹² But there was an irony inherent in these and other metrics generally used to assess both warfighting and reconstruction success—metrics frequently reported by SIGAR over the years. While, in aggregate, such metrics were broadly useful in concluding that there was a *lack of* progress, the absence of measures that could be interpreted and used to assess *forward* progress became a familiar source of critique.¹¹⁹³

The assessment itself seemed to vacillate on the utility of available metrics. On one hand, the assessment concluded that “many indicators suggest the overall situation is deteriorating,” implying that current metrics could be used to determine how well warfighting and reconstruction were progressing, or failing to.¹¹⁹⁴ But in another section, the assessment called on ISAF to “identify and refine appropriate indicators to assess progress, clarifying the difference between operational measures of effectiveness critical to practitioners on the ground and strategic measures more appropriate to national capitals.”¹¹⁹⁵ In short, the assessment used available indicators to conclude that the situation in Afghanistan was deteriorating while at the same time calling for better indicators to measure effects.

In the words of one researcher writing nearly two years after the assessment was produced, “criticisms of our ability to measure and demonstrate progress in a clear, credible, and transparent manner have only increased. . . . By far the most popular criticism . . . is that we do not have the right metrics for Afghanistan.”¹¹⁹⁶ The highlight below discusses the strengths and weaknesses of what became one critical, campaign-level metric: district control.

DISTRICT CONTROL: A CASE STUDY OF THE UTILITY OF A KEY CAMPAIGN-LEVEL METRIC

The main goal of warring parties is frequently control of territory and of the populations within.¹¹⁹⁷ In Afghanistan, where warfighting and reconstruction have been inextricably linked, district control became one first-order measure of how reconstruction was proceeding.¹¹⁹⁸

Territorial and population control can reveal how well combatants are performing: Simplistically put, the greater its degree of control, the better off one side is.¹¹⁹⁹ Trend lines showing how territorial control has shifted over a specified period can reveal which side in a conflict has momentum.¹²⁰⁰ In addition to purely military outcomes, a government's or an insurgent group's control of territory may also indicate that it is administratively functional in an area and able to provide services like education and healthcare.¹²⁰¹

For several years, Resolute Support (the current NATO-led military mission in Afghanistan) used variants of district control—including population and territorial control—to assess the overall progress of the conflict.¹²⁰² At the beginning of 2019, Resolute Support informed SIGAR that “it [was] no longer assessing district-level insurgent or government control or influence.”¹²⁰³ According to Resolute Support, “the district-level stability assessments were “of limited decision-making value to the [Resolute Support] Commander.”¹²⁰⁴ Data available at the time—the number of attacks initiated by insurgents, Afghan security force casualty trends, and overall security incidents—showed that Afghanistan was experiencing heightened insecurity.¹²⁰⁵ The discontinuation of the collection of district control data raised concerns that Resolute Support was depriving the Congress and the public of a key metric useful for assessing the effectiveness of reconstruction.¹²⁰⁶

Control remains an important implicit metric. In July 2020, DOD stated that one sign of progress in the Afghan government's ability to meet shared security objectives was that “no provincial capitals [were] captured and minimal district centers [were] seized by [the] Taliban.”¹²⁰⁷

How Control Was Assessed

Resolute Support never provided clear, unclassified criteria for how, specifically, it measured district control, making it difficult to validate its methodology. The information that was available showed that it used a number of factors to subjectively assign districts to one of five categories: insurgent control, insurgent influence, contested, Afghan government influence, and Afghan government control.¹²⁰⁸ To determine the control status of a district, evaluators from Resolute Support's Train, Advise, and Assist Command accounted for governance, security, infrastructure, the economy, and communication.¹²⁰⁹ Inputs included information on Afghan military operations and “feedback from the Afghan population,” among other sources.¹²¹⁰ Although Resolute Support did not explicitly define government control, as of February 2016, evaluators were directed to consider infrastructural and economic factors in their assessments.¹²¹¹



Afghan Commandos prepare for clearing operations in the Mohmand Valley in February of 2018.
(U.S. Army photo)

Since the unit of analysis was the district, population control was calculated by summing the estimated number of Afghans living in districts controlled by one side. In sum, Resolute Support produced three different measures of district control: (1) the number of districts controlled; (2) the total land area implied by the number of districts controlled; and (3) the total population controlled based on the number of districts controlled.¹²¹²

Limitations

Although it was a useful metric in many ways, district control had several limitations:

Violence was not directly accounted for. For security, an early definition of full Afghan government control was “ANDSF dominant. [Insurgent] attacks are rare and ineffective.”¹²¹³ However, in practice, violence levels could vary significantly even as control remained unchanged.¹²¹⁴ The best example was Kabul, which the Afghan government controlled, but where levels of violence eventually grew very high.¹²¹⁵

Data inputs did not always keep pace with rapid changes. Significant, frequent internal migration may have caused population data to lag behind sometimes rapid shifts in the actual number of Afghans present in a district. This raised the possibility of potential inaccuracies in assessments of population control, especially given that the most recent population density data for a given district was several years old.¹²¹⁶ What’s more, increased levels of population control did not necessarily signal greater security. For example, migration from contested areas to government-controlled areas because of rising violence increased the population (not the territory) under government control—while at the same time reflecting rising overall insecurity.¹²¹⁷

The unit of analysis may have concealed nuance. Because the unit of analysis for district control assessments was the district (rather than smaller geographic areas), nuance was sometimes concealed. Although, according to the assessments, the Taliban have generally controlled a fairly extensive amount of territory, the ANDSF have generally successfully

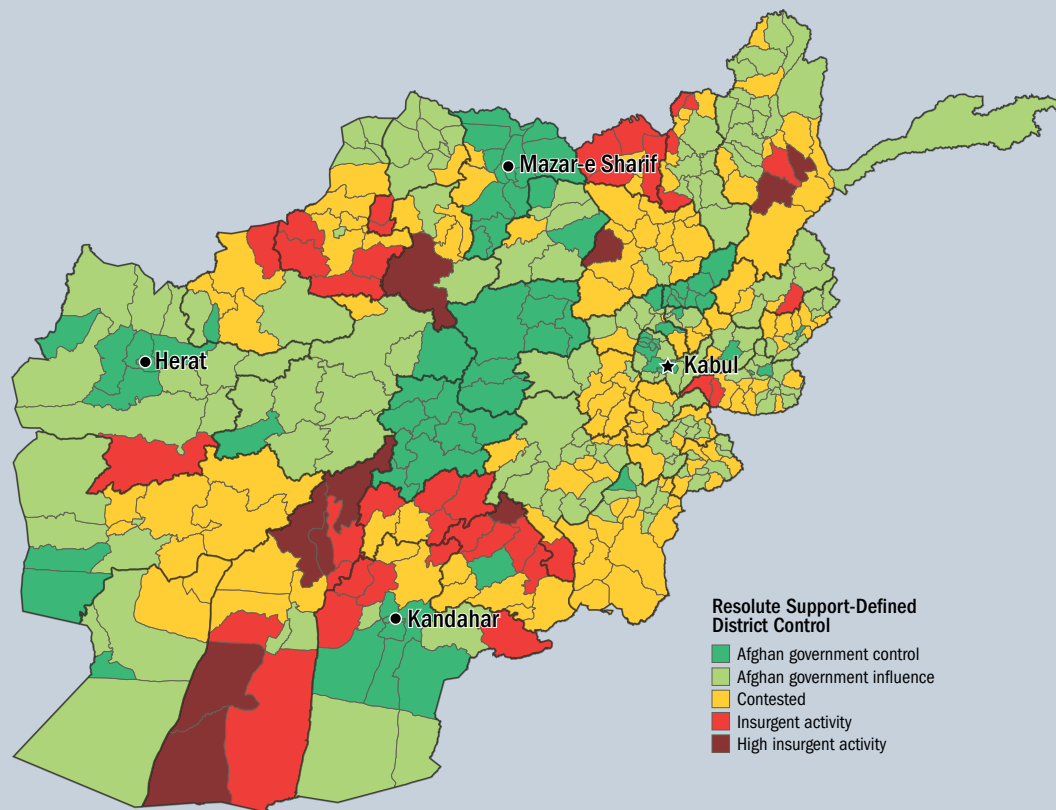
defended geographically smaller, but heavily populated, urban areas.¹²¹⁸ The way in which the data was most prominently presented—in the form of a color-coded map (Figure 8 below)—tended to mask this distinction. Colors shifted over district boundaries that, though perhaps administratively significant, were largely arbitrary for the purpose of the control analysis.

Limitations Aside, District Control Was a Conceptually Useful Metric

Despite the limitations described above, district control assessments provided critical information on the overall progress of warfighting goals to which contracted reconstruction efforts have contributed.¹²¹⁹ But it was not always clear which specific indicator—the number of districts, the geographic area, or the proportion of population under the control of one side—was most significant. Drawing broad conclusions from any or all of these macro-level indicators (such as numbers of districts controlled) is best done with the caveat that they may conceal granular but important variations in underlying detail, or partially omit factors (such as changing levels of violence) that may provide key context.¹²²⁰ Even so, territorial and population control remain fundamental (though very basic) indicators of military success. Despite some limitations, therefore, district control assessments provided useful longitudinal data to evaluate warfighting and reconstruction progress.¹²²¹

FIGURE 8

EXAMPLE OF A DISTRICT STABILITY ASSESSMENT



Note: SIGAR originally published this graphic in the October 2018 quarterly report. Each district was assigned a district stability level based on its overall trend of land-area and population control. The district stability levels listed in the key of this map do not correspond exactly to the terminology used in Resolute Support's September 2018 narrative response for district control (i.e., High Insurgent Activity vs. Insurgent Control).

Source: Resolute Support, response to SIGAR data call, September 19, 2018.

CHAPTER SUMMARY AND KEY FINDINGS

To achieve reconstruction objectives in Afghanistan, DOD made extensive use of contracts and grants. Without a standardized monitoring and evaluation system that could be applied uniformly to all its activities, DOD's M&E efforts evolved in piecemeal fashion in response to the challenges of each individual program. Like its counterpart agencies, DOD in many cases limited measurement to inputs and outputs. The reasons for limited measurement of outcomes and impact varied by program and by line of effort. In some cases, as with assessing the capacity of the ANDSF, what DOD was attempting to measure was inherently abstract and difficult to capture. Rating systems changed over time, offering less longitudinal consistency, but signaling a willingness to adapt. Still, some contracts intended to contribute to increasing the ANDSF's capacity lacked metrics altogether. In those cases, return on investment could not be ascertained, with funds spent in exchange for undetermined effects.

Lack of measurement in other areas, such as in the Commander's Emergency Response Program, reflected some combination of the difficulty of measuring impact in a complex environment and a lack of capacity to fully assess results—within the broader context of a belief system that seemed to assume effects were being achieved. However, in CERP's case, impact was rigorously assessed after the height of the program by a wide array of academic researchers and by a DOD-funded study. In the end though, even sophisticated regression analyses could not fully parse out CERP's effects. Optimists could claim that the research signaled the possibility of *specific* circumstances under which CERP may have produced favorable impacts. But taken as a whole, the research painted a picture of impact that was at best mixed and at worst conflicting.

The Task Force for Business and Stability Operations presented another example of unclear impact. The task force was conceived in the same spirit as CERP, working towards the objective of promoting stability. It relied on an unstructured form of M&E that was largely ad hoc and often impressionistic. Determination of the task force's effects was left to an economic "impact assessment" that in actuality merely projected *future* economic impact. Moreover, the assessment did not attempt to determine the impact of task-force projects on Afghanistan's stability, avoiding a walk all the way down the causal pathway presumed by the assumptions guiding the task force's work.

For the majority of their implementation, no attempts were made to examine the impact of other programs, such as the Afghanistan Infrastructure Fund. While some intermediate counterinsurgency benefits were assumed, they were never fully validated. Meanwhile, the military's overarching, M&E-like process—operations assessment—seemed to repeatedly fail. Entrenched overoptimism and a lack of clear metrics—both issues of which were reflected in then International Security Assistance Force Commander General Stanley McChrystal's 2009 assessment of the war—contributed to that failure.

As pressure for greater oversight of funds and demonstration of impact grew, M&E at DOD evolved and became more demanding over time, culminating in the introduction of an M&E policy for security cooperation. The policy introduced the possibility of greater M&E standardization and rigor. Overall, DOD's experience with M&E in Afghanistan demonstrates both the challenges of measurement in a complex environment and the imperative of at least attempting to try.

Key Findings

- For most of the Afghanistan reconstruction, DOD lacked the kind of standardized M&E systems used by State and USAID. However, it did have analogous systems and processes.
- Although lack of standardization does not necessarily signal inherent weakness in M&E, it makes these systems and processes more difficult to compare and evaluate.
- Still, numerous parallels in implementation are evident. Like its counterpart agencies, DOD struggled to measure impacts, instead taking stock mostly of inputs and outputs.
- Those efforts to measure impact that did occur were often either unsatisfying or inconclusive. For example, while DOD's impact assessment of the Commander's Emergency Response Program was useful in some ways, it could not determine the independent effects of the program.
- Operation assessments hold the potential to be a valuable way to measure the aggregate effects of military operations, including the effects of contracted activities. However, their usefulness in Afghanistan was limited by a complex environment that made it difficult to gather actionable data and identify useful metrics. At times, the credibility of these assessments was undermined by overoptimism and other analytical pitfalls.
- In cases where contracted activities (or grants) were intended to produce counterinsurgency outcomes (as with the Commander's Emergency Response Program), key assumptions and theories of change were never fully validated, but instead were often simply assumed to be true.



CHAPTER 5

INTERNATIONAL EXPERIENCE AND PRACTICES

USAID photo

THE UNITED STATES WAS NOT ALONE IN STRUGGLING WITH M&E

In Afghanistan, M&E challenges were not unique to the U.S. government. Most other donors had documented guidelines stressing data integrity, continuous evaluation, and appropriate course correction. How such policies played out on the ground, however, is a separate question. This chapter surveys the M&E practices of both multilateral institutions and international donors, and examines techniques that may be useful in Afghanistan or in other volatile contingency environments.

The panoply of international, bilateral, and NGOs working in Afghanistan reflected wide-ranging missions, mandates, philosophies, and operational scales—differences which to some extent dictated their approach to M&E. For example, the UN put M&E in the context of risk management, while the UK's Department for International Development took more of a “value for money” approach that emphasized aid efficiency.¹²²² Some smaller donors, such as Denmark, delegated monitoring to implementing partners.¹²²³

Yet the variation in approaches could not fully resolve the fundamental challenge of practicing M&E in insecure environments. In a 2016 survey of M&E experts and practitioners from non-governmental and UN organizations operating in four conflict-

affected countries (including Afghanistan), respondents reported that the practice of M&E was “problematic.”¹²²⁴

In Afghanistan, respondents reported that the biggest problems were lack of M&E capacity in local partners and aid agencies (such as a dearth of M&E personnel and low levels of technical know-how in existing personnel), a lack of time to plan and implement M&E, and a general lack of willingness among agencies to share data or lessons with one another.¹²²⁵ They also cited “project-based and short-term funding structures” and “overly complex systems and tools for insecure contexts” as significant issues. Unsurprisingly, given all of these constraints, the last major problem reported by respondents was inflated expectations about what M&E could actually achieve.¹²²⁶

In short, the challenge of developing and implementing effective M&E systems in an environment like Afghanistan was not unique to the U.S. government. The first part of this chapter explores various M&E techniques employed by multilateral institutions, NGOs and other donor countries. It is organized by institution and country (with one subsection summarizing the experience of NGOs). The next part of the chapter explores common themes observable across those countries and institutions. The chapter concludes with suggested best practices emerging from the international experience.

The World Bank: Despite a Layered M&E System, Complicated Lines of Accountability and Gaps in Oversight and Performance Measurement

To oversee the **Afghanistan Reconstruction Trust Fund** (ARTF) the World Bank employed two types of third-party monitors: a monitoring agent for recurring civilian operating expenses, such as government salaries, and a supervisory agent for development projects.¹²²⁷ Under the ARTF, third-party monitoring began with tracking the progress of major infrastructure projects, but was later expanded to capacity-building activities

Defining Key World Bank Terms

The **Afghanistan Reconstruction Trust Fund** is a pool of money created by donations from the United States and 33 other countries that provides on-budget financing to the Afghan government.¹²²⁸

Self-evaluation is a systematic assessment of the design, implementation, and results of an ongoing or completed World Bank project written by or for the operational department in charge of the project.¹²²⁹ Within the Bank’s system, **independent evaluations** are intended to provide an additional layer of rigor by validating the results of self-evaluations, either through a direct examination of a self-evaluation or via a separate assessment of the relevance, efficacy, and efficiency of the project.¹²³⁰

Grievance mechanisms offer a way for people to lodge complaints about, and resolve disputes related to, development projects implemented in their communities. Such complaints and disputes can range from land, water, or labor issues to fraud and corruption.¹²³¹



Afghan girls attend a science class supported by the Education Quality Improvement Program in Herat. (World Bank photo)

(including assistance to Afghan ministries to better track and report on ARTF-supported projects), and social and environmental monitoring—for example, ensuring that a project did not have any adverse effects on the local community.¹²³²

According to World Bank policy, monitoring is supplemented by **self-evaluation** by the Bank, in conjunction with the Afghan government, with additional **independent evaluation**.¹²³³ The Afghan government's approach includes community monitoring (as in the National Solidarity Program, discussed below), **grievance mechanisms**, and use of information and communications technology.¹²³⁴ The Bank distinguishes impact evaluations, which aim to establish causality and are done more selectively, from independent evaluations.¹²³⁵

The Bank's Multilayered M&E System Was Still Not Always Enough to Prevent Lapses in Oversight and Performance Measurement

The Bank's multitiered M&E system gives the impression of a well-oiled machine positioned to prevent lapses in oversight. However, in the unique environment of Afghanistan, it was not always up to the task. For example, as SIGAR reported in October 2019, a fiduciary review of the Bank's \$418 million **Second Education Quality Improvement Project** revealed significant weaknesses in the record-keeping practices of Afghanistan's Ministry of Education and identified several instances of potential procurement fraud.¹²³⁶ According to the Bank, the review was prompted by allegations of possible corruption in the education sector and problems in the fiduciary arrangements of the education project.¹²³⁷ Out of a sample of \$156.5 million in project expenditures, \$21.9 million (14 percent) lacked adequate documentation,

The **Second Education Quality Improvement Project** aimed to increase access to basic education in Afghanistan, particularly for girls. It provided school grants, teacher training, and technical assistance to the Ministry of Education. While it was being implemented, it was the Bank's flagship education project.

World Bank, "Afghanistan - Second Education Quality Improvement Program," n.d., accessed December 1, 2020.

\$2.2 million (1.4 percent) did not comply with World Bank procedures, and \$37.2 million (23.8 percent) in procurement contracts were identified as potentially fraudulent.¹²³⁸ The possibly fraudulent contracts involved a local non-governmental organization that was supposed to provide teacher training services.¹²³⁹ The expenditures for which there was inadequate documentation all related to salary payments made to teacher trainers.¹²⁴⁰

The sample of examined expenditures represented 37.4 percent of the total project cost and the total amount of potentially compromised expenditures was 39.2 percent of the sample.¹²⁴¹ Although the Bank emphasized that the results could not necessarily be extrapolated out to all project expenditures because sample expenditures were not randomly selected, it noted that if expenditures for the entire project had been inadequately documented at the same rate, the total amount of compromised funds would be approximately \$165.0 million.¹²⁴² As a result of the review, and to its credit, the Bank said it had implemented additional measures intended to bolster fiduciary oversight, including enhanced third-party monitoring and procurement assessments and in-depth reviews of other high-risk projects.¹²⁴³ Still, the magnitude of possibly compromised funds raised serious concerns about whether, at the ground level, donor monies provided to the ARTF were being spent as intended.¹²⁴⁴

There were also gaps in performance measurement. A 2018 SIGAR audit criticized the Bank for inconsistent compliance with its own guidance for measuring and reporting on the outcomes of ARTF development projects. Specifically, in several instances examined by SIGAR, the Bank did not develop performance indicators in line with guidance, did not provide specific quantitative or qualitative status information for performance indicators, or did not provide clear support or justification for the performance and progress ratings it gave projects. In one example, a performance report for the aforementioned Second Education Quality Improvement Project did not provide current information for 21 of the project's 46 performance indicators. In another example, a performance report for an ARTF agriculture project, presented an overall progress rating of "Moderately Satisfactory" despite the fact that the project had not made any progress on nearly 70 percent of its performance indicators. It was unclear how a project that was not making progress on more than two thirds of its performance indicators could be facing only moderate shortcomings.¹²⁴⁵

Under the ARTF, Third-Party Monitoring Was Expensive and Best Suited to Simple Data Collection

Gaps in oversight and performance measurement were not the only challenges. Although the principle of third-party monitoring—providing an external, objective view—had advantages, it also had drawbacks. The first was its cost. A 2014 World Bank review of third-party monitoring of ARTF-funded programs concluded that third-party monitoring was useful but expensive (\$11 million per year), partly because it required maintaining a 16-member staff to provide technical expertise in Kabul.¹²⁴⁶

The report also warned against a misperception that third-party monitoring was the primary monitoring mechanism, as opposed to the projects' own systems—a misperception that could encourage donors to discount internal project monitoring data and potentially overlook the opportunity to compare it against data collected by the third-party monitor.¹²⁴⁷ Additionally, it concluded that third-party monitoring was better suited to physical asset monitoring (such as assessing the quality of building materials or the maintenance of facilities) and other quantitative questions (such as water quality measures), and less suited for collecting more complex social indicators.¹²⁴⁸

The ARTF's Structure as a Pooled Trust Fund Complicated Donor Efforts to Monitor How Money Was Spent

The complicated lines of accountability created by the structure of the ARTF represented an additional challenge—though for donors rather than the Bank itself.¹²⁴⁹ USAID, for example, donated funds in accordance with its own goals and strategies, and was answerable to the Congress for how its money was spent. But once USAID money went into the fund, the agency could not always track how it was used.¹²⁵⁰ The previously mentioned 2018 SIGAR audit of the ARTF criticized the World Bank's reluctance to provide access to information—a violation of its own policies—as one reason why SIGAR could not assess the extent to which the Bank was measuring and reporting on progress.¹²⁵¹ The report also noted that the structure of the ARTF actually prohibited the World Bank and donors from withholding funds or taking other steps when projects performed poorly.¹²⁵² Two of SIGAR's recommendations were that the Bank provide donors with complete versions of the monitoring agent's reports and that the monitoring agent and supervisory agent themselves be periodically evaluated.¹²⁵³

The United Nations Tries to Meet Demand for Accountability

Much of the UN's work is done in fragile states, and donors are increasingly demanding a greater focus on resource stewardship. The UN's response has included the establishment of risk management units—initially in Somalia and then, starting in 2014, in Afghanistan.¹²⁵⁴ The purpose of the Afghanistan risk management unit was to advise UN agencies on how to better monitor their projects and produce reports that addressed, for example, obstacles to effectively mitigate the risks of corruption and fraud.¹²⁵⁵

The risk management unit was useful as a way of coordinating and promoting minimum monitoring standards across disparate UN agencies. For example, risk management units developed minimum due-diligence standards designed to help UN entities validate data and information received from their contractors and implementing partners.¹²⁵⁶ They also surveyed risk-management practices across UN agencies that, among other results, revealed significant differences in patterns of fraud reporting: although 5 agencies reported 77 cases of suspected fraud, 7 agencies reported none.¹²⁵⁷ That disparity suggested further research into whether fraud detection mechanisms were adequate and whether, within some UN agencies operating in Afghanistan, fraud was underreported.¹²⁵⁸ The risk management unit also found uneven use of information and

communications technology, with various UN agencies responding in ad hoc ways that were not fully integrated into monitoring processes.¹²⁵⁹

The Law and Order Trust Fund for Afghanistan: Verdict Still Out on Efforts to Use M&E for Evidence-Based Decision Making

One of the most prominent UN aid mechanisms was the Law and Order Trust Fund for Afghanistan (LOTFA), which was created in 2002 primarily as a way to pay the salaries of the Afghan police. In 2011, its mission expanded to broader institutional strengthening of the Ministry of Interior and other entities, including the establishment of human resource and payroll systems.¹²⁶⁰ Its mission was expanded again in 2015 to support the transition of the Afghan police from counterinsurgency to civilian policing.¹²⁶¹

The fund was administered by the UN Development Programme, which, as did the ARTE, employed a monitoring agent. LOTFA's financial accountability was generally regarded as problematic. After a series of negative media reports that included wide coverage of the UN's firing of three senior LOTFA officials due to corruption concerns, including the fund's finance officer, and after several critical performance assessments, the UN Development Programme increased its level of monitoring.¹²⁶²

More recently, during 2017 and 2018, the UN Development Programme rolled out a more sophisticated M&E system that attempted to transform LOTFA's M&E from “a function to fulfill corporate reporting requirements and to please donors” to one “central to evidence-based programming and decision-making.”¹²⁶³ The new M&E system consisted of three components: “gadgets” (mobile data collection tools, a centralized impact-and-results platform, interactive dashboards and data visualizations, and mapping tools to present data), “voices” (surveys and citizen feedback mechanisms), and “fun” (“data parties” and other applications that allowed various stakeholders to engage with the data).¹²⁶⁴ While these various sources supplement each other in producing a general picture of attitudes towards the police and of other indicators, they were not designed for triangulation of specific data, as was the purpose of USAID's multitiered monitoring system, which similarly intended to draw information on projects from a variety of sources.¹²⁶⁵ The UN Development Programme has reported that the new tools work well in Afghanistan, but it remains to be seen how they—and other new forms of information technology—will be used for adaptive management, and what their ultimate impact on outcomes will be.¹²⁶⁶

The World Food Programme Further Demonstrates the Limits of Third-Party Monitoring

The UN World Food Programme—which receives funds from USAID among other donors—distributes emergency food assistance to populations in need. It has relied on third-party monitoring since 2007 to monitor its partners' activities, a practice it uses globally in less accessible project areas. As of 2014, the third-party monitors were a mix of Afghan and international firms and NGOs.¹²⁶⁷ Because it often serves populations in insecure, hard-to-reach areas, its work is subject to numerous risks, including



Afghans pump water from a well funded by the World Bank and facilitated through the National Solidarity Program in 2008. (World Bank photo)

diversion of food supplies by insurgent or criminal groups.¹²⁶⁸ Consequently, continual monitoring of operations is especially important.

Despite the use of these teams, monitoring was constrained by many of the same factors that affected other such efforts: lack of technical capacity, problems with data verification and management, and inability to measure outcomes.¹²⁶⁹ A 2012 report contracted by the program noted a “clear overreliance of WFP’s area offices on [third-party monitors] for M&E despite their questionable records in terms of monitoring,” and instances where third-party monitoring teams colluded with local communities or leaders to cover up fraud. The report also noted that the monthly reports focused mainly on counting numbers of beneficiaries served and volume of food distributed, rather than making larger observations about program progress and challenges.¹²⁷⁰ A 2019 SIGAR audit report echoed this observation, noting that updates provided to USAID by the World Food Programme were limited mostly to administrative matters rather than progress on achieving objectives.¹²⁷¹

NGOs Attempt to Balance Accountability with Political Neutrality

Many NGOs have been working in Afghanistan (and even in the same communities) for decades. Some come out of the emergency humanitarian-assistance tradition, in which monitoring and accountability consist essentially of maintaining close engagement with communities by establishing a reputation as nonpolitical actors. Although much of the collected information is informal and ad hoc, some NGOs have developed more formal systems, such as compliance checklists for beneficiary selection and accountability, quality standards, and organizational policy.¹²⁷²

The anticorruption NGO Integrity Watch Afghanistan (IWA) developed a community-based monitoring program for infrastructure, mining, schools, and the courts, in which volunteers from project areas were trained to use focus groups, surveys, public meetings, and social media to identify corruption and encourage accountability.¹²⁷³ For infrastructure, IWA provided technically trained monitors to assess building techniques and material quality; for schools, local monitors conducted baseline surveys to ascertain the current state of education in a community, then used monitoring forms to submit weekly and quarterly information.¹²⁷⁴ NGOs who were facilitating partners for the Government's National Solidarity Program and its similar successor program, Citizens' Charter, also used a system of community-based monitoring.¹²⁷⁵

A basic problem these NGOs faced, however, was maintaining their stance as nonpolitical actors. Given the large foreign military presence and the political nature of the international community's state-building focus, insurgent groups increasingly targeted these aid agencies as allies of the Afghan government or foreign "invaders." As insecurity increased and access became more constrained, NGOs lost the ability to closely oversee their programs. In response, from 2014 onwards, NGOs increasingly relied on remote management and third-party monitoring.¹²⁷⁶

Germany Takes a Well-organized Approach to M&E; Still, Many Evaluations Focused Mostly on Completed Tasks

The program planning and M&E practices of Germany, Afghanistan's second-largest donor of development assistance, were similar in spirit to USAID's. A short policy paper described the M&E system's five components: individual projects' results framework-based M&E; a centralized database for project information (the Development Tracker for Afghanistan, widely known as the DevTracker); a separate monitoring, evaluation and communication project; third-party assessments and reviews of the portfolio of German development assistance that included a longitudinal impact study; and support for large-scale perception surveys.¹²⁷⁷ The system was rolled out in 2009 in response to public debate in Germany about spending in Afghanistan.¹²⁷⁸

The DevTracker was a data warehouse for activities and outputs related to German development assistance, which ranged from vocational training to efforts to improve governance. The intent of the DevTracker was similar to that of USAID's Afghan Info, and it was used for a variety of purposes, including providing data for third-party assessments and for reporting to policy makers. However, although the DevTracker has a large volume of activity and output data, it does not reflect outcomes or impact, which are left to individual projects to measure—which they do, at various levels of quality. Additionally, because the systems that monitor outcomes are only in place for the project's duration, it is very difficult to look at outcomes or impact beyond that period, a recurrent issue among development agencies.¹²⁷⁹

A 2014 review commissioned by the German Institute for Development Evaluation to look at the "types, scope, quality, and usefulness of all evaluations" of German development cooperation in Afghanistan concluded that project-level evaluation reports

were of “relatively good quality” and that the results were used by officials and other actors. But, the review noted, those evaluations tended to focus on outputs, such as successful installation of new drinking-water systems, and considered outcomes or impacts only in passing or not at all.¹²⁸⁰

According to the review, “outcomes and impacts are either just assumed”—meaning projects’ theories of change went unquestioned by evaluations—“or, at best, reviewed for plausibility.”¹²⁸¹ The review concluded that, although outcome measures appeared to be mostly positive, “this positive result contrasts with the fact that there is hardly any empirical proof for the effects (outcomes and impacts) of the [development cooperation] projects” or even for entire programs focused on priority areas like governance and education.¹²⁸² Moreover, it attributed the lack of substantive content to domestic political factors: the need to “continually show progress and report quick successes,” which caused M&E to focus on measuring completed tasks. This focus often prevented the kind of critical learning that was crucial to improving project performance.¹²⁸³

One interesting exception to the broad emphasis on outputs—and a notable component of the German approach—was a multiyear study of the causal impact of aid in six districts in three northern provinces, where much of Germany’s assistance was focused. Four biannual surveys tracked public perceptions of security, attitudes towards foreign military forces and western development actors, and views of local Afghan government officials.¹²⁸⁴ In a challenge to the widespread claim that determining impact in Afghanistan

A Metareview Commissioned by the German Government Was a Useful and Unique Approach to M&E

A recent, unique component of Germany’s approach to M&E was a 2020 metareview of international development assistance to Afghanistan.¹²⁸⁵ Commissioned by the German Federal Ministry for Economic Cooperation and Development, the metareview analyzed 148 program and project evaluations and other reports (including SIGAR’s) published between 2008 and 2018.¹²⁸⁶ The objective was to “learn more about what has worked, what has not worked, and why.”¹²⁸⁷

The metareview made a significant contribution to the M&E of Afghanistan reconstruction. Looking widely across programs and projects allowed the metareview to draw broad conclusions about efficacy. For example, it found that modest, locally embedded projects with immediate, tangible results worked best, and that complex projects aimed at building capacity and changing behavior rarely worked.¹²⁸⁸

Equally useful was the metareview’s conclusion that interventions in some sectors were more effective than efforts in others. For example, the metareview concluded that health, education, and basic livelihood interventions led to results. By contrast, it found that capacity-building, rule of law, and gender interventions rarely worked and that stabilization projects were “mostly ineffective.”¹²⁸⁹

was prohibitively difficult, the review concluded that “conducting methodologically sophisticated impact analyses of development aid including extensive household surveys is also possible in Afghanistan.”¹²⁹⁰

The United Kingdom Strives to Move “As Far Down as Feasible on the Causal Chain”—Though It Sometimes Falls Short

The UK Department for International Development (DFID), whose programming in Afghanistan ranged from capacity-building at the Ministry of Interior to support for girls’ education, was the aid agency for Afghanistan’s fourth-largest donor of development assistance (DFID was merged into the Foreign, Commonwealth and Development Office in 2020). DFID had many elements in common with its U.S. counterpart, USAID. These included project logical frameworks (“log frames”—planning tools that articulate a project’s goal, activities, and anticipated results), use of third-party monitoring, and contracted third-party evaluation. Unlike USAID, DFID had no specific requirements built in for evaluations, although its staff reported that riskier and larger investments had a higher probability of being evaluated. Also unlike USAID, the UK agency procured an M&E supplier for individual projects based on perceived needs, rather than contracting with a single M&E provider to cover multiple projects—in part because it wanted flexibility and “didn’t want to put all of [their] eggs in one basket.”¹²⁹¹

As with other donors, DFID eventually shifted more to remote monitoring and use of local nationals, especially after the drawdown of British troops. Kabul-based staff expressed an interest in experimental designs, and stated that their evaluations tried to go “as far down as feasible on the causal chain.” Although they believed that establishing impact is usually not possible in Afghanistan, in some cases they used “theory-based” evaluations to look at whether a project’s original theory of change has held up in order to establish whether there was a plausible contribution to impact.¹²⁹²

However, these mechanisms have not always been up to the task of effectively revalidating the theories underpinning projects. Some reports from the UK Independent Commission for Aid Impact—formed in 2011 “to provide independent evaluation and scrutiny of the impact and value for money of all UK government [official development assistance]”—underscore instances where theory was insufficiently scrutinized, limiting the effectiveness of both DFID’s projects and its M&E systems.¹²⁹³ The commission has at times been very critical: A March 2014 review of the department’s efforts to build physical infrastructure and provide vocational training to Afghans so that they could obtain better jobs “[lacked] strategic coherence.”¹²⁹⁴ The commission added that significant weaknesses in project design, including “unproven theories of change,” made it difficult to assess the impact of projects against the UK’s overall strategic objective for Afghanistan, which was to create a “viable state.”¹²⁹⁵ Pointing to the “weak evidence base of ‘what works’ in Afghanistan,” the commission said that DFID needed to improve its M&E systems.¹²⁹⁶

DFID’s most ambitious M&E effort was the 2009 Helmand Monitoring and Evaluation Plan, designed “to improve the delivery and effectiveness of [the UK’s] stabilisation

efforts” in Helmand Province.¹²⁹⁷ Among other components, the Helmand Monitoring and Evaluation Plan involved quarterly perception polls of Afghans residing in targeted areas in order to monitor the effectiveness and impact of programming.¹²⁹⁸ According to a 2014 UK government review, this was the largest survey ever carried out in Helmand.¹²⁹⁹ But the review also said it was overly complex, and showed “how an M&E system will struggle to demonstrate impact if its overarching purpose is ill defined” and “how one M&E system has struggled to be everything for everyone.”¹³⁰⁰

Still, the UK’s experience with the Helmand Monitoring and Evaluation Plan yielded some useful lessons: According to the review, “Intervention hypotheses based on a well thought out [theory of change] should be tested through surveys and analysis to validate the intervention plan before undertaking the action or activities.”¹³⁰¹

MANY KEY THEMES FROM THE INTERNATIONAL EXPERIENCE REINFORCE THOSE FROM THE UNITED STATES

Several themes emerge from the international experience with M&E.

Sound M&E Policies Do Not Necessarily Equate to Sound M&E

Just because sound M&E systems exist on paper does not mean they will be implemented in useful ways. For example, Germany’s foreign assistance to Afghanistan is subject to guidelines for bilateral cooperation that require implementing agencies to regularly conduct evaluations of projects.¹³⁰² Although a 2014 review commissioned by the German Institute for Development Evaluation concluded that evaluations of the country’s development projects were “of relatively good quality,” they generally failed to consider outcomes and impacts.¹³⁰³ Consequently, the Institute said, evaluations offered limited use for strategic planning and decision-making.¹³⁰⁴

Nor will layers of M&E necessarily assure that funds spent will be spent effectively to achieve their intended purpose. The World Bank employs third-party monitoring and several different forms of evaluation to assess whether donor monies provided to the Afghanistan Reconstruction Trust Fund are achieving results.¹³⁰⁵ The Afghan government itself also performs M&E of Trust Fund projects; additionally, certain projects are subject to community monitoring by local Afghans.¹³⁰⁶ On top of all of these systems, the Bank is required to provide Trust Fund donors and the public with monitoring and performance information.¹³⁰⁷ However, as SIGAR reported in 2018, the Bank did not always develop performance indicators in line with its own guidance, provide specific status information for performance indicators, or provide clear support for the performance ratings it gave projects.¹³⁰⁸

Outcome and Impact Measurement was Lacking

The international experience with M&E shows that the United States was not alone in struggling with outcome and impact measurement. For example, in its review of German development assistance, the German Federal Ministry for Economic Cooperation and Development observed that “major donors have ... experienced deficiencies in



Workers on a National Solidarity Program road construction project move boulders. (World Bank photo)

their collection of data at the impact level.”¹³⁰⁹ Similarly, a March 2014 report by the UK Independent Commission for Aid Impact gave repeated examples of its inability to tell if a given program had achieved its desired impact—either because indicators being tracked were irrelevant or vague, or because results were not validated by implementing partners.¹³¹⁰ Moreover, a 2017 Australian government report on LOTFA noted that the reporting process captured progress on improving the Ministry of Interior’s payroll function. However, it added that the reporting process was too output-focused to measure capacity-building and institution-strengthening objectives intended to transition Afghanistan’s police from a largely militarized counterinsurgency force to a civilian police organization.¹³¹¹

Third-party Monitoring Can Help Address Site Access Challenges, but Has Limitations

In response to poor security conditions, many organizations turned to third-party monitoring, which a 2016 report by the Berlin-based Global Public Policy Institute called a “sizable industry,” with annual spending of around \$200 million in Afghanistan alone. Funds went to a mix of international and Afghan and for-profit and not-for-profit organizations, although the line between local and international, for-profit and not-for-profit was not always clear, nor was the line between third-party monitors and implementing partners.¹³¹²

Third-party monitoring was largely focused on compliance and output verification, mainly because data collectors’ ability to make judgements about what they were seeing or even to collect relevant information was limited by their lack of knowledge of a project’s logic or theory of change.¹³¹³ Monitors also risked undermining the reputation of the agency or an implementing partner through their behavior—for example, by

looking the other way on shoddy performance. This was especially problematic given that the highly localized relationships in Afghanistan meant that persons from the project area were the most likely—perhaps the only—people to get access to the community, creating the potential for the collection of incomplete or inaccurate data, conflicts of interest, or even corruption.¹³¹⁴ Finally, outsourcing monitoring tended to reduce the implementing partner or agency’s knowledge and institutional memory, as the staff responsible for the activity were then less likely to visit the project site area.¹³¹⁵

Experimental Designs Can Test Theories Linking Projects to Presumed Impact

Perhaps the most well-known exception to the lack of examination of actual results was a randomized controlled trial on the impact of the National Solidarity Program, an ambitious development program funded by the World Bank.¹³¹⁶ The study randomly assigned villages to two groups, one receiving the development aid and one not.¹³¹⁷ The purpose of the study was to test the theory of change associated with the “hearts and minds” model—namely, that “development projects will increase economic welfare, improve attitudes to government, and reduce insurgent violence.”¹³¹⁸ By testing the theory and ultimate outcome, the evaluation avoided the trap of many M&E efforts, which, by focusing on activities and outputs, risk giving an “A” grade to projects where the wrong thing was done perfectly.

Notably, the evaluation concluded that “development projects can prevent the spread of violence in relatively secure regions, but they are not effective in reducing violence in regions already experiencing significant security problems.”¹³¹⁹ While positive, long-run effects on violence were observed in regions with “moderate” levels of violence, there were “no effect[s] on attitudes toward government or security” in areas where violence was high.¹³²⁰ Hence, the National Solidarity Program (and by extension the “hearts and minds” model) appeared to be somewhat effective, but only under certain circumstances—an important conclusion made possible by the use of an experimental evaluation design. Moreover, in particularly insecure areas, the National Solidarity Program had a positive effect on economic outcomes, *without* an associated reduction in violence—a finding with significant implications for future project design.¹³²¹

KEY FINDINGS

- Sound M&E policies do not always equate to sound M&E practices.
- Outcome and impact measurement was lacking.
- Third-party monitoring can help address site access challenges but risks creating its own set of problems, including imprecise analysis and poor-quality reporting.
- Experimental designs can be effective in testing the theories tying projects to presumed impact.

A photograph of a U.S. Army helicopter in flight over a dusty, hilly landscape. In the foreground, several soldiers in full combat gear are seen from behind, walking along a path. The scene is set in a dry, arid environment with sparse vegetation and distant mountains under a clear sky.

CHAPTER 6

CONCLUSION

U.S. Army photo

The intervention in Afghanistan, along with Iraq, created a vast market for reconstruction services.¹³²² Although there is a tendency to think of these services as a discrete aspect of the overall effort to stabilize Afghanistan, contracting was so widespread that it wove its way into the fabric of the broader reconstruction and warfighting efforts. In many areas of reconstruction over the last two decades, implementing partners and other contractors acted as the primary agents of U.S. agencies. Their representatives on the ground and the entities physically implementing projects and providing services in support of U.S. strategic objectives were, in effect, the face of the U.S. government.

The previous chapters discussed the special challenges the Afghan environment posed for monitoring and evaluation and overall accountability, the uneven responses of the agencies, the external pressures to produce information, and the introduction of increasingly complex M&E systems and tools in response to criticism about widespread waste and inefficacy.

Overall, M&E of the reconstruction contracts, grants, and cooperative agreements falling within the scope of this report was not strong. Indeed, much of the report conforms very closely to findings and lessons from SIGAR's larger body of work. Touching on every major reconstruction sector—health, education, rule of law, women's rights, infrastructure,

security assistance, and others—that work 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.¹³²³

This does not necessarily mean that there were not attempts to improve M&E over time, some of which were significant. Innovations like multitiered monitoring aimed to diversify and prioritize sources of data and information. Third-party monitoring allowed agencies to continue to track the progress of projects even as security conditions restricted the ability of U.S. government personnel to move around the country. Meanwhile, M&E policies were introduced or further developed. At State, M&E systems began to look more like USAID's. At DOD, M&E procedures for the Commander's Emergency Response Program were refined and an assessment, monitoring, and evaluation policy for security cooperation was eventually developed. The end result is that all three agencies examined in this report—DOD, State, and USAID—now have M&E processes in place that are theoretically robust.

But an important contribution of this report is that the existence of policies and procedures does not necessarily translate into sound M&E. Whether it was the result of political pressure, a tendency towards overoptimism, or inadequate implementation, M&E as it was practiced consistently distorted reality. At USAID and State, program and project evaluations often obfuscated or equivocated. At DOD, programs like CERP and projects financed by the Afghanistan Infrastructure Fund continued to be held up as important tools for stabilization or counterinsurgency despite an utter lack of M&E. Across all three agencies, programs and projects were often either not assessed, or poorly assessed, against the political objectives they sought to achieve. These are but a few examples.

Ultimately though, what stands out the most is the extraordinarily delicate interdependence of M&E and the theories and assumptions underlying programming. Where those theories and assumptions proved to be shaky—as they often did—M&E as practiced generally did a very poor job of identifying and underscoring that core flaw. This created the very high risk of doing the wrong thing perfectly. Where the evidence base underlying a strategy or program was weak, M&E was ill equipped to pick up on that weakness. Where performance indicators were poorly selected and turned out to be irrelevant to programmatic objectives, M&E that relied on those flawed metrics was unable to credibly establish whether objectives had been met. Where the problem a project was trying to solve was heavily dependent on deeply entrenched external factors (such as corruption or even the war itself) that were difficult to materially influence, M&E tended to err on the side of positivity, failing to adequately consider evidence of stagnation or regression. Where M&E relied on flawed data, it came to ill-informed conclusions. All too often, the results were unsubstantiated assertions of efficacy.

The findings and lessons presented in this chapter touch on these and other core problems in further detail.

But the irony is that significant reform of existing M&E policies is not necessarily required. On the contrary, were DOD, State, and USAID merely to adhere more closely to the spirit of their current policies, the result would be fundamentally improved M&E. This is why many of the recommendations offered by this report simply ask agencies to practice what they preach.

Recent developments present a unique opportunity for agencies to do just that. As policymakers begin the process of resetting U.S. aspirations based on what can be reasonably accomplished, now is the time to rethink how progress towards new objectives should be measured. Unless policymakers determine that a failed Afghan state is an acceptable outcome, the United States may have to provide significant assistance for years to come. Our hope is that this report helps policymakers and practitioners avoid another two decades of poor M&E.

Beyond Afghanistan, this report also has further-reaching application in two important ways. First, that the United States periodically becomes involved in large-scale reconstruction efforts is almost axiomatic. Should policymakers choose to launch another major reconstruction effort—even several years or decades from now—this report’s core lessons will likely remain highly relevant.

Second, even in the present, many of the report’s observations are likely generalizable to other contexts. Although the sheer scale of U.S. efforts in Afghanistan renders them somewhat distinguishable, 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, that logic remains remarkably constant across U.S. involvement in numerous other countries and regions characterized by instability. The implication is that properly applying M&E to the innumerable contracts, grants, and cooperative agreements that are intended to advance strategic stabilization objectives is vital. In sum, for as long as the United States continues to pursue its foreign policy interests through a combined “defense, development and diplomatic” approach that seeks fundamentally to promote stability, many of the findings, lessons, and recommendations presented in this chapter are likely to remain highly relevant.

FINDINGS

We identify 11 major findings from our analysis of M&E of contracting efforts in Afghanistan:

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

The report identifies 10 lessons to inform future M&E efforts in Afghanistan:

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. USAID has well-developed M&E policies that date back decades. Many of State's current policies reflect USAID's. DOD has an Assessment, Monitoring, and Evaluation policy for security cooperation, doctrinal M&E-analogue processes, and, other M&E guidelines specific to Afghanistan reconstruction programs. The core tenets of these systems are generally sound. Policy and doctrine prescribe that the theories of change underlying projects must be evidence-based and capable of being revalidated; that specific, relevant, and measurable indicators be used to assess performance and effectiveness; and that rigorous monitoring and evaluation inform learning at both the programmatic and strategic levels (Appendix B details existing M&E systems at DOD, State, and USAID).

Yet, what exists on paper was not always put into practice. Examples include:

- **USAID:** According to the Automated Directives System, evaluations “should be readily understood and should identify key points clearly, distinctly, and succinctly.” Findings and conclusions should be “specific” and “concise.” Moreover, evaluations should “objectively evaluate the strategy, project, or activity.”¹³²⁴ Yet evaluations sometimes presented conflicting findings, were phrased in dense bureaucratese, or based findings on bad data—thus failing to provide a clear answer to a very simple question: “Is the project working?”
- **State:** Policy encourages “incorporating regular reviews of progress” that are intended to help program teams “test theory of change hypotheses” and “modify or eliminate what is not working.”¹³²⁵ However, some periodic reviews seemed perfunctory.¹³²⁶
- **DOD:** Operation assessments are supposed to determine whether “objectives (strategic and intermediate) [are] achievable given changes in the [operating environment] and emerging diplomatic/political issues,” and to determine whether “the current plan [is] still suitable to achieve the objectives.”¹³²⁷

However, as one expert put it, “the theoretical utility of operations assessment is rarely realized in practice.”¹³²⁸

Without meaningful implementation, M&E is unlikely to achieve its full potential. 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.¹³²⁹ For example, a senior civil affairs officer said his division staff would regularly tell the Commander of the International Security Assistance Force that “[The] Commander’s Emergency Response Program is a terrible development tool, but it’s a great stabilization tool, *but we never knew if it was true* [emphasis added].”¹³³⁰ As our analysis of USAID’s current strategy demonstrates, the evidence base justifying entire portfolios of contracted projects can be relatively shaky.

Despite uncertainty about what worked, impact was rarely assessed. At times, it was simply assumed. In part, this was because M&E systems were not always well positioned to validate the fundamental theories and assumptions tying projects to broader goals. Instead, M&E functioned as a “work-maker” rather than as a legitimate process capable of determining what worked and what did not. In Afghanistan and elsewhere, the answer to that fundamental question depends heavily on whether the theories and assumptions guiding programming are valid. M&E must therefore continually reassess that validity.

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

This is a corollary of Lesson 3. 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.

Two prominent permutations of irrelevant metrics emerge from the Afghanistan case. First, output indicators such as the number of project beneficiaries who were trained, can be met without actually achieving project objectives. In this example, if training is intended to increase the capacity of institutions whose officials receive the training, that objective is unlikely to be achieved if the quality of training is low. It follows that measuring only against output indicators creates the risk of assessing effectiveness against the wrong metrics.

Second, even metrics that do capture outcomes are only relevant if those outcomes are closely tied to program objectives. For instance, an alternative development program that seeks to reduce opium poppy cultivation by providing beneficiaries with high-value, economically competitive crops may increase beneficiary income. The increase in income is an outcome. But because the program's core objective is to reduce opium poppy cultivation, measuring only changes in household income is insufficient. It was often assumed that once a farmer planted orchards or vineyards, these crops would result in a lasting shift away from opium poppy.¹³³¹ But that assumption was not always valid.¹³³² The result was that programs that did not directly measure changes in poppy cultivation failed to test their fundamental theory of change.¹³³³ This can be a fatal M&E flaw.

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. Such pressure sometimes superseded planning requirements, resulting in implementation of projects that were unlikely to succeed. As one senior State official said, "The political imperative . . . often can dictate the strategic planning of issues rather than evidence that shows a certain strategy or program can be more productive than others."¹³³⁴ At least two programmatic categories—stabilization and counternarcotics—appear to have failed entirely.¹³³⁵

Yet, although goals were not always achievable or objectives realistic, claims of success, progress, or a high probability of either remained common: A corrections system program was contributing to Afghanistan's stability, power-sector projects were contributing to the counterinsurgency effort, the Commander's Emergency Response Program was a vital tool, trade shows were increasing Afghanistan's exports, or a new strategy and more resources would make success achievable. Such claims were often advanced without sufficient evidence, even as many available metrics painted a very different picture. This is a clear disconnect. Where real or perceived pressure incentivizes selective or inaccurate reporting, it can render meaningful M&E very difficult.

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.¹³³⁶

Yet the veneer of quantifiable precision can obscure important nuance or qualification. School enrollment figures are misleading when they are conflated with attendance rates, which are much lower.¹³³⁷ Assertions of seemingly unprecedented improvements to life expectancy and maternal mortality turned out to be based on data that was faulty, potentially anecdotal, or that otherwise lacked proper qualification.¹³³⁸ More importantly, 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 Afghanistan. Those objectives have included increasing the legitimacy of the Afghan government, a goal that health and education indicators alone cannot capture.¹³³⁹

Measurability can also determine which metrics are prioritized. Such decisions are not always deliberate, particularly if a measurable indicator is already being tracked. The most infamous example of this is the burn rate—the rate at which money is being spent—as an indicator of program effectiveness, despite the fact that money spent is merely an input and, on its own, has little assessment value. As one expert explained, “Comptrollers track burn rate as a matter of course, so in the utter absence of any other performance metric, burn rate became the only default.”¹³⁴⁰ In short, measurability is not always a good proxy for efficacy.

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

The use of anecdotes is a widespread technique that is not unique to Afghanistan reconstruction. 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. For example, especially in the area of stabilization, there emerged some “darling districts” (such as Nawa) that were held up as symbols of success.¹³⁴¹ Yet less anecdotal and more representative evidence would later show that stabilization programming turned out to be relatively unsuccessful. As one former USAID staffer summarized, “Anecdotal evidence and stories crowded out more sober, systematic evaluation efforts.”¹³⁴²

The experience of Afghanistan reconstruction shows that anecdotes presented as evidence of success deserve careful scrutiny, particularly where there is pressure to demonstrate progress. 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 explicit political objectives—in particular, “stability.” For example, USAID’s Plan for Transition, which guided the agency’s efforts from 2015–2018, was based on the theory that USAID-funded development projects would help the Afghan government better deliver key services, which, “along with security and the equitable rule of law, will ultimately generate increased confidence in the legitimacy and effectiveness of the government and, in turn, will foster stability.”¹³⁴³ Currently, even as reconstruction shifts to a new phase, stability remains a core aspect of USAID’s intervention logic.¹³⁴⁴ Yet many programs and projects were not assessed against this ultimately political objective, raising questions about whether they actually contributed to it. In sum, where programs or projects aim to achieve political effects, progress against those effects must be measured.

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

Virtually every report written on reconstruction contracting has pointed out that contract oversight personnel are overworked, overburdened, or too few in number.¹³⁴⁵ One contributing factor is high spending levels that outpaced the growth of oversight personnel. The imbalance persists in part because it has been politically easier to increase spending on programs than to increase the number of oversight personnel, particularly those posted at high-risk missions like Afghanistan.¹³⁴⁶ The problem was reportedly so bad that at one point, USAID’s Director of the Office of Acquisition and Assistance determined that, in order to meet the U.S. government’s average ratio of dollars to contracting officers, USAID would have to send nearly its entire overseas workforce to work only in Afghanistan.¹³⁴⁷

SIGAR’s research for this report confirmed that this issue remains a core feature of Afghanistan’s reconstruction contracting landscape. For example, according to Combined Security Transition Command-Afghanistan staff, one significant risk to DOD’s train, advise, and assist contract with DynCorp was “not enough contracting officer representatives to get out and do quality checks.”¹³⁴⁸ Similarly, an official from State’s Bureau of International Narcotics and Law Enforcement Affairs described staffing for contracting officer’s representative positions as a “leaky glass.”¹³⁴⁹ Other reports have also highlighted such challenges, but it is worth underscoring their persistence as a key lesson.

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.¹³⁵⁰ At least two programmatic categories—stabilization and counternarcotics—appear to have failed entirely.¹³⁵¹ 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, we make 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. Numerous SIGAR reports have underscored the extent to which agencies have made claims of success that were either exaggerated, improperly qualified, or based on suspect data.

For example, in 2016, SIGAR expressed concerns about USAID's presentation of Afghan student enrollment, which USAID used as evidence of overall progress in the sector.¹³⁵² As SIGAR continues to report, these figures are misleading. Afghanistan's Ministry of Education counts students who have been absent for up to three years as enrolled because, it says, they might eventually return to school. This means attendance rates are far lower than enrollment rates.¹³⁵³ A 2017 SIGAR audit identified similar problems surrounding USAID's efforts in Afghanistan's health sector.¹³⁵⁴

These audits are part of a broader body of SIGAR work documenting a lack of sound performance measurement in every major reconstruction sector. Additional reports include SIGAR's 2015 audit of DOD, State, and USAID efforts to develop rule of law in Afghanistan, SIGAR's 2018 audit of Promote (USAID's largest single investment to advance women globally), and U.S. efforts to develop Afghanistan's power sector—all of which point to unclear outcomes or impact, inadequate assessment of effects, or the use of incomplete or faulty data.¹³⁵⁵ Taken as a whole, SIGAR's body of work raises serious questions about many claims of success advanced by agencies.¹³⁵⁶

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.**

The United States has spent two decades and \$144 billion attempting to reconstruct Afghanistan.¹³⁵⁷ This extraordinary investment of time and resources reflects perceptions of the importance of the mission. It also demands a serious analysis of which programs and projects worked and which did not.

SIGAR has tried in the past to catalyze just this sort of high-level M&E exercise. In March 2013, SIGAR requested that DOD, State, and USAID list their 10 most and least successful reconstruction projects or programs and provide evaluation criteria for their choices. The request remains unfulfilled after more than eight years.¹³⁵⁸ Through this report, SIGAR is renewing its request.

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.

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

Randomized Controlled Trial and Quasi-experimental Design

A **Randomized Controlled Trial** is an experiment in which individuals are randomly assigned to a control group, which does not receive an intervention, or to a treatment group, which receives the intervention.¹³⁶⁰ In such an experiment, the only possible cause for outcome differences between the treatment and control groups is the treatment itself, or in this case, the program or project.¹³⁶¹

A **Quasi-Experimental Design**, also known as a natural experiment, is an opportunity to study the effects of an intervention in an already occurring scenario.¹³⁶² Randomness is introduced by some prior event or intervention that affected one group and not the other, which has the effect of assigning different treatments to different individuals ‘as if’ the treatment is randomly assigned.¹³⁶³ Quasi-experiments are a practical alternative to randomized controlled trials, which are impossible to execute in most research environments.¹³⁶⁴

desired effects? In particular, DOD, State, and USAID should consider more regularly implementing **randomized controlled trials** or conducting studies that use a **quasi-experimental design**.

By design, randomized controlled trials—which try to tease out impact by assessing the effects of an intervention on a “treatment” group, compared to a “control” group that does not receive the intervention—can credibly show impact. Randomized controlled trials are not always feasible: They were difficult to implement in Afghanistan because of insufficient resources and the heightened complexity they entail, particularly in insecure environments. There were also arguably ethical constraints: By necessity, a control group consisting of non-beneficiaries does not receive an intervention, such as access to improved healthcare, that could better their lives. However, where such constraints can be overcome, and if they are implemented properly, randomized controlled trials are arguably the most effective way to evaluate whether an intervention is actually achieving its desired impact. The study of the National Solidarity Program described in Chapter 5 is perhaps the most prominent example of a randomized controlled trial conducted in Afghanistan.

Where those are not feasible, agencies would do well to use quasi-experimental designs to evaluate impact. Opportunities for quasi-experimental impact evaluations arise when interventions unintentionally create *effective* treatment and control groups. For example, Mercy Corps evaluated the stability effects of a technical vocational education and training program in Helmand Province by administering surveys to a treatment group consisting of recent graduates of the program and a control group consisting of incoming students who had not yet begun classes.¹³⁵⁹ USAID’s Measuring Impacts of Stabilization Initiatives also used a quasi-experimental design: While USAID did not program in villages randomly,

MISTI created a control group by identifying comparison villages whose key characteristics matched those of treatment villages.¹³⁶⁵

Agencies should be on the lookout for other opportunities to use these kinds of experimental or quasi-experimental designs to evaluate the impact of current and future programs.

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.

As detailed earlier in this chapter, 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.¹³⁶⁶ A 2012 SIGAR audit of the Local Governance and Community Development project is illustrative. The audit noted the volume of administrative paperwork expected to be performed by a single contracting officer's representative: certifying that contractor services had been provided, submitting payment forms to USAID, and reviewing invoices and their often voluminous supporting documentation. In theory, certifying services required travel to project sites. For the Local Governance project, there were 608 subprojects in 22 provinces, with a typical invoice covering up to 210 subprojects. The sheer number of subprojects made it impossible for the contracting officer's representative to visit all subprojects within the periods covered by the invoices. In addition to the extensive invoice verification, the contracting officer's representative was also responsible for performance monitoring.¹³⁶⁷

At the time of SIGAR's audit, the results of the \$398 million Local Governance and Community Development project were uncertain, in part because USAID struggled to clearly define expected outcomes.¹³⁶⁸ Meanwhile, a 2011 USAID presentation called into question the project's underlying assumptions, referring to them as "myths."¹³⁶⁹ Although spending significant amounts of time meeting administrative requirements can help prevent fraud, money can also be wasted if the basic assumptions underlying a program or project are weak.¹³⁷⁰

This raises 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.¹³⁷¹ However, key assumptions presented in USAID's Country Development Cooperation Strategy for Afghanistan are not always supported by sufficient evidence. For example, the strategy claims that "the struggle to find a job can lead to economic marginalization and instability, which, in turn, makes one susceptible to violent extremism."¹³⁷² This quote is cited to a 2017 Brookings Institution study.¹³⁷³ As a source for the statement made in the strategy, the Brookings study is problematic for several reasons. Perhaps most prominently, the study's own literature review acknowledges that "previous empirical work has failed to demonstrate any link between unemployment and radicalization."¹³⁷⁴ Other issues 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.”¹³⁷⁵

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.

USAID policy states that the agency “has a responsibility to monitor progress, performance, and operational context of the implementation of [Country Development Cooperation Strategies].” M&E, the policy states, should be “adequate to facilitate strategic learning.”¹³⁷⁶

Two of USAID’s M&E processes have the potential to contribute to learning at a strategic level: a strategy’s annual portfolio review and its midcourse stocktaking exercise. According to USAID policy, “Portfolio reviews are opportunities for missions . . . to periodically examine all aspects of the mission’s strategy, projects, or activities. Missions must conduct at least one portfolio review per year that focuses on progress toward strategy-level results.”¹³⁷⁷ Substituting for the annual portfolio review at least once during the implementation of a country-level strategy, a midcourse stocktaking exercise is performed “with the objective of better aligning the implementation of the mission’s programs with changes to the context and agency direction, as well as with emerging knowledge and lessons learned.”¹³⁷⁸ These are strong processes. However, in implementation, they are unlikely to realize their full potential value for learning and adaptation if the conclusions emerging from these reviews are unclear, equivocal, or otherwise lack credibility.

For example, the phrasing of a mid-course stocktaking exercise examining USAID’s 2015–2018 strategy for Afghanistan proved difficult to navigate. According to the document produced by the exercise, “Key conditions of the development hypothesis [the theory about which specific actions would help achieve strategic goals] did not prove to be valid, and most critical assumptions that underpinned it have not held true throughout implementation.”¹³⁷⁹ Yet the same document said that key components of the strategy’s results framework—defined by USAID policy as “representing the development hypotheses underlying the mission’s strategy”—“*are still valid*” (emphasis in original source).¹³⁸⁰

The assessors attributed “this somewhat contradictory finding” to “some modifications” to the results framework.¹³⁸¹ However, they added, “some senior USAID officials . . . questioned the validity of the framework goal, ‘Afghan-led, sustainable development,’ noting that in mid-2017 it seems to be ‘aspirational.’”¹³⁸² But according to USAID’s current policy, “the goal should be practical.”¹³⁸³ Given the serious questions about the achievability of USAID’s strategic goal expressed by some senior USAID officials at the time the stocktaking exercise was conducted, it

was difficult to understand the assessors' insistence on the supposed validity of the strategy's results framework.

USAID's current Country Development Cooperation Strategy for Afghanistan presents an interesting situation. Shortly after the strategy was finalized and the mission director who signed off on the strategy retired, the now-former mission director immediately undercut the strategy in an op-ed that argued that it was "time to leave" Afghanistan and added that, despite high hopes for "the use of tons of development dollars . . . the reality is little has changed."¹³⁸⁴

The former director's op-ed underscored the importance of conducting a thorough review of USAID's current strategy (essentially an entire portfolio of contracts, grants, and cooperative agreements). SIGAR recommends that USAID conform closely to both the letter and spirit of its M&E policies to ensure that strategic reviews fully embrace bad news.

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."¹³⁸⁵ However, there are instances of evaluations which have seemingly contradictory findings and conclusions, or which minimize significant qualifications.

For example, in March 2017, USAID conducted a midterm evaluation of its \$142.0 million Power Transmission Expansion and Connectivity Commercialization Activities project, which aimed to turn Afghanistan's national power utility into a more commercially viable business entity.¹³⁸⁶ In its summary of findings, the evaluators, Checchi and Company Consulting, Inc., stated, "Overall, the team found . . . that achievement of the component's intended results was mixed."¹³⁸⁷ Later, the assessors stated, "Overall, the PTEC's Commercialization component *did not achieve* its intended results, including in reducing power losses, increasing revenues, and creating jobs" [emphasis added].¹³⁸⁸ Although it is possible to achieve some results and not others, *overall* results cannot be both mixed and unachieved.

USAID has good, clear criteria to help ensure that evaluation reports are of high quality. Those criteria should be more consistently followed. In sum, every evaluation should clearly, accurately, and succinctly answer a basic question: "Is the project working?"

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.**

Among other objectives, State’s Program Design and Performance Management Toolkit aims to “help bureaus, offices, and posts design their activities in a way that effectively advances broader strategic goals.”¹³⁸⁹ To foster analysis and learning, the toolkit encourages “incorporating regular reviews of progress,” which can help program teams to “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.”¹³⁹⁰ 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.¹³⁹¹

In practice, however, reviews do not always fulfill these objectives. For example, the Corrections System Support Program’s 2015 annual review equates “contribution to stability and confidence” with program activities, including advising that the Afghan government transfer “national security threat prisoners” to different prisons.¹³⁹² While such a step could theoretically contribute to stability, providing advice—an input—is conflated with impact (“contribution”).¹³⁹³ Additionally, the program’s theory of change hypotheses were left unaddressed in the annual review, despite the toolkit’s explicit emphasis on this aspect of analysis and learning.

These kinds of errors occur when M&E is purely programmatic and does not explicitly focus on broader questions of relevance and achievability, or the sufficiency of the existing evidence base to justify theories of change. Key processes must be faithfully followed for M&E systems to realize their full potential. 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 that evaluations

“should be ‘evidence based.’”¹³⁹⁴ However, adherence to the written policy is uneven, and there are cases in which evaluations are less useful or rigorous than required by policy.

Illustratively, a 2014 midterm evaluation of State’s \$47.7 million Justice Training Transition Program, which aimed to improve the legal knowledge and skills of Afghan justice-sector officials, exhibited key flaws.¹³⁹⁵ For example, some of the evaluation’s conclusions and recommendations were puzzling. On one hand, the evaluators concluded that the training program was designed in a way that adequately addressed the needs of the Afghan justice sector, with “needs” sweepingly defined as the “formation of [a] competent rule of law system needed for achieving peace and security, poverty reduction, economic growth, and the protection of fundamental rights in Afghanistan.”¹³⁹⁶ On the other, the evaluators noted that key conditions like “political commitment to improve [the] justice sector system” were omitted from the program’s logic model (which described the relationship between program activities and outputs with presumed outcomes and impact).¹³⁹⁷ These conflicting observations raised questions about whether the program was in fact adequately designed, as claimed by the evaluators.

The methodological rigor of the evaluation was equally tenuous. Despite pointing out significant omissions in the training program’s logic model and never establishing that the program was likely to impact the end-goal of restoring public confidence in the Afghan justice system, the evaluators recommended that the program be continued and extended by six months.¹³⁹⁸

The midterm evaluation of State’s Justice Training Transition Program was not the only evaluation that did not meet the standards of usefulness and methodological rigor described in State’s M&E policy.¹³⁹⁹ 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. DOD’s series of ministerial capacity-building contracts with DynCorp illustrate the problem.

For example, as of August 2018, the cumulative value of DOD’s advising contracts with DynCorp was \$1.62 billion.¹⁴⁰⁰ Yet, despite the high value of the contracts, there were times when DOD seemed to be out of compliance with Federal

Acquisition Regulation standards requiring the use of “measureable performance standards” to judge contractor performance.¹⁴⁰¹ Contracted advisors were not solely responsible for achieving desired outcomes—there were also military and civilian advisors.¹⁴⁰² However, other processes that could have assessed the collective efforts of all DOD advisors were not fully used.¹⁴⁰³

This recommendation echoes prior SIGAR calls for DOD to ensure that security capacity-building contracts are achieving the results for which they aim.¹⁴⁰⁴

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.

In Afghanistan, DOD implemented projects that did not fall under its typical mission. As part of counterinsurgency doctrine, reconstruction and development—digging drainage canals, repairing bridges, buying textbooks for a village school in the name of winning hearts and minds—became part of the overall war effort.

But despite the importance of such efforts, DOD struggled to implement M&E systems that could have measured whether desired effects were being achieved. For example, observing that M&E of projects supported by the DOD’s Task Force for Business and Stability Operations was poor, SIGAR recommended that any future task force-like entity develop monitoring and evaluation plans for all projects so that their impact could be assessed.¹⁴⁰⁵

In another example, DOD invested more than \$450 million to expand Afghanistan’s Northeast Power System and its Southeast Power System.¹⁴⁰⁶ Financed through the Afghanistan Infrastructure Fund, the projects primarily involved the rehabilitation and construction of electrical transmission lines and substations.¹⁴⁰⁷ During the high-water mark of the Afghanistan intervention (from 2010 to 2012), DOD used a counterinsurgency justification for these efforts: increased access to electricity would bolster the Afghan’s population’s confidence in the Afghan government, the reasoning went, thereby reducing support to the Taliban.¹⁴⁰⁸

Although DOD monitored construction progress, which was eventually completed, DOD did not collect or report strategic-level performance data for its projects because it had no requirements to do so. Nevertheless, DOD claimed that AIF projects advanced counterinsurgency goals via socioeconomic development. Although USAID eventually initiated an evaluation to assess the impact of U.S. efforts to increase access to electric power, the extent to which the evaluation will thoroughly examine longer-term counterinsurgency impacts is unclear.¹⁴⁰⁹

Important questions remain unanswered because there was no intermediate assessment. For example: Can projects reap counterinsurgency benefits upon

announcement or during implementation? Or: Do delays in project completion have adverse effects on counterinsurgency outcomes?¹⁴¹⁰ In the future, DOD should ensure that impacts are assessed—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.**

Intended as a “feedback mechanism throughout campaign planning and execution,” 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.¹⁴¹¹ When conducted at the campaign level, it is the U.S. military’s highest-order M&E-like mechanism for assessing progress. Although overtly unrelated to contracting, contracted components of reconstruction were inextricably linked to U.S. military objectives in Afghanistan, where military operations were combined with security-force capacity building, and development and governance initiatives in order to counter the insurgency.¹⁴¹²

A key flaw of operation assessment was a tendency towards overoptimism. As one author put it, “Each new general in Afghanistan from 2002 to 2011 suggested that the situation he had inherited was dismal; implied that this was because his predecessor had had the wrong resources or strategy; and asserted that he now had the resources, strategy and leadership to deliver a decisive year.”¹⁴¹³ But that decisive year remained elusive.

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.¹⁴¹⁴ 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.¹⁴¹⁵ 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.¹⁴¹⁶

Matters for Consideration for 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. An increased emphasis on impact evaluation could improve the efficacy of reconstruction efforts in Afghanistan. Because the strategic logic of intervention in other contexts where the United States aims to promote stability is often very similar to that of Afghanistan, the Congress may also wish to consider providing funds to conduct impact evaluations of other U.S. foreign assistance programs, including security-sector assistance programs, implemented around the world. In lieu of appropriating additional funds, the Congress could mandate that a certain proportion of funds appropriated to these agencies be used for impact evaluation.¹⁴¹⁷



USAID photo

APPENDICES & ENDNOTES

APPENDIX A: METHODOLOGY

SIGAR conducts its lessons learned program under the authority of Public Law 110-181 and the Inspector General Act of 1978, as amended. This report was completed in accordance with the Council of the Inspectors General on Integrity and Efficiency’s Quality Standards for Federal Offices of Inspector General (commonly referred to as “the Blue Book”). These standards require that we carry out our work with integrity, objectivity, and independence, and provide information that is factually accurate and reliable. SIGAR’s lessons learned reports are broad in scope and based on a wide range of source material. To achieve the goal of high quality and to help ensure our reports are factually accurate and reliable, the reports are subject to extensive internal and external review, including by relevant U.S. government agencies.

The M&E of contracting research team drew upon a wide array of sources. Some of the team’s documentary research focused on material provided by USAID, State, and DOD. All of this material is unclassified, though some is not publicly available. These agencies provided documents to SIGAR in response to requests for information specific to this lessons learned report. Many additional documents and data were provided in response

to information requests for other SIGAR products, including quarterly reports and audits. SIGAR downloaded certain USAID documents from Afghan Info, the agency's central repository for information relating to programs and projects implemented in Afghanistan, and to which USAID granted SIGAR access. SIGAR also consulted narrative responses submitted by agencies in reply to specific questions, both for this report and for the quarterly report, and used publicly available material from department and agency websites.

These official sources were complemented by hundreds of additional publicly available sources, both governmental and nongovernmental. These included books, think tank reports, press reports, academic studies, audits conducted by the State and USAID inspectors general, and congressional reports, including those produced by congressional commissions established over the years to examine various aspects of defense and development contracting.

Finally, the team also drew from SIGAR's own work, embodied in audits, quarterly reports to the Congress, and prior lessons learned reports. Much of SIGAR's prior work has touched either directly or tangentially on M&E and contracting. This report both distills and builds on that previous work.

To both supplement and bring to life the documentary research, the team interviewed more than 100 individuals with experience applying M&E to defense and development contracts, grants, and cooperative agreements. Interviews were conducted with both government entities and contractors, including U.S. military and civilian personnel at USAID, State, and DOD who have been directly involved in contract oversight and program management in Afghanistan; contractors who directly implemented programs and projects; and European civilian officials.

Interviews provided valuable insights into the challenges, limitations, and frustrations of conducting oversight and implementing M&E in environments like Afghanistan. Due in part to the potentially sensitive nature of sometimes very candid comments, many of the interviewees wished to remain anonymous. Protecting confidentiality for both U.S. government personnel and implementing partners was particularly important to us. To preserve anonymity, our interviews often cite some permutation of an "official" affiliated with a particular agency, an "implementing partner staff member," or a "contractor." In some but not all cases, we include an additional association in the citation, such as the bureau in which a current U.S. official was working at the time of the interview, or the name of the private-sector organization where a contractor worked. We conducted in-person interviews in Kabul and Washington, DC. We also conducted several interviews by phone.

Conducting interviews on a topic like monitoring and evaluation sometimes presents unique challenges. The field of M&E experts is relatively narrow and, as this report demonstrates, M&E systems themselves are sometimes byzantine and difficult to comprehend. Few people have expertise in the M&E policies and practices of every

agency covered by this report. Moreover, some officials, contractors, and subject-matter experts we interviewed may have an interest in selectively presenting facts. To the best of our ability, we sought to control for possible biases and knowledge gaps by triangulating information, either through information obtained through other interviews, through our own examination of program and project documents, or through other reports on the programs, projects, and concepts explored here.

The report underwent an extensive review process. In particular, we shared the report with various subject-matter experts in SIGAR's lessons learned, research and analysis, and audits and inspections directorates. Additionally, USAID, State, and DOD were given an opportunity to formally review and comment on a draft of the report. Department and agency officials provided both technical and substantive feedback. We incorporated their comments where appropriate. However, the analysis, conclusions, and recommendations of this report remain SIGAR's own.

APPENDIX B: LEGISLATION, AUTHORITIES, AND POLICIES GOVERNING CONTRACTING AND M&E

This technical appendix briefly describes the legislative and regulatory landscape for federal contracting and discusses DOD, State, and USAID M&E policies in detail. The information presented here provides a useful benchmark against which the actual practice of M&E is assessed throughout this report.

THE GUIDING LEGISLATION FOR CONTRACTING AND M&E IS A MIXTURE OF BOTH OVERARCHING AND ONE-OFF LAWS

The ability of the U.S. government, the world's largest buyer of goods and services, to enter into contracts is ultimately derived from the Constitution.¹⁴¹⁸ Today, federal contracting practices are governed by a considerable body of laws and regulations. Annual congressional **authorizing** and **appropriating** legislation, including the National Defense Authorization Act, contains a range of directives on both contracting and M&E. Some appropriations legislation, for example, stipulates that a certain amount of funding be designated for M&E activities.¹⁴¹⁹ In other cases, annual appropriations acts contain ad hoc restrictions on, or reporting requirements for, how foreign-assistance funds are spent, including in Afghanistan. Some prohibitions on spending have implicit rather than overt implications for M&E. For example, the 2017 Consolidated Appropriations Act prohibited funding for any project or activity that “is not regularly accessible for the purposes of conducting effective oversight,” or that “legitimizes the Taliban or other extremist organizations.”¹⁴²⁰ The implication was that agencies needed to have monitoring systems in place capable of detecting whether projects supported by appropriated funds were in fact legitimizing insurgent groups in Afghanistan.

Aside from authorizing and appropriating legislation, the Congress has enacted other statutes that more comprehensively impact how U.S. agencies monitor and evaluate foreign assistance. (Table 9 on the next page presents select examples of legislation affecting the M&E of contracted activities in Afghanistan.) The most significant of these is The Foreign Aid Transparency and Accountability Act of 2016, which required the Office of Management and Budget to prescribe guidelines for the establishment of goals, performance metrics, and monitoring and evaluation plans for covered U.S. foreign assistance.¹⁴²¹ In January 2018, OMB released a memorandum to “set forth key monitoring and evaluation principles to guide each agency and provide specific direction on content for agencies to include in their own policies.”¹⁴²²

The Congress funds federal agencies through a two-step process. **Authorizing** legislation approves the creation or continuation of an agency or action. Separately enacted **appropriating** legislation provides funds to the agency, program, or activity that has been authorized.

CRS, “Overview of the Authorization-Appropriations Process,” RS20371, November 26, 2012, p. 1.

TABLE 9

SELECT EXAMPLES OF LEGISLATION IMPACTING CONTRACTING AND M&E	
Broad, Overarching Laws with Implications for all Federal Contracting or for Foreign Assistance	
Legislation	Relevance to Contracting or M&E
Federal Grant and Cooperative Agreement Act of 1977	Provided for the government wide standardization of acquisitions and contracting and was also meant to promote a better understanding of federal spending.
Office of Federal Procurement Policy Act Amendments of 1979	Established the Federal Acquisition Regulation, which sets rules for how the federal government acquires goods or services via contracts and other legal agreements.
Foreign Aid Transparency and Accountability Act of 2016	Required the executive (president) to establish measurable goals, performance metrics, and monitoring and evaluation plans to test the effectiveness of foreign assistance.
Global Fragility Act of 2019	As part of a broader, ten-year “global fragility strategy,” instructed agencies to describe the criteria, metrics, and mechanisms for monitoring and evaluation of programs related to the strategy, to report clear metrics for each country or region covered by the strategy, and to biennially report progress made towards achieving specific targets, metrics, and indicators for each priority country and region.
One-Off Laws with Implications Specifically for Afghanistan Reconstruction	
Legislation	Relevance to Contracting or M&E in Afghanistan
National Defense Authorization Act, 2013	Section 846 required DOD to report on the potential risks of using contractors in overseas contingency operations.
National Defense Authorization Act, 2014	Section 831 prohibited contracting with “the enemy” but left the identification of enemy persons or entities to the Secretary of Defense. The Secretary of Defense was directed to update the DOD supplement to the FAR accordingly.
National Defense Authorization Act, 2015	Section 1225 required the Secretary of Defense to submit a semiannual report to the Congress on security and stability in Afghanistan. This included developing performance metrics for programs intended to increase the capability of the Afghan security forces.
Consolidated Appropriations Act, 2017	Prohibited funds appropriated through the act from being spent on any project or activity that “is not regularly accessible for the purposes of conducting oversight,” or that “legitimizes the Taliban or other extremist organizations.”
Consolidated Appropriations Act, 2018	Required the Secretary of State and the Administrator of USAID to confirm to the Congress that projects or activities funded by monies appropriated through the act, and implemented in areas that cannot be monitored, do not legitimize the Taliban or other extremist organizations.

Source: CRS, “The Federal Acquisition Regulation (FAR): Frequently Asked Questions,” R42826, February 3, 2015; The Federal Grant and Cooperative Agreement Act of 1977, Pub. L. 95-224, 31 U.S.C. §§ 6301-6308 (1977); Further Consolidated Appropriations Act for Fiscal Year 2020, Pub. L. No. 116-94, Title V, §§ 504, 506, 508 (2020); National Defense Authorization Act for Fiscal Year 2013, Pub. L. No. 112-239, § 846.c.4a (2012); National Defense Authorization Act for Fiscal Year 2014, Pub. L. No. 113-66, § 831 (2013); National Defense Authorization Act for Fiscal Year 2015, Pub. L. 113-291, § 1225 (2014); Consolidated Appropriations Act for Fiscal Year 2017, Pub. L. No. 115-31, § 7044 (2016); Consolidated Appropriations Act for Fiscal Year 2018, Pub. L. No. 115-1625, § 7044 (2018).

Efforts to increase the efficacy of foreign aid—and to leverage M&E to tease out what kinds of interventions actually work—have not been limited to the Congress. In 2010, President Barack Obama issued a Presidential Policy Directive on Global Development that called for “a more substantial investment of resources in monitoring and evaluation, including with a focus on rigorous and high-quality impact evaluations.”¹⁴²³ The dearth of impact evaluations in Afghanistan is a key theme cutting across every agency examined in this report.

THE FEDERAL ACQUISITION REGULATION SETS THE RULES FOR U.S. GOVERNMENT PROCUREMENT

Below the level of legislation, the most important rules impacting contracting are articulated in the Federal Acquisition Regulation, which applies to all U.S. government executive agencies. The FAR was conceived by the Office of Federal Procurement Policy Act Amendments of 1979, which sought to establish uniform procurement regulations for the U.S. federal government.¹⁴²⁴ Standardization was a response to the perception that federal procurement rules in force at the time were a confusing patchwork. Indeed, the regulations, laws and directives that had proliferated in the wake of the federal government's expansion of the 1960s created "a burdensome mass and maze," according to a 1972 report by the Commission on Government Procurement, created by the Congress to improve the system.¹⁴²⁵

The FAR applies to most goods and services provided to executive branch agencies. As summarized by the Congressional Research Service, depending on the topic, the FAR describes:¹⁴²⁶

1. "the government's basic policy (e.g., small businesses are to be given the 'maximum practicable opportunity' to participate in acquisitions);
2. any requirements agencies must meet (e.g., obtain full and open competition through the use of competitive procedures);
3. any exceptions to the requirements (e.g., when and how agencies may waive a contractor's exclusion); and
4. any required or optional clauses to be included, or incorporated by reference, in the solicitation or contract (e.g., termination for convenience)."

The FAR is granular in some places and broader in others. It lays out in detail the process that a contracting officer must follow in order to properly enter into a contract, outlining standards ranging from required market research to the application of labor laws.¹⁴²⁷ Some sections of the FAR are explicitly relevant to M&E. For example, when contracting for services, there are specific requirements for the kinds of metrics agencies must include in performance work statements (core documents of performance-based contracts that describe the work contractors are supposed to perform). Specifically, agencies must "to the maximum extent practicable . . . describe the work [to be performed by the contractor] in terms of the required results rather than either 'how' the work is to be accomplished or the number of hours to be provided."¹⁴²⁸ In addition, agencies must incorporate "measurable" and "structured" standards against which the performance of a contractor can be assessed.¹⁴²⁹ The FAR, in other words, explicitly underscores the importance of measuring outcomes, with clear metrics, rather than merely tracking activities or outputs. This is a theme to which this report repeatedly returns.¹⁴³⁰

Though it is by many measures quite specific, the FAR also establishes an overarching framework for federal contracting by setting forth certain broad guiding principles.¹⁴³¹ Most significantly, the FAR states that contracts must "fulfill public policy objectives."¹⁴³² Government contracting, in other words, is intended as a non-arbitrary exercise that aims to further public goals—goals that may involve more than simply ensuring that

work is completed on time and to standard. Throughout this report, we examine the implications of this guiding principal for M&E.

M&E POLICIES AT STATE, USAID, AND DOD VARY IN COMPLEXITY AND ELABORATION

Individual agency policies supplement the laws and regulations governing contracting and M&E. Among all U.S. government agencies, USAID is considered to set the highest bar for M&E policies (though, as this report underscores in Chapter 2, USAID's policies were often unevenly applied). By contrast, State has only recently adopted some of the more complex M&E processes that are the hallmark of USAID. DOD, meanwhile, has a tradition of using M&E-like processes to measure warfighting progress.

Ultimately, each agency has its own ways of measuring effectiveness. While all three are working to advance the interests of the United States, they engage in different activities and pursue different strategies for accomplishing objectives—which has obvious implications for measurement. For example, assessing the effectiveness of a project that has economic goals, as USAID does, is a different exercise than assessing the capacity of the Afghan National Defense and Security Forces to conduct operations independent of U.S. advisors. Still, as Table 10 shows, while there is some variation in the definition of key monitoring and evaluation terms, there are also many core similarities.

TABLE 10

WHAT M&E MEANS TO EACH AGENCY			
	USAID	State	DOD
Assessment	"A forward looking process that may be designed to examine country or sector context to inform project design, or an informal review of projects. It is distinct from evaluation."	While the term "assessment" is not listed in State's glossary of terms, the meaning is to some extent implied by State's definition of "evaluation": "An assessment may be designed to examine country or sector context to inform program or project design."	"Systematic analysis to provide an understanding of the context, conditions, partner capabilities, and requirements to inform security cooperation planning and implementation. Assessments are generally conducted in advance of security cooperation activities, but may be repeated to update analysis and identify mid-course corrections of security cooperation activities."
Monitoring	"The ongoing and systematic tracking of information relevant to USAID strategies, projects, and activities."	"An ongoing system of gathering information and tracking performance to assess progress against established goals and objectives."	"A continuous process designed to provide regular feedback on the extent to which expected outputs and outcomes are being achieved to inform decisions or corrective actions. In general, results measured in monitoring are the direct and near-term consequences of initiative activities that provide opportunities to validate the theory of change throughout implementation and an early indication of the likelihood that expected results will be attained."
Evaluation	"The systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or inform decisions about current and future programming. Evaluation is distinct from assessment, which may be designed to examine country or sector context to inform project design, or an informal review of projects."	"The systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or inform decisions about current and future programming. Evaluation is distinct from assessment, which may be designed to examine country or sector context to inform project design, or an informal review of projects."	"A systematic collection and analysis of information and evidence about the characteristics and outcomes of an ongoing or completed initiative, and its design, implementation, and results. Evaluations determine relevance, value, effectiveness, efficiency, sustainability, and impact as a basis for improving effectiveness and to inform decision makers regarding future plans, programs, and activities. Evaluation, distinct from assessment and monitoring, focuses on documenting the achievement of outcomes and results and in some cases the value of continuing the investment."

Source: USAID, "Glossary of ADS Terms," April 18, 2018, pp. 23, 95; CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, pp. 3, 5, 6; DOD, "Instruction 5132.14: Assessment, Monitoring, and Evaluation Policy for the Security Corporate Enterprise," January 13, 2017, pp. 20, 21; USAID, "ADS Chapter 201: Program Cycle Operational Policy," June 11, 2019, p. 14; State, "18 Foreign Affairs Manual 301.4," Program and Project Design, Monitoring and Evaluation, p. 2.

The remainder of this appendix discusses each agency's M&E policies in detail, providing a benchmark that we use to analyze the actual practice of M&E in chapters 2, 3, and 4. These policies also serve as the basis of several recommendations made in the final chapter of this report.

USAID: An Extraordinarily Complex M&E System

USAID refers to M&E as MEL—for “monitoring, evaluation and learning,” activities that the agency explicitly links together.¹⁴³³ USAID's highly elaborated M&E policies are detailed in Automated Directives System, Chapter 201, which describes the way the agency plans and delivers development programming.¹⁴³⁴ (For simplicity's sake, this report generally refers to ADS Chapter 201 as either “USAID policies” or, where more appropriate, “USAID policy.”) Programming is conceived, implemented, and modified through what USAID calls the “program cycle,” which consists of three main stages: strategic planning, design and implementation, and monitoring and evaluation.¹⁴³⁵ While M&E features as a distinct component in the program cycle, feedback from M&E systems is intended to inform all other stages.

The pages that follow explain these stages and describe the ways in which M&E figures into each. It is easy to get lost amid the plethora of technical terms and Byzantine processes that are the hallmarks of USAID's policy. Graphics throughout this section explain key documents, definitions, and procedures with illustrative examples. In this appendix, we use the version of Automated Directives System Chapter 201 dated June 11, 2019 as a reference. Elsewhere in the report, we refer to the July 23, 2020 version of the chapter, as appropriate.

At USAID, M&E Is Organized around Strategy

M&E systems at USAID begin with strategy—specifically, a Country Development Cooperation Strategy, which articulates a **mission's** objectives for a given country and the planned approach for achieving them.¹⁴³⁶ The strategy is mapped out in something called the results framework—a visual representation of a hierarchical sequence of outcomes that are presumed to produce strategic effects.¹⁴³⁷ It describes how intermediate outcomes, if achieved, will result in the attainment of increasingly broader objectives, stretching all the way along an assumed causal hierarchy towards an overarching, country-level goal.¹⁴³⁸ For example, in Afghanistan, USAID assumes that:

- increasing the commercial viability of Afghan businesses and the productivity of key agricultural crops will -
- improve the competitiveness of Afghan value chains, which in turn will -
- accelerate exports and economic growth -
- thereby rendering Afghanistan a more “economically viable” country that is better positioned to be a national security partner of the United States.¹⁴³⁹

Understanding the structure of a results framework is crucial for comprehending how USAID's M&E systems are supposed to work. In general, the closer an outcome is to the level of the framework on which actual implementation occurs, the more

A USAID **mission** is made up of USAID personnel working together within a given assistance-recipient country.

USAID, “Glossary of ADS Terms,” April 18, 2018, p. 165.

accountable USAID holds itself for results. The graphics on pages 202–203 visualize the results framework for USAID’s current Country Development Cooperation Strategy for Afghanistan, provide illustrative M&E requirements at each level of the framework, and summarize some of the outcomes USAID currently hopes to achieve.

Built around the Results Framework, the Performance Management Plan Is a USAID Mission’s Highest-level M&E Tool

Structured around the results framework, the country-level performance management plan is intended to be a mission’s primary tool for managing M&E processes.¹⁴⁴⁰ The plan provides a sweeping view of what a mission aims to accomplish in a given country, reiterates the hierarchy of outcomes presented in the results framework, describes the portfolios of projects and activities that aim to support those outcomes, and explains how M&E will be used to track progress at each level of the hierarchy, including what indicators will be used. In other words, the performance management plan is an overarching, strategic-level M&E document to which programmatic M&E efforts are

FIGURE 9

THE PERFORMANCE MANAGEMENT PLAN IN ACTION: AN EXAMPLE FROM USAID’S 2011–2015 PMP

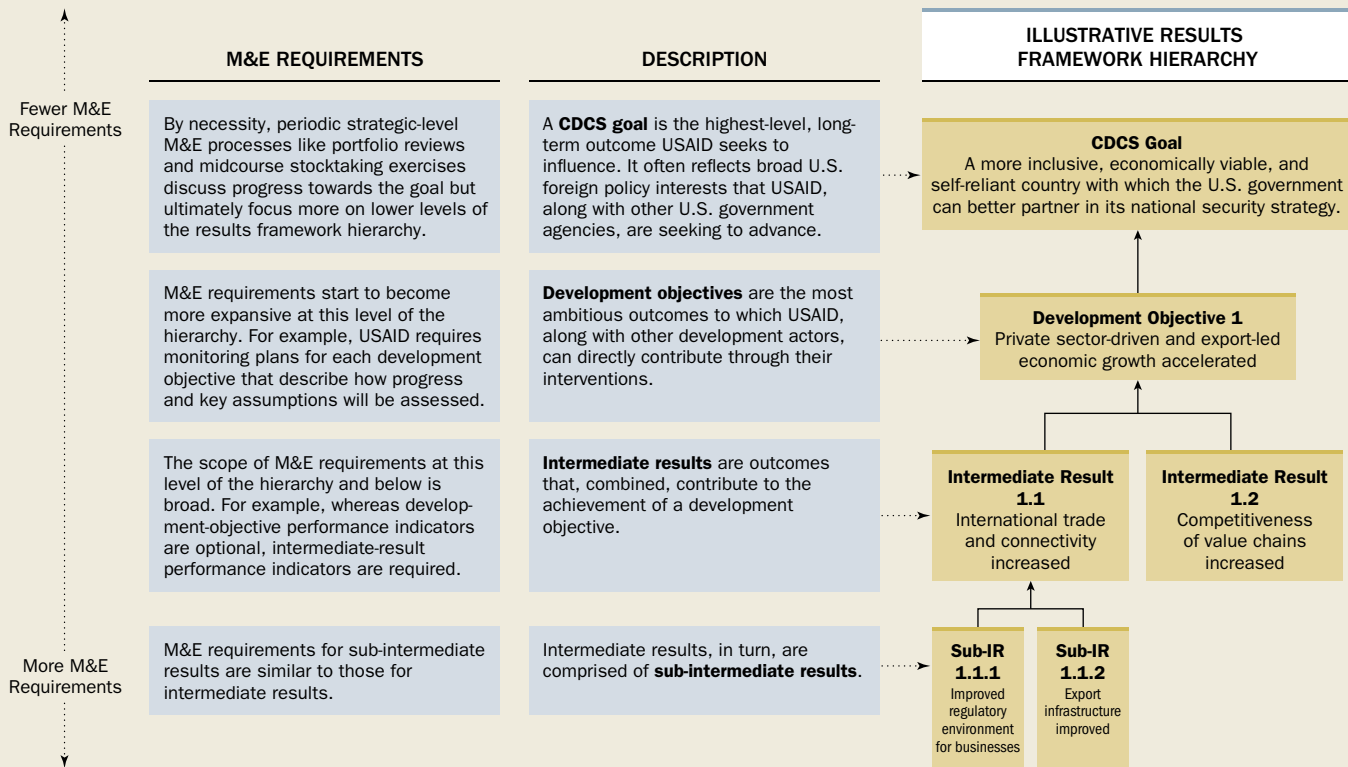
	FUNCTION	EXAMPLES FROM THE PMP
	Articulating what USAID was trying to accomplish	In its opening pages, the 2011–2015 PMP provided an overview of what USAID and its agency counterparts were hoping to accomplish in Afghanistan. In essence, the overall objective was to partner with the Afghan government “in order to build its capacity to provide its people a stable future.” The PMP also included a complex results framework.
	Describing the portfolio of projects intended to achieve outcomes	The PMP described active and planned projects supporting each intermediate outcome presented in the results framework. For example, in order to strengthen “governance and service delivery at [the] national and sub-national levels,” the PMP detailed a USAID program intended to make Afghanistan’s legislature more effective.
	Explaining how M&E will be implemented	The PMP described when activity level data would be collected and consolidated by the mission —“during the first quarter” of every year—in order to meet annual reporting requirements. The PMP also explained who was responsible for collecting this data.
	Describing relevant indicators	The PMP presented the indicators to be used at each level of outcome specified in the results framework. For example, the indicator “number of women holding seats in national parliament” was intended to capture progress towards the broad outcome “improved performance and accountability of governance.” Lower down, the indicator “number of judges trained with [U.S. government] assistance” was intended to help measure the subordinate outcome “formal rule of law system improved.”

Source: USAID, “U.S. Foreign Assistance for Afghanistan Post Performance Management Plan 2011–2015,” October 2, 2010, pp. 2–3, 10, 12; USAID, “U.S. Foreign Assistance for Afghanistan Post Performance Management Plan 2011–2015: Annex I - Assistance Objective 1: Improved Performance and Accountability of Governance,” October 2, 2010, pp. 7, 13.

FIGURE 10

AT USAID, THE COUNTRY-LEVEL RESULTS FRAMEWORK HELPS ORGANIZE M&E

USAID's Country Development Cooperation Strategy is mapped out in something called the **results framework**—a visual representation of outcomes that, combined, are presumed to produce strategic effects. It describes how lower-order outcomes relate and contribute to broader objectives, moving upward towards an overarching, country-level goal. USAID uses specific terms to describe each level of outcome. For example, a **CDCS goal** is the broadest anticipated outcome of a strategy and typically reflects U.S. foreign policy objectives to which USAID, along with other agencies, are expected to contribute. By contrast, at the lowest level of the hierarchy, a **sub-intermediate result** reflects a much more discrete outcome that an individual intervention is intended to produce. In general, M&E requirements are more expansive at lower levels of the results framework hierarchy. The figure below illustrates USAID's results framework using an example from the agency's current Country Development Cooperation Strategy for Afghanistan. The table to the left of the figure describes the type of outcome or impact depicted in each level of the hierarchy and provides example M&E requirements associated with each tier. As this example represents only one branch of the results framework, the graphic on the adjacent page provides a more complete picture.



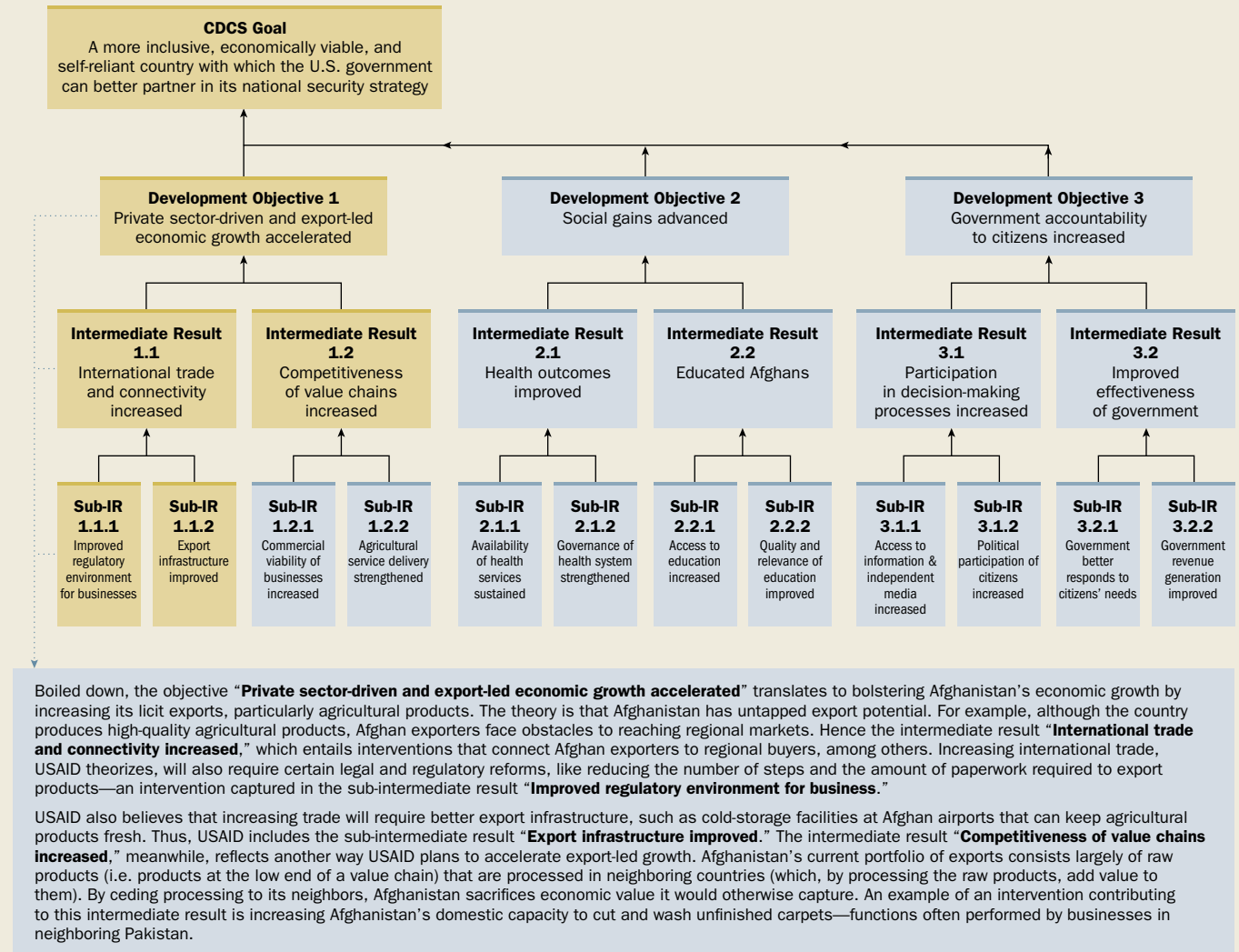
Note: USAID's country-level results framework for Afghanistan is large. Consequently, some intermediate objectives and sub-intermediate results are not depicted.

Source: USAID, "Afghanistan Country Development Cooperation Strategy FY 2019–2023," September, 2018, p. 27–30; State, "Afghanistan: Integrated Country Strategy," September 2018, p. 7; USAID, "Glossary of ADS Terms," April 18, 2018, pp. 40, 76, 139, 235, 262; USAID, OAPA, "Fixed Amount Award No. 72030618FA00006, under Annual Program Statement (APS) Number APS 306-17-000003 and RFA-306-17-0000012 (Establishing Kabul Carpet Export Center (KCEC)," June 6, 2018, p. 28.

FIGURE 11

VISUALIZING USAID'S RESULTS FRAMEWORK

The figure on the left depicts only one branch of the results framework for USAID's current Country Development Cooperation Strategy for Afghanistan. Below is a more complete visualization. Further descriptions of some outcomes and examples of interventions supporting them are presented beneath the results framework.



supposed to connect.¹⁴⁴¹ Figure 9 on page 203 illustrates how USAID’s performance management plan for the 2011–2015 period performed these functions.¹⁴⁴²

USAID Projects are Designed to Support Outcomes Identified in a Strategy

Once a mission has produced a results framework, it must define how it will achieve those results.¹⁴⁴³ This occurs through the project design process, which guides the execution of a strategic approach.¹⁴⁴⁴ Projects, which are generally aligned with a single intermediate result in the results framework, consist of multiple activities. These activities directly carry out an intervention via contracts, grants, or cooperative agreements.¹⁴⁴⁵ The culmination of the project design process is the project appraisal document.¹⁴⁴⁶

The project appraisal document presents a project’s theory of change—a statement that explains why implementing the project will, in theory, lead to desired outcomes—and an associated implementation plan, including descriptions of the project’s activities.¹⁴⁴⁷ It also includes the project’s monitoring, evaluation, and learning plan, which lays important groundwork for future M&E actions and builds on the performance management plan.¹⁴⁴⁸

Activity MEL plans flow from project MEL plans and must be in place before major implementation actions begin.¹⁴⁴⁹ Among other purposes, activity MEL plans describe the monitoring approaches to be applied during implementation and define relevant performance indicators consisting of activity-level outputs and outcomes.¹⁴⁵⁰ Activity MEL plans fulfill many of the same functions as performance management plans, though they are applied to individual awards as opposed to a strategic portfolio of contracts. For example, the MEL plan for USAID’s Initiative for Hygiene, Sanitation, and Nutrition activity, which aimed to expand nutrition and water, sanitation and hygiene interventions in Afghanistan, contained:¹⁴⁵¹

- a narrative describing how the nutrition and sanitation activity was believed to contribute to USAID’s overarching strategy,
- an activity-level results framework that described presumed connections between the activity’s objectives and relevant outcomes and impact, and
- a list of performance indicators to track progress

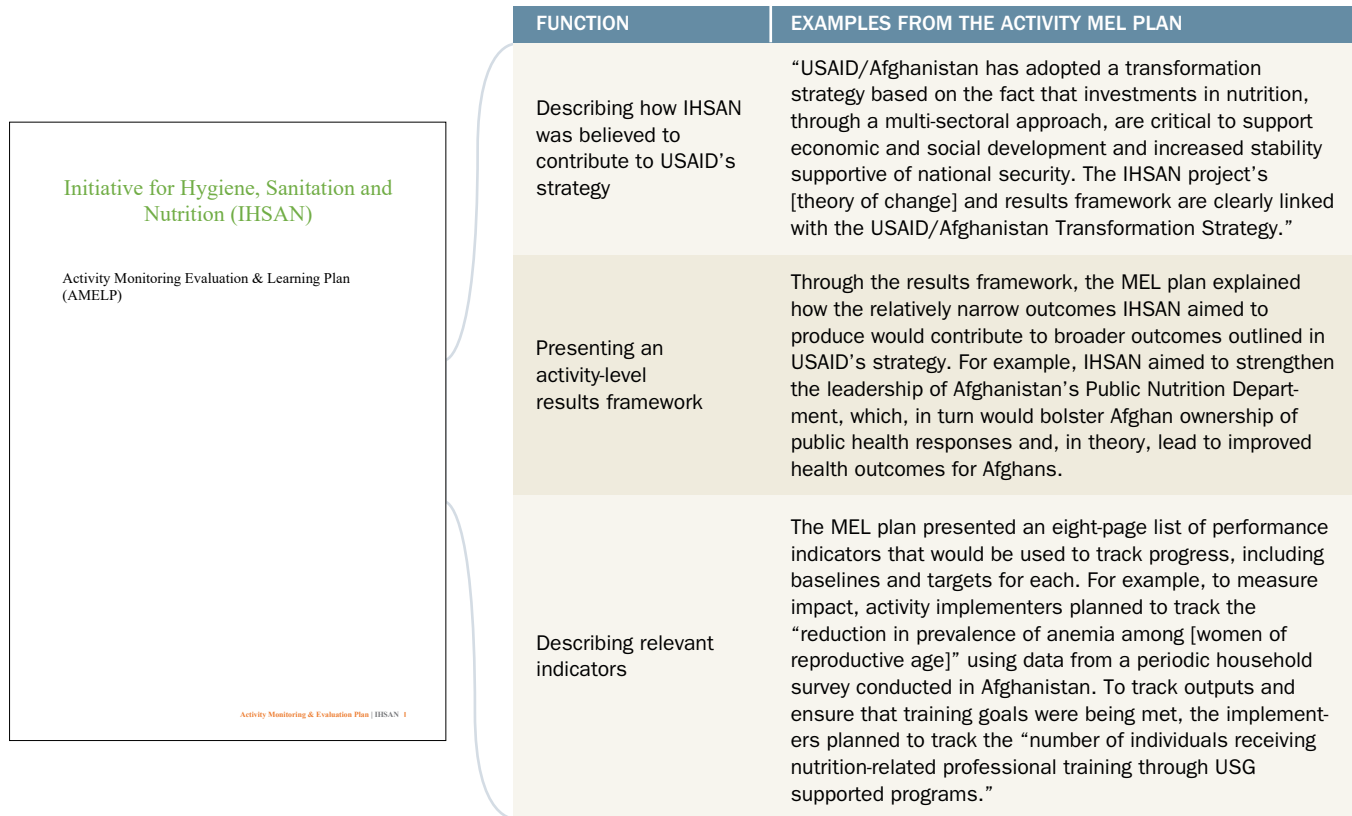
Figure 12 summarizes the various functions of an activity MEL plan, using the Initiative for Hygiene, Sanitation and Nutrition activity as an example.

Definitions, Requirements, and Principles of Monitoring and Evaluation

After strategic planning, project design and implementation, and activity design and implementation, the last major component of the program cycle is monitoring and evaluation—the purpose of which, according to USAID policy, “is to apply knowledge gained from evidence and analysis to improve development outcomes and ensure accountability for the resources used to achieve them.”¹⁴⁵²

FIGURE 12

ILLUSTRATING THE FUNCTIONS OF AN ACTIVITY MEL PLAN



Source: FHI 360, "Initiative for Hygiene, Sanitation and Nutrition (IHSAN) Activity Monitoring Evaluation & Learning Plan (AMELP) FY 2020," November 13, 2019, pp. 10–11, 14, 16.

Monitoring Is Routine Tracking of Progress Towards Targets

Defined as "the ongoing and systematic tracking of data or information relevant to USAID strategies, projects, and activities," monitoring is the bread and butter of USAID's M&E system.¹⁴⁵³ While monitoring may sound like merely collecting information, it is intended to be an active process that involves planning, selecting which data to collect, ensuring that it is reliable, properly collecting and maintaining it, and then analyzing it to make decisions and adapt methods.¹⁴⁵⁴ During implementation, monitoring aims to promote accountability and provide data for evaluations where appropriate.¹⁴⁵⁵

USAID missions conduct two types of monitoring.¹⁴⁵⁶ Performance monitoring collects indicator data and other information that shows whether implementation is on track and whether an activity is achieving expected results; it directly informs judgments about particular approaches.¹⁴⁵⁷ USAID recommends comparing actual results as measured by these indicators to initially expected targets, then comparing that analysis with the assumptions underlying the intervention logic of each level of the results framework.¹⁴⁵⁸

Context monitoring involves collecting information about external conditions that may affect a mission’s strategy, projects, and activities.¹⁴⁵⁹ It, too, is meant to be used to assess the assumptions of the results framework and the intervention logic of projects and activities.¹⁴⁶⁰

Evaluation Focuses on Whether and Why Results Were Achieved

Defined as “the systematic collection and analysis of information about the characteristics and outcomes of strategies, projects, and activities,” evaluation is monitoring’s complement.¹⁴⁶¹ While monitoring is an ongoing tracking process, USAID’s evaluations usually happen at defined intervals, and normally incorporate data that have been collected through monitoring and other sources. However, just as with monitoring, the purpose of evaluation is to inform decisions that will improve program effectiveness and accountability.¹⁴⁶² USAID provides a similar set of principles to guide evaluations: Evaluations should be planned during the design of strategies, projects, and activities; unbiased in measurement and reporting; relevant; based on best methods to generate the most credible evidence; oriented toward reinforcing local ownership; and transparent.¹⁴⁶³ USAID categorizes evaluations into two types: **impact** and **performance**.¹⁴⁶⁴

Types of Evaluation

Impact evaluations measure the change in a development outcome that is directly attributable to an intervention. As changes to outcomes may occur due to reasons other than an intervention, impact evaluations attempt to control for all relevant factors that might also account for observed changes.¹⁴⁶⁵

For example, an observed outcome for a vocational-training intervention may be an increase in employment among beneficiaries. However, looking only at the employment status of beneficiaries would not be sufficient to establish impact. Other factors, such as a growing local economy coincident with the intervention, could also account for the change. Consequently, an impact evaluation would have to define and evaluate a counterfactual—in this case, by determining the change in employment status for a group of non-beneficiaries whose demographics and circumstances resemble those of the beneficiaries. The counterfactual, in other words, attempts to answer the question, “what would have happened to the beneficiaries had they not received the intervention?”

Performance evaluations focus on what a particular program or action has achieved, how it was implemented, whether the expected results occurred, and other questions pertinent to project design, management and decision making.¹⁴⁶⁶ Performance evaluations may address any cause-and-effect, descriptive, or normative questions (for example, how a project is perceived or valued).¹⁴⁶⁷ Performance evaluations “often incorporate before-after comparisons but generally lack a rigorously defined counterfactual.”¹⁴⁶⁸

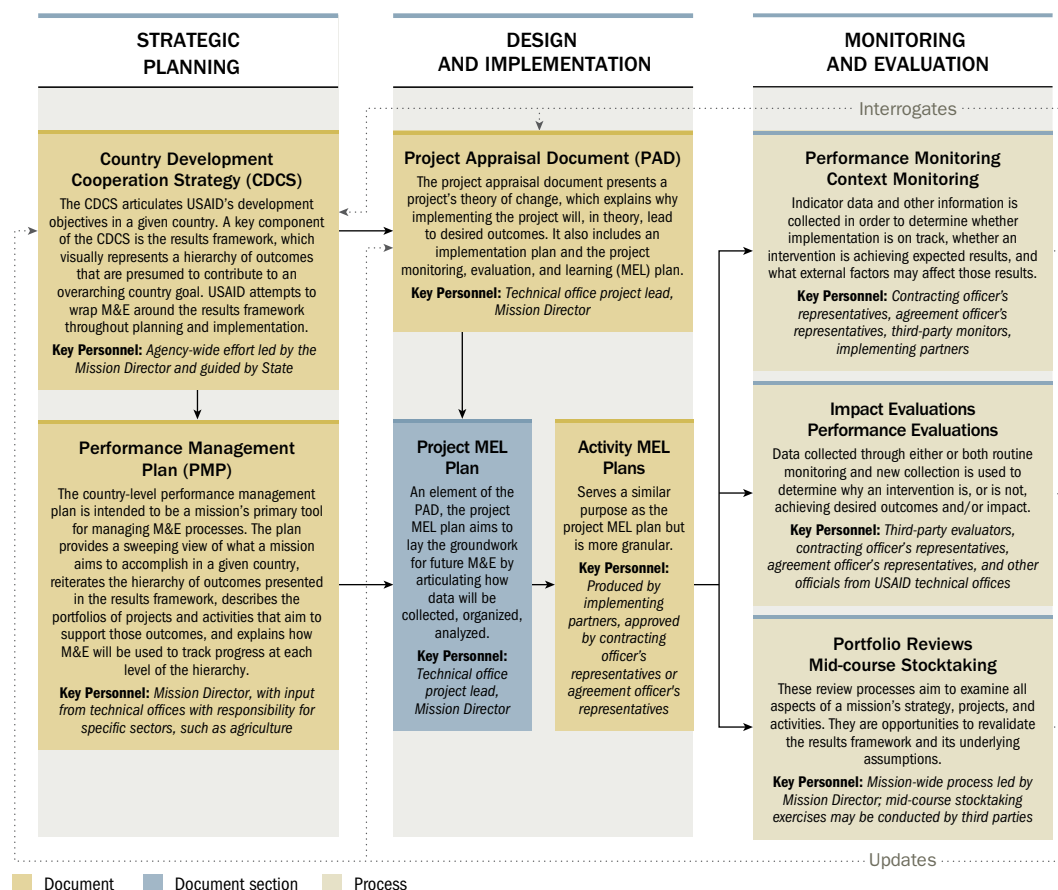
Strategic-Level Evaluation Zooms Out from Individual Projects to Focus on the Bigger Picture

Within USAID, the term “evaluation” generally refers only to evaluations of projects and activities. However, evaluation-like processes periodically occur at the strategic level. The most important are portfolio reviews and mid-course stocktaking exercises.¹⁴⁶⁹

Portfolio reviews give missions an opportunity at least once a year to examine “all aspects of the mission’s strategy, projects, or activities” and review progress toward strategic results.¹⁴⁷⁰ The portfolio review assesses how well projects are progressing toward achieving the objectives of the Country Development Cooperation Strategy, any changes to beginning assumptions, changes in context, what features of projects might need to change, and the status of post-evaluation action plans.¹⁴⁷¹ These reviews aim both to update current strategy, and to inform the development of future strategic documents.¹⁴⁷²

FIGURE 13

M&E IN USAID'S PROGRAM CYCLE: KEY DOCUMENTS AND PROCESSES



Note: As used here, the term “technical office” refers to a group of USAID personnel responsible for implementing and overseeing activities specific to an individual sector, in an individual country. In Afghanistan, USAID’s technical offices include those for health, education, economic growth, infrastructure, and governance, among several others. A contracting officer’s representative or agreement officer’s representative that oversees an individual activity typically comes from the technical office responsible for the activity.

Source: USAID, “ADS Chapter 201: Program Cycle Operational Policy,” partial revision dated June 11, 2019; USAID, “Glossary of ADS Terms,” April 18, 2018, pp. 57, 126, 191, 200, 235.

Mid-course stocktaking exercises, which must happen at least once during the course of a strategy's implementation, are essentially a more expansive version of portfolio reviews.¹⁴⁷³ The most important functions of this exercise are to revalidate the results framework and its underlying assumptions, and to identify learning from monitoring data and evaluations.¹⁴⁷⁴ An additional purpose of the exercise is to provide Washington-based USAID personnel with a comprehensive update on strategic progress to date.¹⁴⁷⁵ In essence, they provide missions with a chance to “pause and reflect” in order to “generate and apply new learning.”¹⁴⁷⁶

Chapter 2 examines causal claims made by USAID in its current Country Development Cooperation Strategy for Afghanistan. The analysis highlights the importance of ensuring the reliability and consistency of the underlying assumptions and evidence base for a strategy implemented mainly by contractors.

Obviously, USAID's M&E processes are highly complex. Figure 13 on the previous page brings various elements of this subsection together, depicting how the parts of USAID's program cycle fit together.

State: Evolving Towards “Managing for Results”

State began to implement its first department-wide policy on evaluation in 2012.¹⁴⁷⁷ Before then, monitoring and evaluation was implemented on a small-scale, ad hoc basis best suited to projects that had already been implemented and tested in other contexts.¹⁴⁷⁸ When funding for Afghanistan reconstruction rapidly increased beginning in 2009, State was compelled to quickly develop procedures for monitoring and evaluating a significantly higher amount of funding.¹⁴⁷⁹ As part of a department-wide effort to improve performance measurement, State issued the Managing for Results framework in 2013.¹⁴⁸⁰ The framework was intended to guide day-to-day performance management, as well as to assist strategic and operational-level decision-making.¹⁴⁸¹

In 2015, the **Bureau of Budget and Planning**, together with the **Office of Foreign Assistance**, published an updated framework to establish objectives against which

Defining Key State Department Entities

In order to submit State's budget requests to the Congress, its **Bureau of Budget and Planning** takes into account both operational needs and program performance. It coordinates with State's **Office of Foreign Assistance** to develop policies, plans, and programs that aim to achieve U.S. foreign policy goals.¹⁴⁸² State's Office of Foreign Assistance has the lead for coordinating all U.S. foreign assistance activities, including policy, planning, and performance management. It also provides strategic direction for both State and USAID.¹⁴⁸³

activities could be measured, and to connect resource requests and activity reviews to those strategic objectives.¹⁴⁸⁴ The framework also included instructions for how to monitor and evaluate progress and results.¹⁴⁸⁵ This instruction tool kit was intended to help program designers and implementers determine what does and does not work effectively, and to incorporate that information into future programs. The tool kit was followed by the 2015 issuance of a second evaluation policy to work “in concert with [State’s] evaluation guidance.”¹⁴⁸⁶ State issued a new monitoring and evaluation policy in 2017. In addition to M&E, the policy covered program and project design. Its primary purpose was to “establish a clear line of sight” from strategic objectives to individual programs, give guidance on how to achieve them, and collect data on whether those efforts were working as intended.¹⁴⁸⁷

The Managing for Results Framework Draws a Line from Programs to Strategy

First developed in 2013, the Managing for Results Framework is intended to connect strategies with the programs supporting them through rigorous monitoring and evaluation practices that improve outcomes and inform decision-making.¹⁴⁸⁸ The goals of the framework are to:

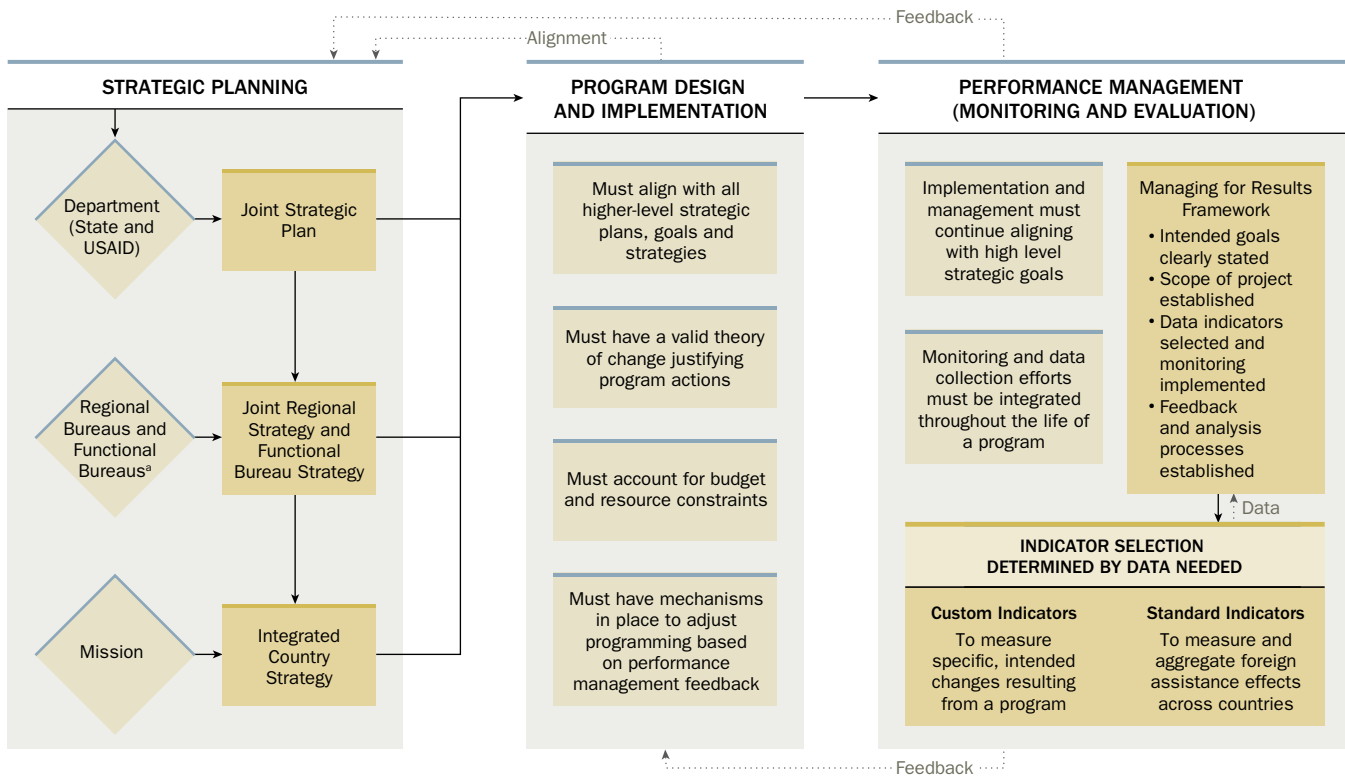
- strengthen the decision-making process on strategic priorities, programming, and budgeting;
- increase transparency by demonstrating how the department spends resources; and
- improve coordination within the department and with external stakeholders.¹⁴⁸⁹

MfR provides the overarching framework for State’s program and project design, as well as monitoring and evaluation policies.¹⁴⁹⁰ During the program design stage, bureaus are required to align programming to strategy, create program goals that advance this strategy, and articulate a specific theory of change describing how program activities will achieve strategic goals. This theory of change is captured in a logic model that links program inputs and activities with desired outputs and outcomes. Cumulatively, these steps are intended to create a foundation against which progress can be monitored and evaluated.¹⁴⁹¹ On the program level, bureaus articulate how they will fulfill MfR’s steps in a performance management plan, which includes descriptions of performance indicators used to track program progress and a framework against which strategic progress can be assessed, among other components.¹⁴⁹²

Overall, the best way to think about State’s M&E system is as a rough analogue to USAID’s, though it is not quite as elaborate. State’s 2012 evaluation policy was modeled on USAID’s 2011 policy.¹⁴⁹³ State’s performance management plans fulfill purposes similar to those of USAID project and activity MEL plans.¹⁴⁹⁴ As State’s “Program Design and Performance Management Toolkit” itself states, “to encourage congruence and a common lexicon, concepts from the USAID Automated Directives System are incorporated where possible.”¹⁴⁹⁵ Figure 14 on the next page summarizes key aspects of State’s program design and performance management cycle.

FIGURE 14

STATE'S PROGRAM DESIGN AND PERFORMANCE MANAGEMENT PROCESS



This is a cyclical process, not a linear one. Step Four feeds back into Step One, but the process often involves all four elements occurring simultaneously.

STEP ONE: STRATEGIC PLANNING

Projects must align with high level strategic goals. Within a four-year Joint Strategic Plan, State sets strategic direction and priorities. Lower-level strategy, including country strategies, must fit into that framework, working all the way down, through embassies, to individual programs.^b This document is focused on macro strategy, not individual program management.

STEP TWO: DESIGN AND IMPLEMENTATION

Given limited resources, projects must be prioritized according to strategic plans. Program design must account for the restrictions of the host country environment. Individual programs must also have an underlying theory of change justifying the resources used.^c Monitoring and evaluation must be integrated throughout program design.

STEP THREE: PROGRAM MANAGEMENT

Program management involves collecting data within the structure of the Managing for Results Framework. Continually monitoring performance indicators allows individual programs to adjust based on shifts in data.

STEP FOUR: LEARNING

Feedback is incorporated back into the strategic planning process. This allows ongoing projects to continually improve and informs future projects.

^a A regional bureau is a State office responsible for a geographic area. The Bureau of South and Central Asian Affairs is the regional bureau that oversees State's work in Afghanistan, among other countries. By contrast, functional bureaus focus on a specific policy area. For example, the Bureau of International Narcotics and Law Enforcement works on issues related to drug trafficking, crime, and criminal justice. Functional and regional bureaus collaborate with each other to achieve broad policy goals.

^b An example of a program is the Corrections System Support Program, which supports the development of a safer, more humane and more transparent prison system in Afghanistan. This aligns with the Integrated Country Strategy objective of "an Afghan government that is more stable . . . and increasingly capable of performing key functions" and the INL Functional Bureau Strategy objective of improving the rule of law capacity of vulnerable countries.

^c In essence, a theory of change is an explanation of why and how a program contributes to broad strategic goals. For example, for the Corrections System Support Program, State theorizes that providing training and advice to Afghan corrections officials will increase the managerial capacity of those officials. This increase in managerial capacity, in conjunction with other program activities like infrastructure improvement, will theoretically result in a "safe, secure, humane, and transparent corrections system." While programs are required by policy to have theories of change, they are not always well-articulated or fully explained.

Source: State, "Proposed Departmental Organizational Chart," February 2020, accessed September 1, 2020; State, "18 Foreign Affairs Manual 301.1," Managing for Results Framework, pp. 1-3; State, "18 Foreign Affairs Manual 301.4," pp. 2, 9; State, "Foreign Assistance Library," accessed November 6, 2020; State, "Program Design and Performance Management Toolkit," pp. 8, 30, 37; State, "Standard Foreign Assistance Indicators," accessed September 1, 2020, p. 1; State, "Diplomacy in Action: Department of State Evaluation Policy," archived content, January 29, 2015, accessed August 31, 2020; USAID, "Performance Monitoring Indicators," accessed September 1, 2020, p. 3; USAID, "Standard Indicators," accessed September 1, 2020, pp. 1-2; State, Bureau of International Narcotics and Law Enforcement Affairs Office of Afghanistan and Pakistan, "Corrections System Support Program Performance Management Plan," June 2017, pp. 5-7; State, "Annual Performance Review for CSSP," January 2019, p. 5.

ROAD TESTING STATE'S PERFORMANCE MANAGEMENT TOOL KIT: THE IMPORTANCE OF CONTINUALLY REASSESSING PROGRAMMATIC ASSUMPTIONS

To demonstrate how it is supposed to work, State's "Program Design and Performance Management Toolkit" walks readers through a hypothetical scenario: A State team is attempting to design a program to reduce air pollution and develop clean energy policies in a fictional country.¹⁴⁹⁶

The hypothetical case study involves a fictitious State employee—Mary—who is leading the design and management of a \$5 million project that aims to help fulfill U.S. government climate-change goals in the country of Freedonia.¹⁴⁹⁷ As the tool kit details, "It is up to Mary and her team to determine how best to use these funds to address climate change," using (as might be expected) "this program design and performance management toolkit to guide the process."¹⁴⁹⁸ On and off through the next 58 pages, Mary dutifully runs through each step of the process. After considerable analysis, Mary designs and implements a program intended to increase the capacity of Freedonia's national government to implement energy regulatory frameworks and to increase the Freedonian public's demand for clean energy.¹⁴⁹⁹ The case study ends with the production of a "strategic progress review framework" designed to assess progress towards program objectives.¹⁵⁰⁰

The technique of using fictional case studies can be helpful. However, this particular one shows how program design can fall victim to faulty premises. In one particular key step—the determination of root causes, performed to ensure the program is "effectively [addressing] the true problem"—the case study seems to suffer from logical fallacies.¹⁵⁰¹

Specifically, Mary performs an analysis called "The Five Whys" in order to track the causal chain from the undesirable symptom her program seeks to address (high rates of greenhouse-gas emissions) down to its "original

cause."¹⁵⁰² In the process, Mary makes an implicit analytical leap that will have significant consequences: After determining that the people of Freedonia "are more concerned about the cost of energy than its source," Mary asks, "Why don't the people of Freedonia care where their energy comes from?"¹⁵⁰³

This question ignores her just-stated conclusion: Freedonians *do* care where their energy comes from. Specifically, they care that it comes from the cheapest source.¹⁵⁰⁴ Mary nevertheless determines the answer to her question is that "the people of Freedonia are unconcerned or unaware of the harmful effects of burning coal or the economic possibilities new businesses could bring to the area," and concludes that a public awareness campaign should be a key aspect of her clean energy program.¹⁵⁰⁵ She has overlooked the very real possibility that Freedonians are well aware of the adverse health effects of using coal, but choose to continue doing so because immediate benefits outweigh long-term costs. Consequently, a substantial part of \$5 million will be spent on a public awareness campaign that may not actually address the core reason why Freedonians do not demand clean energy.

This flaw emphasizes what the tool kit states elsewhere: that a program's "theory of change . . . should be reviewed periodically to determine if it should be modified."¹⁵⁰⁶ As various parts of Chapters 2, 3 and 4 underscore, the assumptions guiding project design must be continually stress-tested by M&E systems. This is particularly true in complex, uncertain environments like Afghanistan, where causal processes are not always well understood.

Indicator Selection within State's M&E System: Innumerable Metrics, Informed by a Few Basic Principles

State defines a performance indicator as “a particular characteristic or dimension used to measure intended changes.”¹⁵⁰⁷ The purpose of indicators is to generate better program design, more useful M&E, informed learning, better choices and outcomes, and to “connect the dots between strategic plans and how to get there.”¹⁵⁰⁸ The most important are the Standard Foreign Assistance Indicators.¹⁵⁰⁹ Informally known as “F-indicators” because they are annually reported to the Office of Foreign Assistance, these indicators were intended to improve coordination between State and USAID by standardizing the administration of foreign assistance programs.¹⁵¹⁰ Partly a response to criticisms that the effects of U.S. foreign assistance were unclear, standard indicators are intended to “measure and illustrate what foreign assistance accomplishes.”¹⁵¹¹

The benefit of the standard indicators is that they can be combined across the various countries and regions in which State operates. On an annual basis, each State

TABLE 11

WHAT DOES A STANDARD FOREIGN ASSISTANCE INDICATOR LOOK LIKE?	
Categories and Examples of Foreign Assistance Indicators	
Standard Indicator Categories	Standard Indicator Examples
Peace and Security (PS) indicators correspond to U.S. counterterrorism, counternarcotics, and stabilization goals, among others.	- PS.4.4-1 Number of individuals who have received U.S. government supported cybersecurity training - PS.5.1-27 Number of survivors of [human] trafficking with sustainable livelihoods
Democracy, Human Rights, and Governance (DR) indicators aim to measure U.S. rule of law initiatives and efforts to bolster civil society.	- DR.3.1-2 Number of groups trained in conflict mediation/resolution skills or consensus building techniques with U.S. government assistance - DR.5.2-2 Number of journalists trained and supported
Health (HL) includes indicators designed to track progress against disease prevention and the effects of programs intended to improve nutrition.	- HL.2.4-1 Number of multi-drug resistant tuberculosis cases detected - HL.3.1-2 Number of malaria rapid diagnostic tests purchased with U.S. government funds
Education (ES) includes indicators that measure progress against both basic and higher education objectives.	- ES.1-3 Number of learners in primary schools or equivalent non-school based settings reached with U.S. government education assistance - ES.2-1 Number of host country higher education institutions receiving capacity development support with U.S. government assistance
Economic Growth (EG) indicators track a wide array of factors ranging from fiscal and monetary policy to agricultural development and workforce development (i.e., building workforce skills to make an economy more competitive).	- EG.2.1-2 Average time (in hours) to trade goods along trade corridor receiving U.S. government assistance - EG.6-3 Number of individuals who complete U.S. government-assisted workforce development programs
Humanitarian Assistance (HA) indicators seek to measure preparedness for natural disasters and migration management, among other areas.	HA.2.1-1 Number of people trained in disaster preparedness as a result of U.S. government assistance
Some Indicators Involve Data From Multiple Categories	
Cross Cutting Indicators measure focus areas across program categories. This creates more data points for certain kinds of analysis. Cross Cutting Indicators measure performance of programs related to gender, youth and scientific development, among others. Several of these categories are described below.	
Gender - To disaggregate beneficiaries by gender, most indicators measuring the “number of individuals” split the category into men and women. Some indicators include a separate category that accounts for lesbian, gay, bisexual, transgender, and intersex (LGBTI) people. Indicator DR.3.1-2 (listed above) can be broken out to account for number of women’s groups, LGBTI groups, indigenous people’s groups and religious freedom groups, all of which may have implications for gender data. The ability to disaggregate can facilitate analysis of how “democracy, human rights and governance” interventions are affecting different beneficiary sub-groups.	
Youth - Many indicators measuring adult recipients of assistance break down the data between “age 15–29” and “age 30+” for a closer analysis of programs meant for younger or older people. Some indicators, including EG.6-3 , can be broken down even further, creating granular age-demographic data.	
Science, Technology and Innovation/Research - Indicators that examine scientific and technological innovation can be applied to developments that happen in any of the six major categories. Training programs across all six areas may involve the use of new technological advancements. Measuring innovation achieved with U.S. government funding could cover anything from an agricultural advancement that falls under “Economic Growth” to a new social media platform that impacts “Democracy, Human Rights and Governance.”	

Source: State, “FY 2019 Master Indicator List,” n.d., accessed September 1, 2020; State, “Standard Foreign Assistance Indicators,” n.d.

mission and bureau implementing foreign assistance aggregates standard indicators in a performance plan and report, intended to convey progress towards strategic objectives.¹⁵¹² Data from the performance plan and report are then compiled in State's annual performance reports submitted to the President, the Congress, and the public.¹⁵¹³ Among other things, annual performance reports aggregate standard indicators in order to "describe progress the department is making toward the achievement of its long-term strategic goals."¹⁵¹⁴ However, the indicators have been criticized for their focus on outputs, or "what USAID and State were *doing* with their budgets rather than what they were *achieving* with their budgets."¹⁵¹⁵ Table 11 on page 214 provides further detail on standard indicators and presents examples by programmatic category.

State's projects (as well as USAID's) are often measured using a combination of standard and custom indicators. The latter are frequently more useful for measuring the effects of programs tailored to a specific context. For example, one significant development priority in Afghanistan has been promoting community-based education, an educational model that delivers instruction in communities where formal schooling is unavailable.¹⁵¹⁶ Consequently, USAID's five-year, \$70 million Afghan Children Read activity, which aims to build the capacity of Afghanistan's Ministry of Education and to pilot an early-grade reading curricula, uses numerous custom indicators to track progress against community-based education-related objectives.¹⁵¹⁷ Whenever possible, however, standard indicators generally take precedence over customized metrics.¹⁵¹⁸ For example, one indicator in trade programs could be the amount of time required to move goods across borders. While there are many potential choices for the unit of time, the standard indicators make that choice for program designers: the standardized unit is days.¹⁵¹⁹

State (and, by extension, USAID) places significant emphasis on indicator selection. According to State's M&E policy, "to the extent practicable, program indicators should be objective, practical, useful, direct, attributable, timely, and adequate."¹⁵²⁰ Table 12 below defines each of these key indicator attributes. Among other topics, chapters 2 and 3 of this report explore USAID and State's use of indicators—a key aspect of agencies' monitoring efforts—in the context of Afghanistan reconstruction.

TABLE 12

INDICATOR SELECTION CRITERIA	
Criteria	Description
Objective	Indicators should unambiguously measure a single thing. Potentially subjective terminology must be clearly defined.
Practical	Indicator selection should consider the feasibility of collecting a particular data point.
Useful	Indicator data points should be relevant to program needs and allow for program adjustments.
Direct	Indicators should directly measure the program result. In some cases, proxy indicators may be used.
Attributable	Short-term indicators that measure results directly attributable to the program should be included.
Timely	Data with an extended lag time is not useful for decision making processes. Indicators should account for the program's timeline.
Adequate	Monitoring plans should not include unnecessary or excess data beyond what is needed for the M&E process.

Source: State, "Program Design and Performance Management Toolkit," March 15, 2017, pp. 43–44.

At DOD, a Patchwork of Oversight and Other Processes Has Fulfilled Purposes Similar to M&E, but Standardization is Emerging

DOD traditionally has not practiced M&E of reconstruction programming in ways directly comparable to USAID and State. That is, until DOD's introduction in 2017 of a standardized assessment, monitoring, and evaluation policy for security cooperation, there were no standardized DOD processes equivalent to the highly complex M&E systems of USAID, or even to those of State, which have trended in the direction of similar elaboration.¹⁵²¹ But DOD has traditionally used various analogous standards and processes that have fulfilled similar purposes.¹⁵²² In particular, DOD has generally tried to track both **measures of performance** and **measures of effectiveness** to determine the outputs, outcomes, and impact of its wide array of programs and contracts.¹⁵²³

Measures of performance assess task accomplishment—for example, measuring the physical quality of newly constructed ANDSF training facilities.

Measures of effectiveness determine activity impact—for example, looking at the impact of the training facilities on security.

SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 95.

DOD's first step towards a formal, standardized M&E policy appears to have occurred in 2008, after M&E requirements and resources were included in DOD's policy guidance for fiscal year 2008 overseas humanitarian assistance.¹⁵²⁴ In 2011, an increased emphasis on assessment across DOD resulted in a prototype M&E handbook developed by the RAND Corporation to guide the assessment of military humanitarian assistance.¹⁵²⁵ Eventually, DOD released its aforementioned AM&E policy for security cooperation—but as a result of pressure from the Congress, not internal pressure, according to one knowledgeable observer.¹⁵²⁶

DOD's 2017 Assessment, Monitoring, and Evaluation Policy Introduced Possibilities for Standardization and Greater Rigor

The 2017 National Defense Authorization Act required DOD to establish an AM&E process for security cooperation programs, which would include public summaries of evaluations and an annual report to the Congress documenting lessons learned and best practices.¹⁵²⁷ In response, DOD issued an AM&E policy for security cooperation in January 2017.¹⁵²⁸ It defined the goal of AM&E as determining return on investment, identifying and improving or eliminating ineffective initiatives, providing credible oversight information, increasing understanding of which security cooperation methods work and why, and applying lessons learned to inform future security cooperation resources and policy decisions.¹⁵²⁹ These goals are similar to M&E as practiced by State and USAID.

The new policy required DOD to:

- Develop an initiative design document that specifies measurable, attainable, relevant, and time-bound objectives to guide all security cooperation initiatives and to help ensure the program is effectively implemented;¹⁵³⁰
- Develop a performance management plan for the security cooperation initiative, which includes a logic framework that maps program goals and objectives to the activities necessary to achieve desired effects; and¹⁵³¹
- Conduct independent evaluations of significant security-cooperation initiatives to determine their relevance, effectiveness, and sustainability.¹⁵³²

Congressional pressure on DOD to institutionalize monitoring and evaluation continued.¹⁵³³ The 2019 NDAA directed DOD to allocate at least \$6 million towards AM&E of security-cooperation activities.¹⁵³⁴ Moreover, the act made 50 percent of defense-wide operation and maintenance spending contingent on DOD's submission to the Congress of "a description of the activities planned for fiscal year 2019 for the evaluation of security cooperation programs" and a description of how DOD planned to organize and provide training in order to fulfill the intent of the AM&E requirement established two years before in the 2017 NDAA.¹⁵³⁵ As the same knowledgeable observer commented, "It seemed like Congress didn't believe that DOD was moving fast enough."¹⁵³⁶ According to DOD, it is "actively implementing a global program of AM&E," and, since fiscal year 2019, has spent \$20 million on AM&E annually.¹⁵³⁷

How DOD Conceptualizes Assessment, Monitoring, and Evaluation in Doctrine and Policy

While DOD doctrine is not technically policy, it contains procedures and describes processes that can be thought of as "policy-like," particularly in relation to M&E. The aspects of doctrine that are described here illuminate M&E-like frameworks that DOD has used in Afghanistan and other contingency environments that existed before the release of its 2017 AM&E policy.

Assessment Determines Progress Towards Operational Goals

According to one of DOD's numerous definitions, assessment is the "determination of progress toward accomplishing a task, creating a condition, or achieving an objective."¹⁵³⁸ The most significant form is operation assessment—the primary process through which the U.S. military determines how to conduct a campaign, and then measures progress against campaign objectives. In Afghanistan, the concept of assessment encompassed not just traditional warfighting activities, but also non-combat reconstruction activities.¹⁵³⁹ Operation assessments are flexible in their conduct and can be tailored to each commander's goals and decision-making style.¹⁵⁴⁰

Operation assessment procedures are detailed in the doctrine that guides DOD's planning process—that is, how the organization defines goals and designs strategies to achieve them. Ideally, planning begins with the definition of specific national strategic objectives and military end states around which actions and resources are focused.¹⁵⁴¹ After goals and strategies are determined, doctrine requires that commanders monitor their operating environment and assess progress toward desired objectives, as well as assess the effectiveness of operations in obtaining desired end states.¹⁵⁴² To perform this assessment, a commander's staff notes changes in the operating environment, identifies and analyzes risks and opportunities, and formally provides recommendations on courses of action.¹⁵⁴³ In theory, the commander will then use this information to adjust plans in order to make successful accomplishment of the mission more likely.¹⁵⁴⁴

In addition to operation assessments, there are also "independent assessments": DOD commissions third parties such as RAND and CNA to determine the effects of DOD programs or efforts, ranging from how development projects funded through

the Commander's Emergency Response Program contributed to counterinsurgency objectives to how DOD's train, advise, and assist efforts should focus on strengthening the ANDSF.¹⁵⁴⁵ Independent assessments bear some similarity to USAID and State evaluations, in that they are advanced analytic exercises measuring and evaluating the effects of DOD programs or efforts, and are conducted by third parties.¹⁵⁴⁶ However, such assessments are sometimes more expansive than USAID or State evaluations, which generally look retroactively at the effects of discrete projects or programs. For example, DOD-funded independent assessments of the ANDSF examined a wide array of factors, including the ANDSF's strength, structure, posture, and ability to provide adequate security. Analysis of the latter involved forward-looking estimates of insurgent actions and future levels of security.¹⁵⁴⁷

DOD Has Used Monitoring to Determine Whether Contractors Are Performing Required Tasks

DOD has defined monitoring as “a continuous process designed to provide regular feedback on the extent to which expected outputs and outcomes are being achieved to inform decisions or corrective actions.”¹⁵⁴⁸ DOD defines the goal of monitoring as measuring “the direct and near-term consequences of initiative activities” in order to “to validate the theory of change throughout implementation and an early indication of the likelihood that expected results will be attained.”¹⁵⁴⁹

However, this definition comes from the 2017 AM&E policy, and does not represent how DOD conceived of monitoring for most of Afghanistan's reconstruction.¹⁵⁵⁰ That is more accurately captured in DOD contracting doctrine, which describes monitoring as a process by which post-contract award oversight and performance monitoring are directly tied to the Federal Acquisition Regulation-based contract administration process—which emphasizes whether contracted work is executed on time, within budget, and to standard.¹⁵⁵¹

According to DOD, monitoring begins as a collaborative process between contractors and DOD. Contractors propose a plan for quality assurance, management, and reporting requirements to DOD.¹⁵⁵² When agreed upon, these reporting requirements, known as a contract data requirements list, are submitted to the contracting officer's representative and are distributed from there to other stakeholders, such as the contracting officer, the program office, and the agency requesting projects or programs.¹⁵⁵³ Commanders and their staff are responsible for assessing the readiness impacts of contracted services, for ensuring there is a sufficient number of contracting officer's representatives to conduct proper oversight, and for taking action to address major concerns about programming.¹⁵⁵⁴ Although DOD contracting doctrine states that the positive effects of monitoring can be hard to detect and that even good monitoring practices do not assure success, it argues that without proper command involvement and post-contract award oversight, the commander's staff is likely to experience higher operational costs and increased risks to forces and the military mission.¹⁵⁵⁵ As Chapter 4 describes, because of its importance to reconstruction, contracting in Afghanistan eventually became “commander's business.”

Evaluation Is Intended to Produce Learning and to Prompt Adaptation

Like other agencies, DOD conceptualizes “evaluation” primarily as a learning process, by which information collected either specifically for evaluation or through prior monitoring is used to determine whether and why (or why not) a program created its intended effects. As with monitoring, DOD has traditionally lacked a formal evaluation system. Instead, DOD has evolved its own measures and processes to evaluate programs—though they are less codified than those at State or USAID. For example, as previously mentioned, DOD formally defines two different categories of indicators: measures of performance focus on task accomplishment (for instance, whether mentoring of the Afghan National Defense and Security Forces is occurring as scheduled), and measures of effectiveness focus on results (for example, looking at how mentoring is affecting the combat capabilities of Afghan troops).¹⁵⁵⁶ While these formal indicator categories can be used to guide evaluation-like exercises, there is no DOD parallel for the extent of department-wide standardization that exists at State and USAID—in other words, no lengthy standard indicator list from which to select metrics and assess progress.

But lack of standardization does not mean that forms of evaluation did not occur even before the release of DOD’s 2017 AM&E policy for security cooperation. Contract management reviews for instance, are roughly analogous to an evaluation of a portfolio of contracts. They allow commanders or relevant stakeholders to review all the contracts under their authority and maintain, modify, or eliminate them as needed in light of campaign goals and current strategy.¹⁵⁵⁷ Evaluation of this sort is typically a commander-centric process.¹⁵⁵⁸

DOD’s AM&E policy requires DOD to conduct evaluations of significant security cooperation initiatives.¹⁵⁵⁹ The policy defines evaluation as “the systematic collection and analysis of information . . . about the characteristics and outcomes of an ongoing or completed initiative, and its design, implementation, and results.”¹⁵⁶⁰ According to the instruction, evaluations have a wide range of goals, including determining the “relevance, value, effectiveness, efficiency, sustainability, and impact” of programs in order to improve effectiveness and inform future plans, programs, and activities.¹⁵⁶¹ The instruction specifically separates evaluation from assessment and monitoring, due to its focus on “documenting the achievement of outcomes and results” and the value of U.S. investment.¹⁵⁶²

Although the stated objective of these evaluations is to “measure the effectiveness and impact” of certain initiatives, DOD does not believe that impact evaluations are possible within the context of security cooperation, as impact is too difficult to ascertain. Specifically, DOD has stated that it does not consider the use of counterfactuals (necessary for assessing impact) to be a feasible methodological strategy. DOD has reported that it plans to “conduct only performance evaluations of security cooperation initiatives” and does not plan to conduct impact evaluations.¹⁵⁶³

While the AM&E policy offers elaborate guidance on evaluation, similar to monitoring, this policy is relatively new, limited to the security cooperation sector, and does not represent how evaluation has been practiced for most of DOD’s participation in Afghanistan

reconstruction. DOD contracting doctrine offers a more representative framework for how evaluation has been conceived of and practiced. Within doctrine, “assessing contract execution readiness impact and determining future need” serves as an M&E analogue.¹⁵⁶⁴ Through this process, the command staff is responsible for analyzing the effectiveness and efficiency of contracted supply or services and their impact on force readiness through normal command readiness assessments and reporting processes, with the ultimate goal of evaluating a program’s impact and value. This process is designed to be ongoing, to ensure that contracting support is matched to changing mission requirements.¹⁵⁶⁵

Contracting doctrine shows that DOD takes the broader effects of contract work into account, stating that operational contract support is “especially important” in counterinsurgency and stability activities.¹⁵⁶⁶ “Contracted support can have a direct strategic impact on civil aspects of the operation,” it adds, and commanders can use operational contract support “to provide a positive economic and social impact on the local populace.”¹⁵⁶⁷ This link between contracting and local community effects—a core tenet of counterinsurgency theory—assumes that contracting can be used to “build the local economy, promote goodwill with the local population, and contribute to long-term . . . economic growth and stability.”¹⁵⁶⁸ To ensure contracting is creating desired effects, DOD doctrine stresses that “all major contract actions must be synchronized, monitored, de-conflicted, and, most importantly, measured.” Contracting doctrine also emphasizes the importance of understanding which actors are benefitting from contracting actions, and ensuring that U.S. contracting dollars “are properly aligned to mission objectives.”¹⁵⁶⁹

Overall, M&E at DOD is best thought of as an emerging discipline likely to become more formalized and standardized over time. But that does not mean, even before the release of its 2017 AM&E policy for security cooperation, that DOD has not engaged in M&E or attempted to hold itself accountable to standards of measurement and assessment similar in spirit and sometimes in form to those employed by State and USAID. Without an M&E policy for the majority of Afghanistan reconstruction, DOD’s M&E standards and practices varied widely. Chapter 4 examines their effectiveness using illustrative case studies of high-dollar programs and contracts.

Appendix Summary

This appendix described how the overarching authorities, guidance, and policies on contracting and monitoring and evaluation are intended to work, noting some of the differences between agencies according to their mission and function. Among the agencies, USAID has what is unambiguously the most elaborate M&E system, within which M&E is intended to emanate from, and wrap around, strategy. DOD, where M&E is not yet widely standardized, represents the opposite end of the spectrum. State, which over time has adopted many of the processes of USAID, sits between the two.

Although much of the information presented in this appendix is too granular for the body of the report, it provides a comprehensive baseline against which the actual practice of M&E is assessed throughout.



APPENDIX C: USAID COUNTRY DEVELOPMENT COOPERATION STRATEGY EVIDENCE GRADES

TABLE 13

COUNTRY DEVELOPMENT COOPERATION STRATEGY EVIDENCE GRADES

Sentence	Grade	Source	Evidence Cited
"The struggle to find a job can lead to economic marginalization and instability, which, in turn, makes one susceptible to violent extremism" (p. 13).	D	Brookings Institution, "How do Education and Unemployment Affect Support for Violent Extremism?" 2017.	The source document provides some evidence that relative deprivation (the absence of opportunities relative to expectations) in the job market can drive radicalization (p. 3).
"Another driver of extremism among youth is poor social status; evidence shows that improving young Afghans' sense of respect in their communities appears to lower their risk of supporting violent groups and causes" (p. 13).	D	MercyCorps, "Does Youth Employment Built Stability?" 2015.	The document states, "Higher respect within the community was the only social factor found to be positively associated with a decrease in support for political violence" (p. 21).
"All development objectives are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address key drivers of conflict that have enabled the Taliban to make gains, including corruption, unemployment, and a lack of government legitimacy. Achieving these objectives will help create the conditions necessary for peace" (p. 26).	B	Stephen J. Hadley, Andrew Wilder, and Scott Worden, "Four steps to winning peace in Afghanistan," <i>Washington Post</i> , 2017.	Policy experts specializing in Afghanistan make this claim directly (electronic p. 2).
"Economic growth is constrained by crippling government bureaucracy and trade policies, poor infrastructure, and infant industries that lack the enabling environment to grow" (p. 28).	A	SIGAR, Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan, 2018.	Research supports each of these conclusions (p. 3, pp 165–166).
"Increasing international trade and connectivity will accelerate private sector-driven and export-led economic growth because it provides business access to more inputs and information and technology, increasing their efficiency and productivity" (p. 29).	D	World Trade Organization, "The WTO can stimulate economic growth and employment."	The source states, "Trade can also be a catalyst for greater efficiency and productivity. This is because companies have access to a wide range of high-quality, affordable inputs . . . Access to technology and quality inputs can boost innovation and creativity in the workplace" (electronic p. 1).
"Trade also promotes investment, which is key to growth" (p. 29).	B	Islamic Republic of Afghanistan Ministry of Industry and Commerce, "Afghanistan's National Export Strategy 2018–2022: Peace through Prosperity, Prosperity through Trade," 2018.	The document states, "trade facilitation . . . stimulates private sector development, regional integration and investment attraction" (p. 2).
"In-depth analysis of the trade barriers reveals that Afghan exporters are at a significant disadvantage when compared to regional counterparts due to the high cost and time of cross border trade, as well as unfavorable regulations as reflected in the Roadmap's low Trade openness score" (p. 30).	B	Islamic Republic of Afghanistan Ministry of Industry and Commerce, "Afghanistan's National Export Strategy 2018–2022: Peace through Prosperity, Prosperity through Trade," 2018.	The document states, "In-depth analysis of the trade barriers reveals that Afghan exporters are at a significant disadvantage when compared to regional counterparts due to the high cost and time of cross-border trade as well as unfavorable regulations" (p. 1).
"Improving Afghanistan's infrastructure through targeted investments with the private sector will help to connect the region and increase international trade" (p. 30).	B	Asian Development Bank, "Afghanistan Transport Sector Master Plan Update (2017–2036)," 2017.	The Plan states, "Infrastructure constitutes an obstacle to trade facilitation," (p. xii) and that, "the high cost of transport logistics reduces the country's trade competitiveness" (p. 19).
A 2018 World Bank report entitled "Jobs from Agriculture in Afghanistan" estimated that interventions in value chains such as livestock and high-value crops created 34,847 new jobs in 2017, and that "Developing agriculture value chains is key to raising productivity and supporting job creation in Afghanistan" (p. 32).	B	World Bank, "Jobs from Agriculture in Afghanistan," 2018.	This source states, "Development of agricultural value chains is key to raising productivity and supporting job creation in agriculture," (p. 2) and presents tables showing that tens of thousands of jobs were created through various interventions (pp. 89–97).

Contradictory Context	Further flaws
However, the source also says that “previous empirical work has failed to demonstrate any link between unemployment and radicalization” (p. 1), and that “there is also no consensus on the relationship between education and violent extremism” (p. 3). Furthermore, the data does not include Afghanistan (p. 5).	The Mercy Corps source, which is also used to support the CDCS, states that “being employed was not significantly associated with either a willingness to engage in or support violence against the state,” and that “[more] business connections significantly increased the probability of being willing to use violence to fight an unfair law or state decision” (p. 20).
The cited sentence is cherry-picked from a paragraph that states, “The analysis finds little evidence that social outcomes can decrease support for political violence” (p. 21). The study also states, “Personal confidence, number of friends, and identifying as an Afghan were found to be highly significant and perversely associated with propensity towards political violence” (p. 23). The analysis makes no claims of causation, restricting itself to noting correlations (p. 21).	N/A
Apart from their own status as “experts” who held meetings with various stakeholders, no outside evidence is presented to support the claims made.	According to the Mercy Corps source: “No relationship was found between confidence in Afghan institutions—including the national, provincial, and local government—and [respondents’] propensity towards political violence” (p. 26).
N/A	N/A
The statement is qualified by adding that “competition from imports can put producers under pressure and lead them to lay off workers,” and “the impact of competition from foreign producers varies . . . across countries” (electronic p. 1).	The source is too general to apply to Afghanistan without deeper analysis.
The source also notes that, with respect to Afghanistan specifically, “Trade is not a panacea and there are complex considerations.” Those considerations include a need to account “for the ever-present scourge of corruption and security challenges” (pp. 10–12).	N/A
The source presents relevant analysis (pp. 59–65). However, the statement is pulled almost verbatim from the source. This gives the impression that it was USAID that conducted the “in-depth analysis” rather than the authors of the cited export strategy. Additionally, the source discusses the significant contextual challenges mentioned above.	N/A
However, the Plan also argues that, “procedures at BCPs [border crossing points] often disrupt trade more severely than BCP facilities and infrastructure,” (p. xii) and recommends extensive government programming instead of just private sector investment (p. 36).	N/A
The number 34,847 does not appear in the report. It is possible that the number is derivative of various employment generation figures presented in the report. However, if that is the case, it is unclear how the figure was calculated.	N/A

Continued on the following page

COUNTRY DEVELOPMENT COOPERATION STRATEGY EVIDENCE GRADES (CONTINUED)

Sentence	Grade	Source	Evidence Cited
"Studies also show that economic growth is positively associated with job creation" (p. 35).	A	The Donor Committee for Enterprise Development, "Link between economic growth and employment."	The source document argues, "Empirical studies highlight that economic growth tends to be positively associated with job creation," and includes links to relevant academic studies (electronic p. 1).
"Corruption also severely impedes economic growth and development progress, affecting the credibility of the government and fueling widespread resentment" (p. 52).	C	Integrity Watch Afghanistan, "National Corruption Survey," 2016.	"The document states, "Corruption also retards investment, making it a significant drag on the economy," (p. 1) and that "more than 70% of respondents said that corruption is now worse than it was two years ago" (p. 1).

Source: SIGAR analysis of USAID, "Afghanistan Country Development Cooperation Strategy FY 2019–2023," September, 2018, pp. 13, 26, 28–30, 32, 35, 52; SIGAR analysis of Kartika Bhatia and Hafez Ghanem, "How do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries," Global Economy and Development Working Paper 102, Brookings Institution, March 22, 2017, pp. 1, 3, 5; SIGAR analysis of MercyCorps, "Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan," January 2015, pp. 20–21, 23, 26; SIGAR analysis of Stephen J. Hadley, Andrew Wilder, and Scott Worden, "Four steps to winning peace in Afghanistan," *Washington Post*, April 27, 2017, p. 2; SIGAR analysis of SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, pp. 3, 165–166; SIGAR analysis of World Trade Organization, "The WTO can ... stimulate economic growth and employment," accessed August 12, 2020; SIGAR analysis of Islamic Republic of Afghanistan Ministry of Industry and Commerce, "Afghanistan's National Export Strategy 2018–2022: Peace through Prosperity, Prosperity through Trade," 2018, pp. 1–2, 7, 10–12, 59–65; SIGAR analysis of Asian Development Bank, "Afghanistan Transport Sector Master Plan Update (2017–2036)," 2017, pp. xii, 19, 35–36; SIGAR analysis of Leao, Izabela, Mansur Ahmed, and Anuja Kar, "Jobs from Agriculture in Afghanistan," World Bank Group, 2018, pp. 2, 89–97; SIGAR analysis of The Donor Committee for Enterprise Development, "Link between economic growth and employment," accessed August 12, 2020; SIGAR analysis of Integrity Watch Afghanistan, "National Corruption Survey," December 2016, p. 1.

Contradictory Context	Further flaws
N/A	N/A
Polling data is presented out of context, and the executive summary to which the strategy cites provides no supporting data for the link between corruption and investment (though such a link is plausible).	N/A

APPENDIX D: ACRONYMS

Acronym	Definition
ADS	Automated Directives System
AIF	Afghanistan Infrastructure Fund
AM&E	Assessment, monitoring, and evaluation
ANDSF	Afghan National Defense and Security Forces
ARTF	Afghanistan Reconstruction Trust Fund
ASOM	Afghanistan Source Operations Management program
AUAF	American University of Afghanistan
CDCS	Country Development Cooperation Strategy
CERP	Commander's Emergency Response Program
COIN	Counterinsurgency
CSSP	Corrections System Support Program
CSTC-A	Combined Security Transition Command – Afghanistan
DABS	Da Afghanistan Breshna Sherkat
DFID	UK Department for International Development
DOD	U.S. Department of Defense
FAR	Federal Acquisition Regulation
GAO	U.S. Government Accountability Office
IDEA-NEW	Incentives Driving Economic Alternatives for the North, East, and West program
IHSAN	Initiative for Hygiene, Sanitation, and Nutrition
INL	Bureau of International Narcotics and Law Enforcement Affairs (State)
ISAF	International Security Assistance Force
IWA	Integrity Watch Afghanistan
LOTFA	Law and Order Trust Fund for Afghanistan
M&E	Monitoring and evaluation
MEL	Monitoring, evaluation, and learning
MfR	Managing for Results framework
MISTI	Measuring Impacts of Stabilization Initiatives program
NDAA	National Defense Authorization Act
NGO	Nongovernmental organization
OED	Office of Education (USAID)
OMB	U.S. Office of Management and Budget
PRM	Bureau of Population, Refugees, and Migration (State)
PTec	Power Transmission Expansion and Connectivity project
SHAHAR	Strong Hubs for Afghan Hope and Resilience project
SIGACT	Significant activity
SIGAR	Special Inspector General for Afghanistan Reconstruction
SIV	Special Immigrant Visa program

Continued on the next page

Acronym	Definition
SUPPORT	Services under Program and Project Offices for Results Tracking project
USAID	U.S. Agency for International Development
WTO	World Trade Organization

ACKNOWLEDGEMENTS

SIGAR sincerely thanks the more than 100 individuals in the United States, Afghanistan, and elsewhere around the world who gave generously of their time and allowed us to interview them at length in person or by phone. For our team's fieldwork in Kabul, we greatly appreciate the hospitality of the U.S. embassy in Kabul and SIGAR's Forward Operations team.

Finally, we are grateful to the many U.S. government officials at DOD, State and USAID 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.

Research Team

Daniel Fisher, Project Lead
Paul Fishstein, Supervisory Research Analyst
Matt Bentrrott, Senior Research Analyst
Anna Andriychuk, Research Analyst
Alyssa Goodman, Research Analyst
Lauren Helinski, Research Analyst
Patrick O'Malley, Research Analyst
Joshua Herman, Student Trainee
Stephanie Pillion, Student Trainee
Holly Ratcliffe, Student Trainee

Lessons Learned Program Team

Nikolai Condee-Padunov, Program Manager
Tracy Content, Editor
Jason Davis, Visual Information Specialist
Vong Lim, Senior Visual Information Specialist
Joseph Windrem, Director

ENDNOTES

1. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 26.
2. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, p. 128; USAID, “Criteria to Ensure the Quality of the Evaluation Report: A Mandatory Reference for ADS Chapter 201,” September 7, 2016, p. 1.
3. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 58.
4. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 6.
5. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 31.
6. As used in this report, “contracts” refers to contracts, grants, and cooperative agreements. Consequently, the term “contractors” refers to both actual contractors and to nongovernmental organizations that implemented projects awarded under grants or cooperative agreements. “Nongovernmental organizations” excludes multilateral institutions like the World Bank and the United Nations. Total spending on contracts, grants, and cooperative agreements directly managed by DOD, State, and USAID is at least \$36.4 billion. This includes approximately \$20.6 billion obligated by DOD, approximately \$11.9 billion disbursed by USAID, and approximately \$4.0 billion obligated by State. The significant difference between total appropriations for reconstruction and total spending on contracts is due primarily to the gap between total security appropriations (\$88.3 billion, as of March 31, 2021) and confirmed DOD spending on contracts. As SIGAR reported in 2015, accounting idiosyncrasies arising from intradepartmental transfers make it difficult to track total spending on DOD’s reconstruction contracts after the transfers occur. For example, a significant portion of Afghanistan Security Forces Fund (ASFF) monies have been transferred to the Defense Security Cooperation Agency (DSCA) to be executed as Foreign Military Sales (FMS) funding. According to DOD, once ASFF funds are transferred to DSCA for FMS execution, the funds are no longer tracked as ASFF funding. Additionally, there are gaps in DOD data for reconstruction funds spent on contracts prior to 2010. For DOD’s and State’s total spending on contracts, grants, and cooperative agreements, this report relied exclusively on two SIGAR special projects published in 2014 and 2015 that reported cumulative obligations on contracts for State and DOD, respectively. USAID was the only agency whose financial data was conducive to a comprehensive, updated analysis of spending on contracts, grants, and cooperative agreements. For that, this report relies on USAID spending data transmitted to SIGAR for its April 30, 2021 quarterly report to the Congress. SIGAR, *Department of Defense Spending on Afghanistan Reconstruction: Contracts Comprised \$21 Billion of \$66 Billion in Total Appropriations, 2002–May 2014*, SIGAR-15-40-SP, March 2015, pp. 5, 8–9; SIGAR, *Department of State Assistance to Afghanistan: \$4 Billion Obligated Between 2002 and 2013*, SIGAR-14-49-SP, April 2014, pp. 5, 7; SIGAR analysis of USAID, response to SIGAR quarterly report data call, April 13, 2021; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 31.
7. SIGAR, *Afghan Air Force: DOD Met the Initial Date for Fielding UH-60 Helicopters, but the Program Is at Risk of Not Having Enough Trained Pilots or the Capability to Maintain Future UH-60s*, SIGAR-19-18-AR, January 2019; SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017; USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Assistance in Building Afghanistan by Developing Enterprises Program,” Audit Report No. F-306-16-002-P, March 14, 2016.
8. Although numerous reports underscore the problems created by contracting, many of those reports also either directly or indirectly point to benefits. For example, a 2011 U.S. Senate report stated, “Contractors support direct-hire personnel, implement assistance projects, and address U.S. government workforce shortfalls.” However, the report also noted, “While there are many good reasons to use contractors in Afghanistan, there are also reasons for concern.” The Commission on Wartime Contracting also highlighted contracting’s costs and benefits—for example, in its observation: “Particularly important is the impact on U.S. objectives resulting from the government’s extensive use of contractors. Using local contractors not only supports the local economy, but often helps the United States develop a good rapport with the host-nation government and communities. However, rapidly pouring large amounts of money into Afghanistan’s local economy, which has limited absorptive capacity, has contributed to inflation, distorted normal economic activity, and encouraged fraud and corruption.” As researchers Richard Fontaine and John Nagl summarized, “When our nation goes to war, contractors go with it. Contractors have become an enduring feature of modern American conflicts, and the United States cannot now engage in hostilities or in reconstruction and stabilization operations without them.” Richard Fontaine and John Nagl, “Contractors in American Conflicts: Adapting to a New Reality,” Center for a New American Security, December 2009, p. 5, 11; U.S. Senate Foreign Relations Committee, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations, United States Senate,” June 2011, p. 15;

- U.S. Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risks,” August 2011, p. 29.
9. Notably, because this figure encompasses only those instances of waste, fraud, and abuse identified through SIGAR’s work, it may undercount the amount. Other oversight entities like GAO and inspectors general at DOD, State, and USAID have also identified numerous instances of waste, fraud, and abuse in Afghanistan reconstruction. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, pp. i, 11; SIGAR, *Review Letter: Waste, Fraud, and Abuse Uncovered by SIGAR*, SIGAR-18-60-SP, July 17, 2018, electronic p. 8.
 10. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. i.
 11. A number of commissions and congressional committees have produced reports on weaknesses in oversight of contracting, including the following: “Transforming Wartime Contracting: Controlling Costs, Reducing Risks. Final report to Congress, Findings and recommendations for legislative and policy changes,” Commission on Wartime Contracting in Iraq and Afghanistan, July 2011; “Urgent Reform Required: Army Expeditionary Contracting: Report of the ‘Commission on Army Acquisition and Program Management in Expeditionary Operations,’” undated; “Warlord, Inc.: Extortion and Corruption Along the U.S. Supply Chain in Afghanistan,” Subcommittee on National Security and Foreign Affairs, Committee on Oversight and Government Reform, U.S. House of Representatives, June 2010; Senate Foreign Relations Committee, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations United States Senate,” July 2011; The United States Commission on Helping to Enhance the Livelihood of People Around the Globe, “Beyond Assistance: The Help Commission Report on Foreign Assistance Reform,” Appendix 10, “Foreign Assistance Grant and Contract Procurement Processes,” December 2007. The United States Commission on Helping to Enhance the Livelihood of People Around the Globe, “Beyond Assistance: The Help Commission Report on Foreign Assistance Reform,” Appendix 10, “Foreign Assistance Grant and Contract Procurement Processes,” December 2007.
 12. Asking difficult and uncomfortable questions is generally not a strength of M&E. Yet doing so may reveal that some programming should never have been implemented in the first place. As Inspector General Sopko stated in reference to SIGAR’s 2013 request to U.S. agencies to provide a list of their 10 most and 10 least successful reconstruction programs and projects during January 2020 congressional testimony: “And I go back to why some of you may have wondered why did I attach all of those letters from 2013 when I asked the [Secretary of Defense, the Secretary of State and the] USAID administrator, what are your 10 best successes and what were your 10 worst failures and why. I firmly believe if they had honestly answered those questions, we wouldn’t be here today. Because what they would have done is it would force them to answer the question, why are we spending \$9 billion on narcotics if it’s a failure?” John F. Sopko “U.S. Lessons Learned in Afghanistan,” testimony before the House Committee on Foreign Affairs, 116th Cong., January 15, 2020, electronic p. 41.
 13. Allison Stanger, *One Nation Under Contract: The Outsourcing of American Power and the Future of Foreign Policy* (New Haven, CT: Yale University Press, 2009), pp. 15, 59.
 14. For example, see: Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risks,” August 2011; The United States Commission on Helping to Enhance the Livelihood of People Around the Globe, “Beyond Assistance: The Help Commission Report on Foreign Assistance Reform,” December 2007; “Warlord, Inc.: Extortion and Corruption Along the U.S. Supply Chain in Afghanistan,” Subcommittee on National Security and Foreign Affairs, Committee on Oversight and Government Reform, U.S. House of Representatives, June 2010; Senate Foreign Relations Committee, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations United States Senate,” July 2011.
 15. See: 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; SIGAR, *Afghanistan’s Health Care Sector: USAID’s Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017; SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015; SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program’s Achievements and the Afghan Government’s Ability to Sustain Them*, SIGAR-18-69-AR, September 7, 2018; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 31, 2017; SIGAR, *Afghanistan’s Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, May 1, 2019.
 16. As one commentator observed, “While no one in the White House or at the State Department, much less the Pentagon, will say so aloud, the fact is that the vaunted U.S. military has suffered a resounding defeat in Afghanistan. There is no disguising the fact that it has failed to accomplish the mission assigned. It surrenders the field of battle to an enemy that now sees final victory in sight. Even without dramatic video footage of overloaded helicopters lifting off from the roof of the US embassy, the likely outcome of the Afghanistan War bears comparison with the Vietnam debacle of a half-century ago.” Andrew Bacevich, “Afghanistan

- the long defeat,” *Spectator*, April 19, 2021; State, “Afghanistan: Integrated Country Strategy,” September 2018, p. 4.
17. According to the World Bank, the Afghan government “will remain highly dependent on aid inflows” through at least 2024. Given that grant needs in that year are projected to be 13 percent of GDP—only 0.4 percentage points lower than the present—it seems reasonable to assume that donors, including the United States, will have to continue to provide significant amounts of aid long into the future, holding constant the current size and reach of the Afghan government. World Bank, “Afghanistan Development Update April 2021: Setting Course to Recovery,” April 2021, p. 23.
 18. State, “Afghanistan: Integrated Country Strategy,” September 2018, p. 4.
 19. State, “Afghanistan: Integrated Country Strategy,” September 2018, p. 2.
 20. State, “Pakistan: Integrated Country Strategy,” April 2019, pp. 2, 6–8.
 21. State, “Niger: Integrated Country Strategy,” August, 2018, pp. 2, 5.
 22. State, “Honduras: Integrated Country Strategy,” September, 2018, p. 4.
 23. National Defense Authorization Act for Fiscal Year 2008, P. L. 110-181, § 123 Stat. 1932-33.
 24. As SIGAR noted in a 2012 audit of the Afghan First Initiative, the program was funded in part by monies from ASFF and ESF. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data Is Limited*, SIGAR Audit-12-6, January 31, 2012, p. 1.
 25. For example, USAID’s “program cycle” describes how USAID “operationalizes development policy.” As described in Appendix B, the process begins with the development of a strategy that then informs the design of projects and activities. Monitoring and evaluation is intended to be “threaded throughout.” Although USAID’s program cycle begins with strategy, the concept is equally applicable to projects and activities, as M&E is intended to help USAID “make decisions, and inform course-corrections as needed (including the adjustment of non-performing awards).” Just as a USAID strategy must be grounded in a “theory of change,” so too must projects. Although DOD and State use other terminology, their policies also describe program and project planning and implementation as cyclical processes that involve continuous assessment and learning. DOD security cooperation initiatives and State programs and projects must be similarly underpinned by a theory of change. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” partial revision dated July 23, 2020, pp. 11–12, 14, 60; DOD, “DOD Instruction 5132.14 Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise,” January 13, 2017, pp. 12–13, 14–15; State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 3.
 26. If the theory is that locals are less likely to join the insurgency when they are employed, the challenge with construction projects may be finding the right mix between heavy machinery and labor. Using heavy equipment might improve the quality and efficiency of construction, but if one purpose of the road project is to maximize short-term employment, a project designer might wish to err on the side of constructing the road using more labor-intensive methods.
 27. To be clear, short-term employment could also be considered an outcome. A project like this speaks to the sometimes blurry line between outputs and outcomes on one hand, and outcomes and impact on the other.
 28. There are exceptions: “End-use monitoring” refers to the practice of ascertaining whether property transferred to a foreign government is being used as intended. For example, CSTC-A has conducted inspections of vehicles provided to the Afghan government to ensure that security forces could account for the equipment. The Office of Management and Budget Monitoring and Evaluation Guidelines (OMB Memorandum M-18-04), cited in General Accounting Office, “Foreign Assistance: Federal Monitoring and Evaluation Guidelines Incorporate Most but Not All Leading Practices,” Appendix II, July 2019, p. 4; SIGAR, *DOD Improved Its Accountability for Vehicles Provided to the Afghan National Security Forces, but Should Follow Up on End-Use Monitoring Findings*, SIGAR Audit-12-4, January 2012, pp. ii, 9–10.
 29. The Office of Management and Budget Monitoring and Evaluation Guidelines (OMB Memorandum M-18-04), cited in General Accounting Office, “Foreign Assistance: Federal Monitoring and Evaluation Guidelines Incorporate Most but Not All Leading Practices,” Appendix II, July 2019, p. 4; USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, p. 123.
 30. For example, USAID’s program cycle policy suggests that USAID missions conduct “a review of evaluations” in order to develop the theory of change for new projects. In USAID parlance, projects are technically distinct from activities, though they have similar design requirements. For instance, USAID recommends that “monitoring data” and “evaluations of relevant and prior activities or projects” be considered when new activities are designed. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” partial revision dated June 11, 2019, pp. 67, 92.
 31. In reference to Afghanistan, one study observed, “In an environment with little reliable quantitative data, with numerous independent variables that make determining correlation (not to mention causality) virtually impossible, and where western-style public opinion polling methodologies may not be reliable, the determination of impact may often have to be more art than science.” Speaking about development more broadly and, in particular, country-level markers of development like state capability and wealth, development economist Lant Pritchett has argued, “Establishing causation amongst aggregate variables is almost impossible—while Denmark or Finland is rich, democratic, highly educated and has high state capability and Nepal or Haiti or Mali are none of those things, it is hard to parse out which are horses and which carts.” Paul

- Fishstein and Andrew Wilder, "Winning Hearts and Minds? Examining the Relationship between Aid and Security in Afghanistan," Feinstein International Center, Tufts University, January 2012, p. 6; Lant Pritchett, "Folk and the Formula: Fact and Fiction in Development," United Nations University, May 6, 2013, p. 16.
32. Carey Luse, Christopher Madeline, Landon Smith, and Stephen Starr, "An Evaluation of Contingency Contracting: Past, Present, and Future," submitted in partial fulfillment of the requirements for the degree of Master of Business Administration, Naval Postgraduate School, December 2005, p. 5.
 33. Richard Fontaine and John Nagl, "Contractors in American Conflicts: Adapting to a New Reality," Center for a New American Security, December 2009, p. 13; Commission on Wartime Contracting in Iraq and Afghanistan, "Transforming Wartime Contracting: Controlling Costs, Reducing Risks," August 2011, p. 41.
 34. Richard Fontaine and John Nagl, "Contractors in American Conflicts: Adapting to a New Reality," Center for a New American Security, December 2009, p. 6.
 35. The United States Commission on Helping to Enhance the Livelihood of People Around the Globe, "Beyond Assistance: The Help Commission Report on Foreign Assistance Reform," Appendix 10, "Foreign Assistance Grant and Contract Procurement Processes," December 2007, p. 128.
 36. Allison Stanger, *One Nation Under Contract: The Outsourcing of American Power and the Future of Foreign Policy* (New Haven, CT: Yale University Press, 2009), p. 15.
 37. Allison Stanger, *One Nation Under Contract: The Outsourcing of American Power and the Future of Foreign Policy* (New Haven, CT: Yale University Press, 2009), p. 59.
 38. Allison Stanger, *One Nation Under Contract: The Outsourcing of American Power and the Future of Foreign Policy* (New Haven, CT: Yale University Press, 2009), pp. 120–121.
 39. Richard Fontaine and John Nagl, "Contractors in American Conflicts: Adapting to a New Reality," Center for a New American Security, December 2009, p. 11.
 40. The United States Commission on Helping to Enhance the Livelihood of People Around the Globe, "Beyond Assistance: The Help Commission Report on Foreign Assistance Reform," Appendix 10, "Foreign Assistance Grant and Contract Procurement Processes," December 2007, p. 47.
 41. Richard Fontaine and John Nagl, "Contractors in American Conflicts: Adapting to a New Reality," Center for a New American Security, December 2009, p. 12.
 42. Gordon Adams and Shoon Murray, eds., *Mission Creep: The Militarization of U.S. Foreign Policy?* (Washington, DC: Georgetown University Press, 2014), pp. 3–4.
 43. OECD, "DAC Peer Review: Main Findings and Recommendations, Review of the Development Co-operation Policies and Programmes of United States" 2006, p. 26. These figures do not appear in the 2011 and 2016 volumes, so it is not possible to analyze for later years.
 44. Management Systems International, "Trends in U.S. Foreign Assistance Over the Past Decade," August 2009, pp. iii, 8.
 45. Regarding CERP's link to warfighting, see endnote 48. Mark Martins, "The Commander's Emergency Response Program," *Joint Forces Quarterly*, Issue 37, Second Quarter (2005), pp. 49–50.
 46. Mark Martins, "No Small Change of Soldiering: The Commander's Emergency Response Program (CERP) in Iraq and Afghanistan," *The Army Lawyer*, Department of the Army Pamphlet 27-50-369, February 2004, p. 11.
 47. Emergency Supplemental Appropriations Act for Defense and for the Reconstruction of Iraq and Afghanistan, 2004, Pub. L. No. 108-106, § 1110 (2003).
 48. As then-Lieutenant Colonel Mark Martins—who at the time was serving as deputy legal counsel in the Office of the Chairman of the Joint Chiefs of Staff—wrote in the February 2004 issue of *The Army Lawyer*, "CERP originated as an effort to provide commanders in Iraq with a stabilization tool for the benefit of the Iraqi people." Martins clearly sees no conflict between the appropriating language for CERP and the concepts of stabilization and counterinsurgency: He describes CERP as "meeting emergency needs of local Iraqi and Afghan civilians" while at the same time referring to it as "a program by which field commanders . . . can fund initiatives to win hearts and minds [and] hunt enemies." However, the appropriating language does not draw an explicit connection to the security objectives implied by stabilization and counterinsurgency. Mark Martins, "No Small Change of Soldiering: The Commander's Emergency Response Program (CERP) in Iraq and Afghanistan," *The Army Lawyer*, Department of the Army Pamphlet 27-50-369, February 2004, pp. 2–3; Mark Martins, "The Commander's Emergency Response Program," *Joint Forces Quarterly*, Issue 37, Second Quarter (2005), p. 49; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 92–93.
 49. This theory of change has been extensively debated, including in the USAID-funded Radha Iyengar, Jacob N. Shapiro, Stephen Hegarty, "Lessons Learned from Stabilization Initiatives in Afghanistan: A Systematic Review of Existing Research," RAND Corporation, April 2017, p. 21; SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, pp. 42, 48–49; SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data Is Limited*, SIGAR Audit-12-6, January 31, 2012, p. 2.
 50. U.S. Congressional Research Service (CRS), "The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions," R42826, February 3, 2015, p. 11.
 51. Federal Acquisition Regulation, foreword, March 2005, electronic p. 5.

52. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 8.
53. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 9; 48 CFR §1.102: “Statement of guiding principles for the Federal Acquisition System,” July 3, 1995, p. 6.
54. Specifically, agencies must “to the maximum extent practicable . . . describe the work [to be performed by the contractor] in terms of the required results rather than either ‘how’ the work is to be accomplished or the number of hours to be provided.” In addition, agencies must incorporate “measurable” and “structured” standards against which the performance of a contractor can be assessed. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 8; Federal Acquisition Regulation, Part 37: “Service Contracting,” Subpart 37.6: “Performance-based Acquisition,” §37.602: “Performance work statement,” §37.603: “Performance standards”; ASI Government, “A COR’s Guide to Statements of Work, Performance Work Statements, and Statements of Objectives,” January 9, 2011, p. 1.
55. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 4, 34–35, 37, 42.
56. U.S. Embassy Kabul, “New approach to the delivery of U.S. assistance to Afghanistan in support of the President’s strategy,” Kabul 001776 cable, July 8, 2009, pp. 7–8.
57. U.S. Embassy Kabul, “New approach to the delivery of U.S. assistance to Afghanistan in support of the President’s strategy,” Kabul 001776 cable, July 8, 2009, p. 5.
58. Again, CERP’s purpose was always somewhat blurry. See endnote 48 above. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 93, 96.
59. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 42.
60. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 56.
61. Reconstruction has tainted the concept of “burn rate.” While measuring success exclusively by the amount of funds spent is unreasonable, it is nonetheless important to keep track of how much and how quickly money is spent. When used as a less crude measure—that is, when expenditure data are combined with, say, data on the kinds of projects that were funded, or the timeliness or magnitude of the spending—expenditures can theoretically be a useful (if still imperfect) tool for gauging outcomes or impact. For example, see: Greg Adams, “Honing the Proper Edge: CERP and the Two-sided Potential of Military-led Development in Afghanistan,” *The Economics of Peace and Security Journal*, vol. 10, no. 2 (2015); 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, pp. 192, 215, 250.
62. See: Aram Roston, “How the U.S. funds the Taliban,” *Nation*, November 11, 2009; Jean MacKenzie, “Funding the Afghan Taliban,” PRI, GlobalPost, August 7, 2009, updated May 30, 2010.
63. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, p. 3; Dexter Filkins, “Depositors Panic Over Bank Crisis in Afghanistan,” *New York Times*, September 2, 2010.
64. Stuart Gordon, “Winning Hearts and Minds? Examining the Relationship between Aid and Security in Afghanistan’s Helmand Province,” Feinstein International Center, 2011, pp. 44–45.
65. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, p. 3.
66. USAID’s 2005 anticorruption strategy was a reference document for USAID policies. The definition presented here conforms to that used in SIGAR’s lessons-learned report on corruption in Afghanistan. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, pp. 2, 100.
67. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, p. 3.
68. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, pp. 2–3.
69. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, p. 4.
70. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, p. 94.
71. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, p. 3.
72. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, p. 3.
73. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, p. 13.
74. USAID, Christopher Bodle and Diane Ray, “Accountable Assistance for Afghanistan,” June 2011, pp. 9, 11, 13.
75. SIGAR, *Oversight Access Inquiry Letter to Department of Defense, Department of State and U.S. Agency for International Development*, SIGAR-14-4-SP, October 10, 2013, p. 1.
76. USAID officials, SIGAR interview, April 2019.
77. John F. Sopko, “Testimony Before the Homeland Security and Governmental Affairs Committee (HSGAC) Subcommittee on Federal Spending Oversight and Emergency Management (FSO) U.S. Senate,” May 9, 2018, pp. 9–10.
78. Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risk,” August 2011, p. 41.
79. Richard Fontaine and John Nagl, “Contractors in American Conflicts: Adapting to a New Reality,” Center for a New American Security, December 2009, p. 13.

80. Federal Acquisition Regulation, Part 7: “Acquisition Planning,” Subpart 7.5: “Inherently Governmental Functions,” §7.503: “Policy.”
81. Some analysts have suggested this lack of clarity could be solved by substituting the concept of inherently governmental functions with a blanket prohibition on outsourcing “core functions,” “mission-essential functions,” or “critical government functions.” But these terms may simply push the confusion into a new framework, as the question of what constitutes a “core function” or “mission essential task” is in many ways equivalent to the question of what is “inherently governmental.” Other proponents of scrapping the “inherently governmental” standard advocate for a maximalist approach to outsourcing, arguing that the boundaries between functions that are inherently governmental and those that contractors are permitted to perform should largely be left to the discretion of agencies, subject perhaps to some very specific congressional restrictions. Richard Fontaine and John Nagl, “Contractors in American Conflicts: Adapting to a New Reality,” Center for a New American Security, December 2009, p. 13.
82. Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risk,” August 2011, p. 3.
83. Richard Fontaine and John Nagl, “Contractors in American Conflicts: Adapting to a New Reality,” Center for a New American Security, December 2009, p. 5.
84. Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risk,” August 2011, pp. 28–32.
85. Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risk,” August 2011, pp. 29, 31.
86. The United States Commission on Helping to Enhance the Livelihood of People Around the Globe, “Beyond Assistance: The Help Commission Report on Foreign Assistance Reform,” Appendix 10, “Foreign Assistance Grant and Contract Procurement Processes,” December 2007, p. 129; Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risk,” August 2011, p. 2.
87. Allison Stanger, *One Nation Under Contract: The Outsourcing of American Power and the Future of Foreign Policy* (New Haven, CT: Yale University Press, 2009), p. 91.
88. For example, see SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, pp. i–ii, 10; 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, p. 5; USAID Office of Inspector General, “Audit Of USAID/Afghanistan’s Assistance in Building Afghanistan by Developing Enterprises Program,” Audit Report No. F-306-16-002-P, March 14, 2016, p. 9; SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, pp. 6–9.
89. For example, see the analysis of USAID and State evaluations in chapters 2 and 3, respectively, and the analysis of the Commander’s Emergency Response Program in chapter 4. The discussions of shortcomings in USAID’s strategic review process in chapter 2 and of State’s Corrections System Support Program and Legal Aid through Legal Education program in chapter 3 also speak to this point. Additional examples can be found throughout the report.
90. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, p. ii.
91. James Dobbins, Michele A. Poole, Austin Long, Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, pp. 91–92.
92. USAID/Afghanistan senior official, SIGAR interview, June 2019.
93. USAID staff in the Bureau for Policy, Planning, and Learning, SIGAR interview, February 2019.
94. USAID staff in the Office of Infrastructure, SIGAR interview, May 2019.
95. Former USAID/Afghanistan senior official, SIGAR interview, June 2019; SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, p. 36.
96. Beth Cole and Cristina Parajon, “The Afghanistan Reconstruction Group: An Experiment with Future Potential,” U.S. Institute of Peace, published on September 12, 2006, accessed on February 8, 2018; Former USAID/Afghanistan senior official, SIGAR interview, June 2019.
97. U.S. Government Accountability Office (GAO), “Afghanistan Reconstruction: Despite Some Progress, Deteriorating Security and Other Obstacles Continue to Threaten Achievement of U.S. Goals,” GAO-05-742, July 28, 2005, p. 51.
98. SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, p. 37.
99. National Defense Authorization Act for Fiscal Year 2008, Pub. L. No. 110-181, § 1229, 123 Stat. 1932–33; National Defense Authorization Act for Fiscal Year 2008, Pub. L. No. 110-181, Title VII, Sub. D, § 841.
100. Pete Troilo, Jeff Tyson, “Dark cloud of legal trouble begins to lift away from Louis Berger,” *Deveez*, August 31, 2015.
101. GAO, “Afghanistan: U.S. Efforts to Vet Non-U.S. Vendors Need Improvement,” GAO-11-355, June 8, 2011, electronic p. 2.

102. GAO, "DOD, State, and USAID Are Taking Actions to Track Contracts and Contractor Personnel in Iraq and Afghanistan," GAO-09-538T, April 2009, electronic p. 2; GAO, "Contingency Contracting: DoD, State, and USAID Contracts and Contractor Personnel in Iraq and Afghanistan," GAO-09-19, October 2008, electronic p. 2.
103. GAO, "Afghanistan: U.S. Efforts to Vet Non-U.S. Vendors Need Improvement," GAO-11-355, June 8, 2011, pp. 9, 18, 21; State Office of Inspector General, "Review of the Bureau of Administration, Office of Logistics Management, Critical Environment Contract Analysis Staff's Counterterrorism Vetting Function (Risk Analysis and Management)," August 2018, p. 15.
104. Henrietta H. Fore, testimony before the Senate Committee on Foreign Relations, Hearing on Foreign Assistance Reform, June 12, 2007; State, Office of the Director of Foreign Assistance, "Foreign Assistance Performance Indicators," January 8, 2007, accessed May 24, 2021; Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, "Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations For Change," May 2009, p. 13; CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 14.
105. U.S. Embassy Kabul, "New approach to the delivery of U.S. assistance to Afghanistan in support of the President's strategy," Kabul 001776 cable, July 8, 2009.
106. White House, Office of the Press Secretary, "Remarks on United States Military and Diplomatic Strategies for Afghanistan and Pakistan," transcript, March 27, 2009, p. 5.
107. U.S. Embassy Kabul, "New approach to the delivery of U.S. assistance to Afghanistan in support of the President's strategy," Kabul 001776 cable, July 8, 2009.
108. White House, memorandum to the National Security Council on the Presidential Policy Directive/PPD-6, "U.S. Global Development Policy," September 22, 2010, p. 6; White House, "Fact Sheet: U.S. Global Development Policy," September 22, 2010, accessed August 2, 2019.
109. State, "Leading Through Civilian Power, The first Quadrennial Diplomacy and Development Review," 2010, electronic p. 4, pp. xi, 18.
110. Ben Connable, "Embracing the Fog of War: Assessment and Metrics in Counterinsurgency," RAND National Defense Institute, 2012, pp. 155–156.
111. Henrietta H. Fore, testimony before the Senate Committee on Foreign Relations, hearing on Foreign Assistance Reform, June 12, 2007; State, Office of the Director of Foreign Assistance, "Foreign Assistance Performance Indicators," January 8, 2007, accessed May 24, 2021; State, "U.S. Foreign Assistance Reform: Achieving Results & Sustainability in Support of Transformational Diplomacy," June 2006.
112. Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, "Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change," May 2009, p. 13.
113. State, "FY 2019 Master Indicator List," 2018, accessed September 1, 2020.
114. Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, "Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change," May 2009, p. 13.
115. Ben Connable, "Embracing the Fog of War: Assessment and Metrics in Counterinsurgency," RAND National Defense Institute, 2012, p. 154.
116. Ben Connable, "Embracing the Fog of War: Assessment and Metrics in Counterinsurgency," RAND National Defense Institute, 2012, p. 154.
117. The assessment was leaked to Bob Woodward of the *Washington Post*, whose September 21, 2009, article on the topic was titled, "McChrystal: More Forces or 'Mission Failure.'" Regarding Afghan government corruption, Woodward wrote, "The assessment offers an unsparing critique of the failings of the Afghan government, contending that official corruption is as much of a threat as the insurgency to the mission of the International Security Assistance Force." Bob Woodward, "McChrystal: More Forces or 'Mission Failure,'" *Washington Post*, September 21, 2009.
118. ISAF, "Commander's Initial Assessment, August 30, 2009, p. 1-1.
119. For Stewart's scathing account of overoptimism in the military, see: Rory Stewart and Gerald Knaus, *Can Intervention Work*, (New York, NY: W.W. Norton & Company, 2011), pp. 50–58. Stewart adds, "Every year from 2005, more and more money was spent, more and more troops were deployed. Each year was to prove decisive. None was. . . . These attitudes were not confined to the center: they echoed down the hierarchy in every province." John Glaser and John Mueller of the CATO Institute similarly describe this phenomenon in an August 2019 policy analysis arguing for an end to the war in Afghanistan. Regarding military overoptimism, Glaser and Mueller write, "Overly optimistic portrayals are partly a result of institutional habits and a view about civil-military relations that calls for focusing on tactical and operational facts on the ground while leaving broader strategic and political assessments of the war to elected leaders. Some military leaders publicly misrepresented the course of the war to avoid the hit to troop morale they expected would result from more honest and critical presentations. Others felt strongly that negotiations with the Taliban should only occur from a 'position of strength,' which they believed was always just around the corner. But sometimes the deception was more flagrant: media reports revealed in 2011 that commanders tasked with briefing congressional delegations in Afghanistan deliberately misled members of Congress about the progress of the war." As Inspector General Sopko said in January 2020 testimony before the Congress, "We have

- incentivized lying to Congress and by that, I mean the whole incentive is to show success and to ignore the failure and when there's too much failure, classify it or don't report it." For more, see John Glaser and John Mueller, "Overcoming Inertia Why It's Time to End the War in Afghanistan," CATO Institute, Policy Analysis No. 878, August 13, 2019, pp. 5–6; John F. Sopko, "U.S. Lessons Learned in Afghanistan," testimony before the Committee on Foreign Affairs, 116th Congress, January 15, 2020, electronic p. 42.
120. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51.
 121. For instance, the CUAT system was more complex than the CM rating system, while the RASR was designed to improve upon the CUAT by measuring the ability of ANDSF units to utilize their staffing, equipping, and training levels. Additionally, while the prior assessment mechanisms measured the ANA down to the kandak (battalion) level, the MAAR assessed the ANDSF only at the headquarters level. The words used to describe each rating level also tended to change from system to system. For example, the lowest rating level in the MAAR was "In development." By contrast, the lowest rating level of the CM was "May be capable of conducting portions of missions, reliant on support." SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. 51, 53–54; SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, pp. 6–7; DOD, "Report on Progress Toward Security and Stability in Afghanistan," November 2013, pp. 75, 78; SIGAR, *High Risk List*, January 2017, p. 15.
 122. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. xii, 51.
 123. The program was called "Services under Program and Project Offices for Results Tracking" and was generally referred to by its acronym—SUPPORT. USAID staff at the Bureau for Policy, Planning, and Learning, SIGAR interview, February 2019; Checchi and Company Consulting, Inc., "Final Report: Services under Program and Project Offices for Results Tracking Project (SUPPORT)," April 2013, p. 1.
 124. SUPPORT eventually expanded into monitoring in addition to other activities as security began to deteriorate midway through the program and officials found it more difficult to conduct site visits. Checchi and Company Consulting, Inc., "Final Report: Services under Program and Project Offices for Results Tracking Project (SUPPORT)," April 2013, pp. 12, 16.
 125. CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 14.
 126. Emphasis on evaluation during this time period was also on display within the U.S. military. The U.S. Army's February 2011 "Afghanistan Provincial Reconstruction Team Handbook," for instance, contained instructions for conducting M&E activities within the District Stability Framework to determine whether programs were contributing to stabilization goals. DOD's language and definitions differed from those of the civilian agencies, which increased the difficulty of making comparisons across agencies. The District Stability Framework was developed by USAID to allow all partners working on a Provincial Reconstruction Team to identify local sources of instability and create programs to counteract them. Gordon Weynand, "Establishing, Monitoring, and Refining an Evaluation Policy at the Department of State," *Dipnote*, March 29, 2018, pp. 1–2; Center for Army Lessons Learned, "Handbook for Provincial Reconstruction Team: Observations, Insights, and Lessons," Document 11-16, February 2011; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 44.
 127. GAO, "Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue," GAO-11-318SP, March 2011, pp. 120–121; USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 5.
 128. SIGAR is not alone in making this observation. For example, see: GAO, "Afghanistan Development: Enhancements to Performance Management and Evaluation Efforts Could Improve USAID's Agricultural Programs," GAO-10-368, July 2010, p. 33; GAO "Afghanistan Development: USAID Continues to Face Challenges in Managing and Overseeing U.S. Development Assistance Programs," GAO-10-932T, July 15, 2010, p. 2.
 129. USAID, "Fact Sheet: Measuring Impacts of Stabilization Initiatives," updated October 2013. SIGAR, *USAID's Measuring Impacts of Stabilization Initiatives: Program Generally Achieved Its Objectives, but USAID's Lack of a Geospatial Data Policy and Standards Affected Its Implementation*, SIGAR-17-10-AR, October 2016, p. 3.
 130. SIGAR, *USAID's Measuring Impacts of Stabilization Initiatives: Program Generally Achieved Its Objectives, but USAID's Lack of a Geospatial Data Policy and Standards Affected Its Implementation*, October 2016, pp. i–ii.
 131. CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 14.
 132. Paul Fishstein, Anna Paterson, and Alan Roe, "Mapping of Research Capacity in Afghanistan," submitted to UK Department for International Development, September 2012, p. 40.
 133. Paul Fishstein, Anna Paterson, and Alan Roe, "Mapping of Research Capacity in Afghanistan," submitted to UK Department for International Development, September 2012, pp. 40–41.
 134. UK Department for International Development, "Data Quality Assessment of the Asia Foundation (TAF) Surveys," September 2010, pp. ii, 1–2; Department for International Development, *Data Quality Assessment on the Theatre Integrates Nationwide Survey (TINS)*, September 2010, pp. 2–3.

135. Paul Fishstein, Anna Paterson, and Alan Roe, "Mapping of Research Capacity in Afghanistan," submitted to UK Department for International Development, September 2012, p. 40.
136. Wilton Park, "Conference report: Workshop on evaluation lessons from development assistance to Afghanistan," February 2013, pp. 1–2; Charles Michael Johnson, Jr., "Afghanistan, Key Oversight Issues for USAID Development Efforts," testimony before the Subcommittee on National Security, Committee on Oversight and Government Reform, House of Representatives, March 13, 2014, p. 9.
137. Charles Michael Johnson, Jr., "Afghanistan, Key Oversight Issues for USAID Development Efforts," testimony before the Subcommittee on National Security, Committee on Oversight and Government Reform, House of Representatives, March 13, 2014, p. 9; Wilton Park, "Conference report: Workshop on evaluation lessons from development assistance to Afghanistan," February 2013, p. 3.
138. As one USAID project manager put it in 2014, "The most effective monitoring tool I use to prove that roads, bridges, schools and irrigation canals are being built is my smartphone." USAID, "True Detectives: Monitoring Projects in Afghanistan," November/December, 2014.
139. USAID, "True Detectives: Monitoring Projects in Afghanistan," November/December, 2014; Elias Sagmeister and Julia Steets, "The use of third-party monitoring in insecure contexts: lessons from Afghanistan, Somalia, and Syria," Resource Paper, Secure Access in Volatile Environments Project, October 2016, p. 21.
140. Monitoring contractor, SIGAR interview, May 2019.
141. International development professional, SIGAR interview, April 2019.
142. Senior USAID official in RC-South, SIGAR interview, August 2016.
143. Former USAID official, SIGAR interview, January 2019.
144. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. 125, 127.
145. 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. 226.
146. 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. 9.
147. 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. 9.
148. 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. 12.
149. 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. 198.
150. 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. 199.
151. Wilton Park, "Conference report: Workshop on evaluation lessons from development assistance to Afghanistan," February 2013, pp. 2–3.
152. Trent Ruder, "Lessons and Opportunities from the Tokyo Mutual Accountability Framework," U.S. Institute of Peace, September 2015, p. 2.
153. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 1.
154. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 7.
155. Dr. Ronald Waldman, Lesley Strong, and Dr. Abdul Wali, "Afghanistan's Health System Since 2001: Condition Improved, Prognosis Cautiously Optimistic," Afghanistan Research and Evaluation Unit, December 2006, pp. 4, 8. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, pp. 1, 8.
156. Charles Michael Johnson Jr., "Afghanistan, Key Oversight Issues for USAID Development Efforts," testimony before the Subcommittee on National Security, Committee on Oversight and Government Reform, House of Representatives, March 13, 2014, p. 8.
157. CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 18; State, 18 FAM 300 Strategic Direction and Management; 18 FAM 301 Principles and Practices; 18 FAM 301.1; 18 FAM 301.1 Managing for Results (MFR) Framework, February 2019.
158. State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Invoice Review Process for Contracts in Afghanistan," Audit Report No. AUD-MERO-18-30, February 2018, p. 2.
159. Adva Saldinger, "US Congress approves long-sought Foreign Aid Transparency and Accountability Act," *Devex*, July 7, 2016, accessed December 23, 2020.
160. National Defense Authorization Act for Fiscal Year 2017, Pub. L. No 114-328, § 1205 (2016).
161. National Defense Authorization Act for Fiscal Year 2019, Pub. L. No 115-232, § 1211 (2018).
162. DOD, ODASD(SC), correspondence with SIGAR, May 13, 2021.
163. Andrew Natsios, "The Clash of the Counter-bureaucracy and Development," Center for Global Development, July 2010, p. 3.

164. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020; Sharif Hassan and Susannah George, “Rival Afghan leaders sign power-sharing agreement, breaking political deadlock,” *Washington Post*, May 17, 2020.
165. World Bank, “The World Bank in Afghanistan,” accessed Jun 23, 2020.
166. USAID, Office of Agriculture, response to SIGAR quarterly report data call, September 21, 2020.
167. USAID, Office of Agriculture, response to SIGAR quarterly report data call, September 21, 2020.
168. DOD, Combined Security Transition Command – Afghanistan, response to SIGAR quarterly report data call, September 20, 2020.
169. DOD, Combined Security Transition Command – Afghanistan, response to SIGAR quarterly report data call, September 20, 2020.
170. DOD, Combined Security Transition Command – Afghanistan, response to SIGAR quarterly report data call, September 20, 2020; USAID, Office of Education, response to SIGAR quarterly report data call, September 21, 2020; USAID, Office of Economic Growth, response to SIGAR quarterly report data call, September 21, 2020.
171. USAID, Office of Education, response to SIGAR quarterly report data call, September 21, 2020; USAID, Office of Economic Growth, response to SIGAR quarterly report data call, September 21, 2020.
172. White House, Office of the Press Secretary, “Remarks by President Biden on the Way Forward in Afghanistan,” transcript, April 14, 2021.
173. World Bank, “The World Bank in Afghanistan,” accessed June 23, 2020.
174. State Department officials, SIGAR interview, April 2019; Think tank researcher, SIGAR interview, February 2019; Sarah Rose, “Establishing USAID as a Leader in Evidence-Based Foreign Aid,” December 2020, pp. 1–2; SIGAR analysis of USAID, “ADS Chapter 201 Program Cycle Operational Policy,” June 11, 2019.
175. SIGAR analysis of USAID, response to quarterly report data call, April 13, 2021. SIGAR conclusion based on the totality of research on USAID’s M&E policies and practices conducted for this report.
176. See: 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; SIGAR, *Afghanistan’s Health Care Sector: USAID’s Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017; SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015; SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program’s Achievements and the Afghan Government’s Ability to Sustain Them*, SIGAR-18-69-AR, September 7, 2018; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 31, 2017; SIGAR, *Afghanistan’s Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, May 1, 2019.
177. Reflecting the focus on implementation, the only two major reconstruction programs initiated by USAID before the first anniversary of the Bonn Agreement in December 2002 were awarded without fully competitive bidding procedures, authorized by waivers obtained by the agency that allowed it to award contracts for the programs within weeks. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, p. ii; GAO, “Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004, pp. 48, 51, 54; SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, p. 33.
178. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, p. 99; GAO, “Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004, p. 48.
179. GAO, “Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004, p. 54.
180. Former USAID mission director, SIGAR interview, June 2019.
181. Former USAID mission director, SIGAR interview, June 2019; James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, pp. 101–102.
182. GAO, “Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004, pp. 45–46.
183. The aforementioned June 2004 GAO report describes at least two problems that plagued USAID throughout the two-decade reconstruction effort: overworked staff too few in number to adequately perform their monitoring responsibilities (including an acting mission director who was also responsible for overseeing what was then the agency’s largest project in Afghanistan), and movement restrictions due to insecurity that precluded access to field sites. At the same time, the report notes, without any obvious irony, “USAID considers monitoring of its assistance effort critical for measuring the achievement of program outputs and results.” GAO, “Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004, p. 46.

184. William Hammink, “USAID in Afghanistan Challenges and Successes,” U.S. Institute of Peace, December 2017, p. 6; Former USAID mission director, SIGAR interview, June 2019; James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, p. 103.
185. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008 pp. 101–103; GAO, “Afghanistan Reconstruction: Despite Some Progress, Deteriorating Security and Other Obstacles Continue to Threaten Achievement of U.S. Goals,” GAO-05-742, July 2005, electronic p. 2.
186. Checchi and Company Consulting, Inc., “Final Report: Services under Program and Project Offices for Results Tracking Project (SUPPORT),” April 2013, pp. 1, 4.
187. SUPPORT II was scheduled to be completed in January 2019, but was extended until July due to the award protest of the follow-on AMELA project. Checchi and Company Consulting, “Services Under Program and Project Offices for Results Tracking Phase II, Final Report, July 5, 2012 To January 4, 2019,” December 2019, pp. 5–6.
188. Jacquelyn Williams-Bridgers, “Foreign Operations: Key Issues for Congressional Oversight,” testimony before the Subcommittee on State, Foreign Operations, and Related Programs, Committee on Appropriations, House of Representatives, March 3, 2011, electronic p. 2; Checchi and Company Consulting, Inc., “Final Report: Services under Program and Project Offices for Results Tracking Project (SUPPORT),” April 2013, p. 2; U.S. Senate Committee on Foreign Relations, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations, U.S. Senate, 112th Congress, 1st session,” June 8, 2011, pp. 14–16.
189. USAID, “Accountable Assistance for Afghanistan,” April 18, 2011, p. 3.
190. USAID, “Accountable Assistance for Afghanistan,” April 18, 2011, pp. 3–4.
191. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs Throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, pp. 1–2, 5.
192. USAID, “Evaluation: Learning from Experience,” January 2011, updated October 2016, p. 1.
193. U.S. Senate Committee on Foreign Relations, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report,” June 8, 2011, p. 16.
194. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 42, 131.
195. SIGAR, *Oversight Access Inquiry Letter to Department of Defense, Department of State and U.S. Agency for International Development*, SIGAR-14-4-SP, October 10, 2013, pp. 1–2; John F. Sopko, “Improving the Effectiveness of U.S. Reconstruction Efforts in Afghanistan by Enhancing Oversight and Addressing Key Areas of High Risk,” testimony before the Senate Committee on Appropriations, Subcommittee on the Department of State, Foreign Operations, and Related Programs, February 24, 2015, p. 15.
196. USAID, “True Detectives: Monitoring Projects in Afghanistan,” November/December, 2014.
197. USAID, “True Detectives: Monitoring Projects in Afghanistan,” November/December, 2014; International development professional, SIGAR interview, April 2019; monitoring contractor, SIGAR interview, May 2019.
198. Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 10; evaluation expert, SIGAR interview, August 2018; Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 10.
199. As USAID’s 2015 to 2018 strategy for Afghanistan asserted, “These key services—particularly in health and education—are the most highly rated elements of government effectiveness. Delivery of these key services, along with security and the equitable rule of law, will ultimately generate increased confidence in the legitimacy and effectiveness of the government and, in turn, will foster stability.” USAID, “The USAID/Afghanistan Plan for Transition 2015–2018,” August 2, 2016, p. 3.
200. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” partial revision dated June 11, 2019, p. 156.
201. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities,” October 2017, p. i; USAID, “Power Transmission Expansion and Connectivity (PTEC) Project,” Fact Sheet, updated May 7, 2019, accessed October 20, 2020.
202. USAID, “Glossary of ADS Terms,” partial revision dated April 18, 2018, p. 8.
203. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities,” October 2017, p. i.
204. Radha Iyengar, Jacob N. Shapiro, and Stephen Hegarty, “Lessons Learned from Stabilization Initiatives in Afghanistan: A Systematic Review of Existing Research,” RAND Corporation, April 2017, pp. 1, 14.
205. Local perceptions were not measured directly by SIKA. Rather, SIKA relied on surveys carried out by USAID’s MISTI project. DAI, “Stability in Key Areas (SIKA) – North, Performance Management Plan,” August 2014, p. 12; USAID, “Stability in Key Areas (SIKA),” factsheet, May 7, 2019, accessed February 12, 2021.
206. SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, p. 52.
207. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 20.
208. The Asia Foundation, “Afghanistan in 2019: A Survey of the Afghan People – Frequently Asked Questions,” December 2019; Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 20.

209. The Asia Foundation, “Afghanistan in 2019: Survey of the Afghan People,” November 2019, p. 257; Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 20.
210. For example, see: DOD, Office of the Assistant Secretary of Defense, Asian and Pacific Security Affairs, “Comments to SIGAR Draft Audit Report,” correspondence with SIGAR in response to SIGAR’s draft CERP audit, April 27, 2018, p. 2; The Asia Foundation, “Afghanistan in 2019: Survey of the Afghan People,” November 2019, pp. 35, 66, 137.
211. UK Department for International Development, “Data Quality Assessment of the Asia Foundation (TAF) Surveys,” September 2010, pp. ii, 1–2.
212. UK Department for International Development, “Data Quality Assessment of the Asia Foundation (TAF) Surveys,” September 2010, p. ii.
213. UK Department for International Development, “Data Quality Assessment of the Asia Foundation (TAF) Surveys,” September 2010, p. 1.
214. UK Department for International Development, “Data Quality Assessment of the Asia Foundation (TAF) Surveys,” September 2010, pp. ii, 1–2.
215. UK Department for International Development, “Data Quality Assessment of the Asia Foundation (TAF) Surveys,” September 2010, p. ii.
216. Dyan Mazurana, Lacey Andrews Gale, and Karen Jacobsen, eds., *Research Methods in Conflict Settings: A View from Below* (Cambridge, UK: Cambridge University Press, 2013), pp. 9, 11.
217. UK Department for International Development, “Data Quality Assessment of the Asia Foundation (TAF) Surveys,” September 2010, p. 1; Dyan Mazurana, Lacey Andrews Gale, and Karen Jacobsen, eds., *Research Methods in Conflict Settings: A View from Below* (Cambridge, UK: Cambridge University Press, 2013), pp. 9, 11.
218. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, pp. 9, 11.
219. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 36.
220. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 15.
221. ARD is a consortium of German public broadcasters. ABC News/BBC/ARD National Survey of Afghanistan, “Views Improve Sharply in Afghanistan, Though Criticisms of the U.S. Stay High,” January 11, 2010, accessed April 14, 2020; Ryan Grim, “Experts on Afghanistan Doubt Survey on Foreign Occupation: Results are Impossible,” *Huffington Post*, March 18, 2010, accessed November 7, 2019.
222. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, pp. 20, 24, 40–41.
223. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 17.
224. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 40.
225. Checchi and Company Consulting, Inc., “Meta-Analysis of Final Evaluations of USAID/Afghanistan Projects, 2010-2015,” January 2016, pp. 11, 13.
226. Checchi and Company Consulting, Inc., “Meta-Analysis of Final Evaluations of USAID/Afghanistan Projects, 2010-2015,” January 2016, p. 14.
227. Evaluation expert, SIGAR interview, August 2018.
228. USAID, “USAID/Afghanistan Plan for Transition Strategy, 2015–2018: Performance Management Plan,” November 2017, pp. 1, 11.
229. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Assistance in Building Afghanistan by Developing Enterprises Program,” Audit Report No. F-306-16-002-P, March 14, 2016, p. 9; SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, p. 13; Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 11.
230. USAID implementing partner chief of party, SIGAR interview, May 2019.
231. USAID, “Glossary of ADS Terms,” April 18, 2018, p. 192.
232. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” June 11, 2019, p. 151.
233. State, response to SIGAR’s draft audit of U.S. rule-of-law assistance in Afghanistan, May 19, 2015, p. 2.
234. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” June 11, 2019, p. 46.
235. USAID, Office of Afghanistan and Pakistan Affairs, “Fixed Amount Award No. 72030618FA00006, under Annual Program Statement (APS) Number APS 306-17-000003 and RFA-306-17-0000012 (Establishing Kabul Carpet Export Center (KCEC)),” June 6, 2018, p. 11.
236. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 126; Stephen Downes-Martin, “Operations Assessment in Afghanistan is Broken—What Is to Be Done?,” *Naval War College Review*, vol. 64, no. 4 (2011), p. 13; Andrew Natsios, “The Clash of the Counter-bureaucracy and Development,” Center for Global Development, July 2010, pp. 2, 24.
237. USAID senior official, SIGAR interview, December 2018.

238. Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*, vol. 64, no. 4, 2011, pp. 7–8.
239. See: SIGAR, *Afghanistan's Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, May 2019, pp. ii, 14–15; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 56.
240. USAID, "Kandahar Food Zone Program (KFZ) – Year 3 Monitoring & Evaluation Plan, Version 2," May 2016, p. 4, 10–11.
241. USAID, "Kandahar Food Zone Program (KFZ) – Year 3 Monitoring & Evaluation Plan, Version 2," May 2016, pp. 4, 10–11.
242. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 117.
243. 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, p. 1.
244. 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.
245. 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, p. 5; David Mansfield, "Examining the Impact of IDEA-NEW on Opium Production Nangarhar: A Case Study," January 2015.
246. 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, p. 5.
247. David Mansfield, "Examining the Impact of IDEA-NEW on Opium Production Nangarhar: A Case Study," January 2015, p. 12.
248. David Mansfield, "Examining the Impact of IDEA-NEW on Opium Production Nangarhar: A Case Study," January 2015, p. 73.
249. Evaluation specialist, SIGAR interview, April 2019; former USAID official, SIGAR interview, June 2019; USAID implementing partner staff, SIGAR interview, May 2019; M&E contractor, SIGAR interview, May 2019.
250. USAID, "ADS Chapter 201 Program Cycle Operational Policy," July 23, 2020, p. 42.
251. Evaluation specialist, SIGAR interview, April 2019.
252. Former USAID official, SIGAR interview, June 2019.
253. Former USAID official, SIGAR interview, January 2019; State, "U.S. Foreign Assistance Reform: Achieving Results & Sustainability in Support of Transformational Diplomacy," June 2006.
254. State, "Standard Foreign Assistance Indicators," n.d., accessed September 1, 2020.
255. Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, "Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change," May 2009, p. 13; M&E contractor, SIGAR interview, May 2019; implementing partner staff, SIGAR interview, May 2019; Andrew Natsios, "The Clash of the Counter-bureaucracy and Development," Center for Global Development, July 2010, p. 33.
256. State, "FY 2019 Master Indicator List," 2018, accessed September 1, 2020.
257. USAID implementing partner, SIGAR interview, May 2019.
258. As a 2015 Mercy Corps impact evaluation summarized, "Promoting stability is one of the primary objectives of nearly all major development actors operating in Afghanistan." Mercy Corps, "Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan," January 2015, p. 7.
259. State, USAID, and DOD, "Stabilization Assistance Review: A Framework for Maximizing the Effectiveness of U.S. Government Efforts to Stabilize Conflict-Affected Areas," 2018, p. 4.
260. U.S. Mission Afghanistan, "U.S. Foreign Assistance for Afghanistan Post Performance Management Plan-2011–2015," October 2010, pp. Annex I-2, Annex II-3; Annex III-3; Annex V-3–4; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 119.
261. USAID, "Plan for Transition 2015–2018," August 2016, p. 3.
262. USAID, "Measuring Impacts of Stabilization Initiatives (MISTI)," October 25, 2013; USAID, response to SIGAR data call, January 1, 2016; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 117.
263. Management Systems International, "MISTI Stabilization Trends and Impact Evaluation Survey Analytical Report, Wave 5: Sep 28 – Nov 3, 2014," April 28, 2015, pp. 3, 6–7, 9, 13; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 119.
264. State, "Afghanistan and Pakistan Regional Stabilization Strategy," February, 2010, p. i; U.S. Mission Afghanistan, "U.S. Foreign Assistance for Afghanistan Post Performance Management Plan-2011–2015," October 2010, pp. 2–4; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 119.
265. U.S. Mission Afghanistan, "U.S. Foreign Assistance for Afghanistan Post Performance Management Plan-2011–2015," October 2010, p. 9; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 119.
266. USAID, Office of Democracy and Governance, response to SIGAR quarterly report data call, December 28, 2015.

267. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. i.
268. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 2.
269. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 10.
270. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” July 23, 2020, p. 146.
271. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” July 23, 2020, p. 37.
272. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” July 23, 2020, p. 38.
273. SIGAR conclusion based on USAID, “ADS Chapter 201: Program Cycle Operational Policy,” July 23, 2020, pp. 38–39.
274. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 10; SIGAR analysis of USAID, “ADS Chapter 201: Program Cycle Operational Policy,” 23, 2020, pp. 38–39.
275. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 16.
276. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 28.
277. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 26.
278. The presumed connection between employment and insurgent recruitment has been the basis of much aid spending in unstable environments. As one study summarized, “The logic is that gainfully employed young men are less likely to participate in political violence.” Although such logic is intuitively plausible, it has been challenged. For example, the same study, which tested the prediction that higher levels of employment correlate with lower levels of insurgent violence using data from Iraq, Afghanistan, and the Philippines, found, “the data emphatically reject a positive correlation between unemployment and attacks against government and allied forces.” The implication was that “aid and development efforts that seek to enhance political stability through short-term job creation programs may well be misguided.” Eli Berman, Michael Callen, Joseph H. Felter and Jacob N. Shapiro, “Do Working Men Rebel: Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines,” *Journal of Conflict Resolution*, March 22, 2011, vol. 55, no. 4, pp. 496, 519; Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 20.
279. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 13.
280. Kartika Bhatia and Hafez Ghanem, “How Do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries,” Brookings Institution, March 22, 2017, p. 1.
281. Kartika Bhatia and Hafez Ghanem, “How Do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries,” Brookings Institution, March 22, 2017, pp. 2–3.
282. Kartika Bhatia and Hafez Ghanem, “How Do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries,” Brookings Institution, March 22, 2017, pp. 4–5.
283. Kartika Bhatia and Hafez Ghanem, “How Do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries,” Brookings Institution, March 22, 2017, p. 31.
284. For example, one of the variables used to measure support for violent extremism was a question about how justified the 9/11 attacks on the World Trade Center were, measured on a scale from 1 to 5. The coefficient on the interaction term “unemployed/underemployed* secondary” in the first reported ordinary least squares regression results was 0.308.
285. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 13.
286. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 21.
287. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 5.
288. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 26.
289. Stephen J. Hadley, Andrew Wilder, and Scott Worden, “Four steps to winning peace in Afghanistan,” *Washington Post*, April 27, 2017.
290. Other examples of uncited statements in the Country Development Cooperation Strategy that claim a causal relationship between employment and extremism include: “The development objectives are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address the security and development challenges that have made Afghanistan a safe haven for terrorists and violent extremist organizations and have driven the conflict with the Taliban, including corruption, unemployment, and a lack of government legitimacy,” (p. 10); “Afghanistan’s high unemployment and underemployment . . . hinders efforts to achieve gender equality, is a potential driver of extremism, and can lead to other negative socioeconomic impacts,” (p. 35); “Another consideration is that the struggle to find a job and the low social status associated with unemployment can lead to economic marginalization and susceptibility to violent extremism, particularly among young women and men,” (p. 36); “USAID also sees secondary benefits from job creation efforts among youth as increased employment and employability give Afghans a safer and more lucrative alternative to working with insurgents,” (p. 28). USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, pp. 9–10, 28, 35–36.
291. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 4.
292. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 6.
293. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, pp. 4–5.

294. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 28.
295. The cited sentence in the CDCS that makes this claim cites to footnote 52. However, the CDCS makes multiple references to this link. For example, in the section describing USAID’s approach near the beginning, the CDCS makes essentially the same statement, along with the specific claim that “increasing the [Afghan government’s] accountability to its citizens will also help improve the country’s stability.” The entire third development objective is focused on increasing government accountability with the goal of increasing self-reliance and stability. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, pp. 9–10, 26, 47.
296. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 26.
297. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 29.
298. World Trade Organization, “The WTO can . . . stimulate economic growth and employment,” accessed August 12, 2020.
299. World Trade Organization, “The WTO can . . . stimulate economic growth and employment,” accessed August 12, 2020. The UN recognizes Afghanistan as a “Least Developed Country,” and of the countries for which the World Bank was able to collect data in 2019, Afghanistan had the sixth lowest GDP per capita and the lowest rate of exports as a percent of GDP. Additionally, the World Bank ranked Afghanistan third to last in its aggregate Logistics Performance Index from 2012 to 2018, which includes factors such as infrastructure, international shipments, and logistics competence. UN Committee for Development Policy, “List of Least Developed Countries (as of December 2018),” December, 2018, p. 1; World Bank, “GDP per capita (current US\$),” accessed August 24, 2020; The World Bank, “Exports of goods and services (% of GDP),” accessed August 24, 2020; World Bank, “Aggregated LPI 2012–2018,” accessed August 24, 2020.
300. World Trade Organization, “The WTO can . . . stimulate economic growth and employment,” accessed August 12, 2020.
301. World Trade Organization, “The WTO can . . . stimulate economic growth and employment,” accessed August 12, 2020.
302. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 38.
303. SIGAR, correspondence with State, USAID, and DOD, July 5, 2013, electronic pp. 2–3; USAID, correspondence with SIGAR, May 9, 2013, pp. 1–2; Greg Huger, testimony before the U.S. House Foreign Affairs Committee, accessed December 21, 2020.
304. USAID, correspondence with SIGAR, May 9, 2013, pp. 1–2.
305. USAID, correspondence with SIGAR, May 9, 2013, p. 1.
306. U.S. Senate Foreign Relations Committee, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations, United States Senate,” June 2011, pp. 2–3, 14; USAID implementing partner chief of party, SIGAR interview, August 2018; evaluation expert, SIGAR interview, August 2018.
307. USAID implementing partner chief of party, SIGAR interview, August 2018.
308. Fahim Abed, “How the U.S. Government Misleads the Public on Afghanistan,” *New York Times*, September 8, 2018; John F. Sopko, “Examining the Trump Administration’s Afghanistan Strategy,” testimony before the House Committee on Homeland Security and Government Affairs Subcommittee on National Security, January 28, 2020, p. 15; John F. Sopko, “U.S. Lessons Learned in Afghanistan,” testimony before the House Committee on Foreign Affairs, January 15, 2020, electronic p. 7.
309. John F. Sopko, “Examining the Trump Administration’s Afghanistan Strategy,” testimony before the House Committee on Homeland Security and Government Affairs Subcommittee on National Security, January 28, 2020, electronic p. 15.
310. USAID, “USAID in Afghanistan: Partnership, Progress, Perseverance,” March 2012, p. 5; SIGAR, *Afghanistan’s Health Care Sector: USAID’s Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 2017, p. i; John F. Sopko, “U.S. Lessons Learned in Afghanistan,” testimony before the House Committee on Foreign Affairs, January 15, 2020, electronic p. 15.
311. In addition to the examples presented in this paragraph and the next, USAID’s more recent claim that its export-led development strategy was succeeding may not have been accurate. SIGAR demonstrated that two of USAID’s key assertions—that Afghanistan’s merchandise exports totaled \$1 billion in 2018 and that Afghanistan’s airborne exports exceeded \$500 million in the same year—were based on flawed methodology and were therefore misleading. In response to SIGAR’s concerns, USAID removed all references to these figures in its public products. For more, see: SIGAR, *Quarterly Report to the United States Congress*, July 30, 2019, pp. 141–143.
312. Rajiv Shah, “Testimony by USAID Administrator Dr. Rajiv Shah on the President’s Fiscal Year 2011 Budget Requests,” before the House Subcommittee on State, Foreign Operations, and Related Programs; Committee on Appropriations, March 4, 2010, accessed May 10, 2021; Greg Huger, testimony before the House Foreign Affairs Committee, November 8, 2017, accessed December 21, 2020; USAID, “Afghanistan–Education,” September 10, 2019, accessed May 4, 2020.
313. USAID, “Afghanistan–Education,” September 10, 2019, accessed May 4, 2020.
314. SIGAR, *Quarterly Report to the United States Congress*, July 30, 2017, p. 182.

315. As Inspector General Sopko said in January 2020 testimony before the Congress, “We talk about some of the statements made by AID about the great success on life expectancy. It was statistically impossible to double the life expectancy in the time given. I think it’s a combination of hubris and mendacity that anybody can do that. I mean, the next thing you know is we are going to be walking on water on an aid program.” John F. Sopko, “U.S. Lessons Learned in Afghanistan,” testimony before the House Committee on Foreign Affairs, January 15, 2020, electronic p. 15; SIGAR, *Afghanistan’s Health Care Sector: USAID’s Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 2017, pp. i, 7; Government of Afghanistan, NSIA, “Afghanistan Demographic and Health Survey 2015,” January 2017, p. ii; USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 36.
316. U.S. Senate Foreign Relations Committee, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations, United States Senate,” June 2011, pp. 2–3, 14; evaluation expert, SIGAR interview, August 2018; DEVEX, “Special Feature: A history of American public opinion on foreign aid,” August 15, 2017.
317. Evaluation expert, SIGAR interview, August 2018; former USAID official, SIGAR interview, January 2019; evaluation consultant, SIGAR interview, January 2019; evaluation expert, SIGAR interview, April 2019.
318. Evaluation consultant, SIGAR interview, January 2019; Yogesh Rajkotia, “Beware of the success cartel: a plea for rational progress in global health,” *BMJ Global Health*, November 21, 2018, pp. 1–2; Commission on Wartime Contracting in Iraq and Afghanistan, “Transforming Wartime Contracting: Controlling Costs, Reducing Risks,” August 2011, p. 33.
319. William Easterly, “The Cartel of Good Intentions,” *Foreign Policy*, November 11, 2009; Yogesh Rajkotia, “Beware of the success cartel: a plea for rational progress in global health,” *BMJ Global Health*, November 21, 2018, p. 1.
320. Yogesh Rajkotia, “Beware of the success cartel: a plea for rational progress in global health,” *BMJ Global Health*, November 21, 2018, pp. 1–2.
321. Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, “Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change,” May 2009, pp. ii–iv.
322. Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, “Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change,” May 2009, p. iv.
323. Evaluation expert, SIGAR interview, August 2018.
324. Former USAID official, SIGAR interview, January 2019; Matt Millham, “Despite lingering reliance on U.S. aid, Afghanistan’s Nawa district a model of ‘transition’,” *New York Times*, May 28, 2011; Frances Z Brown, “Afghanistan’s need for reform: We have seen the enemy, and it is our anecdotes,” *Foreign Policy*, October 2012.
325. Former USAID official, SIGAR interview, January 2019.
326. Frances Z. Brown, “Afghanistan’s need for reform: We have seen the enemy, and it is our anecdotes,” *Foreign Policy*, October 2012; Management Systems International and D3 Systems, Inc., “MISTI Stabilization Trends and Impact Evaluation Survey Analytical Report, Wave 5: Sep 28 – Nov 3, 2014,” November 2015.
327. International development professional, SIGAR interview, April 2019.
328. USAID implementing partner chief of party, SIGAR interview, May 2019.
329. 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; SIGAR, *Afghanistan’s Health Care Sector: USAID’s Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017, p. i.
330. USAID, “Fact Sheet: Measuring Impacts of Stabilization Initiatives,” October 2013, accessed February 2, 2018.
331. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 228–229.
332. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 131–132, 136.
333. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 229.
334. James Risen, “Afghans See Taliban as a Key to U.S. Aid Projects, Study Finds,” *New York Times*, December 5, 2015; SIGAR, *Quarterly Report to the United States Congress*, October 30, 2015, pp. 143–144.
335. Management Systems International and D3 Systems, Inc., “MISTI Stabilization Trends and Impact Evaluation Survey Analytical Report, Wave 5: Sep 28 – Nov 3, 2014,” November 2015.
336. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 129.
337. Evaluation specialist, SIGAR interview, April 2019.
338. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 229.
339. Former USAID official, SIGAR interview, January 2019.
340. USAID implementing partner official, SIGAR interview, May 2019.
341. USAID implementing partner official, SIGAR interview, May 2019.
342. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 6.

343. Implementing partner staff, SIGAR interview, May 2019; USAID implementing partner chief of party, SIGAR interview, May 2019; USAID, “Afghanistan Trade Event Drives Over \$55 million in New Business,” October 2, 2019.
344. Checchi and Company Consulting, Inc., “Assessment of the Passage to Prosperity: India Afghanistan Trade and Investment Show,” December 2018, p. x.
345. Checchi and Company Consulting, Inc., “Assessment of the Passage to Prosperity: India Afghanistan Trade and Investment Show,” December 2018, pp. x–xi.
346. USAID implementing partner official, SIGAR interview, May 2019.
347. USAID implementing partner official, SIGAR interview, May 2019.
348. Paul Fishstein and Andrew Wilder, “Winning Hearts and Minds? Examining the Relationship between Aid and Security in Afghanistan,” Feinstein International Center, Tufts University, January 2012, p. 6; Radha Iyengar, Jacob N. Shapiro, and Stephen Hegarty, “Lessons Learned from Stabilization Initiatives in Afghanistan: A Systematic Review of Existing Research,” RAND Corporation, April 2017, p. 22.
349. Andrew Natsios, “The Clash of the Counter-bureaucracy and Development,” Center for Global Development, July 2010, p. 3.
350. Former USAID official, SIGAR interview, June 2019.
351. Paul Fishstein, Anna Paterson, and Alan Roe, “Mapping of Research Capacity in Afghanistan,” submitted to UK Department for International Development, September 2012, p. 18.
352. World Bank, “World Bank Releases Randomized Impact Evaluation of Afghanistan’s National Solidarity Programme,” September 29, 2013, accessed June 8, 2020; Dana Burde and Leigh L. Linden, “Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools,” *American Economic Journal: Applied Economics* 2013, vol. 5, no. 3, pp. 27–40.
353. Senior evaluation consultant, SIGAR interview, January 2019.
354. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. 4.
355. Former USAID official, SIGAR interview, January 2019.
356. Former USAID official, SIGAR interview, January 2019.
357. UK Department for International Development, “Prime Minister announces merger of Department for International Development and Foreign Office,” June 16, 2020; Oxford Policy Management, “Independent Evaluation of African Risk Capacity (ARC): Final Inception Report,” January 2017, p. 38.
358. Oxford Policy Management, “Independent Evaluation of African Risk Capacity (ARC): Final Inception Report,” January 2017, p. 38.
359. Implementing partner chief of party, SIGAR interview, August 2019. These assessments had been included in the RFP issued by USAID.
360. Former USAID official, SIGAR interview, January 2019.
361. Senior evaluation consultant, SIGAR interview, January 2019; evaluation expert, SIGAR interview, May 2019; implementing partner staff, SIGAR interview, May 2019.
362. USAID official, SIGAR interview, May 2019.
363. USAID official, SIGAR interview, May 2019; implementing partner chief of party, SIGAR interview, May 2019.
364. USAID officials, SIGAR interview, February 2019.
365. Implementing partner chief of party, SIGAR interview, May 2019; Monitoring Support Program contractor staff, SIGAR interview, July 2019; implementing partner chief of party, SIGAR interview, May 2019; U.S. Senate Committee on Foreign Relations, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations,” 112th Congress, 1st session, June 8, 2011, p. 2.
366. Implementing partner chief of party, SIGAR interview, May 2019.
367. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, p. 5.
368. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs Throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, p. 5.
369. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs Throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, pp. 2, 5, 7.
370. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs Throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, p. 5.
371. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs Throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, p. 6.
372. USAID Office of Inspector General, “Audit of USAID/Afghanistan’s Strategy for Monitoring and Evaluating Programs Throughout Afghanistan,” Audit Report No. F-306-16-001-P, December 10, 2015, p. 6.
373. In response to a draft of this report, USAID stated, “In 2015, Afghan Info underwent substantial process and system improvements to include mandatory reporting per the Performance Monitoring Mission Order (PMMO), system dashboard upgrades to include quarterly status updates, and quarterly compliance report submissions to senior leadership.” USAID, “USAID Afghanistan suggested changes/clarifications to SIGAR Lessons Learned Draft Report LL-11 Monitoring and Evaluation of Contracting,” April 27, 2021.
374. SIGAR’s quarterly report team uses the database as a source of periodic project documents. The team that produced SIGAR’s lessons-learned report on U.S. support for gender equality also used the database. USAID

- implementing partner staff, SIGAR interview, May 2019; USAID implementing partner staff, SIGAR interview, May 2019.
375. USAID implementing partner staff, SIGAR interview, May 2019.
 376. USAID implementing partner staff, SIGAR interview, May 2019.
 377. SIGAR observation based on its experience with Afghan Info.
 378. USAID, "Evaluation: Learning from Experience," October 2016, p. 1.
 379. USAID, "Evaluation: Learning from Experience," October 2016, p. 1.
 380. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs Throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, pp. 1–2, 4; SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 6; SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, p. 4; Checchi and Company Consulting, Inc., "Meta-Analysis of Final Evaluations of USAID/Afghanistan Projects, 2010-2015," January 2016, p. 13; international development professional, SIGAR interview, April 2019; senior evaluation consultant, SIGAR interview, January 2019; Monitoring Support Program contractor staff, SIGAR interview, July 2019.
 381. Implementing partner staff, SIGAR interview, May 2019; former USAID official, SIGAR interview, June 2019; USAID officials, SIGAR interview, February 2019.
 382. Evaluation expert, SIGAR interview, May 2019; USAID officials, SIGAR interview, February 2019.
 383. GAO, "Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed," GAO-04-403, June 2004, p. 46; SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, p. 12.
 384. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, p. 12; Ben Arnoldy, "Afghanistan War: How USAID loses hearts and minds," *Christian Science Monitor*, July 28, 2010, accessed May 26, 2020.
 385. USAID officials, SIGAR interview, May 2019; SIGAR, *High-Risk List*, January 2017, pp. 47–49; John F. Sopko, "Improving the Effectiveness of U.S. Reconstruction Efforts in Afghanistan by Enhancing Oversight and Addressing Key Areas of High Risk," statement for the record, Senate Committee on Appropriations, Subcommittee on the Department of State, Foreign Operations, and Related Programs, February 24, 2015, p. 15.
 386. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 1.
 387. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 1.
 388. According to USAID, "During the clearance process of the Mission Order, concerns were raised that the word 'tier' may not be the most accurate term to describe the mission's monitoring approach, since the approach does not preference one tier over the others. However, due to the extensive usage and socialization of the 'tier' concept among internal and external stakeholders, and because the monitoring approach described in this mission order directly iterates on the current MTM approach, the Mission has opted to continue to use the word 'tier' at this time. Alternative word choices for 'tier' will be explored going forward, as deemed necessary by mission management." USAID, "Mission Order on Performance Monitoring 201.05," September 20, 2017.
 389. SIGAR interpretation based on USAID, "True Detectives: Monitoring Projects in Afghanistan," November/December, 2014.
 390. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs Throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 2.
 391. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs Throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 7.
 392. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs Throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, p. 2.
 393. Senior evaluation consultant, SIGAR interview, January 2019.
 394. USAID staff from the Office of Project and Program Development, SIGAR interview, May 2019.
 395. USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, pp. 1, 8; John F. Sopko, "Oversight of USAID and State Department Reconstruction Efforts in Afghanistan," testimony before the House Subcommittee on the Middle East and North Africa, Committee on Foreign Affairs, June 10, 2014, electronic pp. 38–39.
 396. USAID, "True Detectives: Monitoring Projects in Afghanistan," November/December, 2014.
 397. For discussion of the differences in approach, see Dr. Ronald Waldman, Lesley Strong, and Dr. Abdul Wali, "Afghanistan's Health System Since 2001: Condition Improved, Prognosis Cautiously Optimistic," Afghanistan Research and Evaluation Unit, December 2006, pp. 1–2, 4, 15.

398. David H. Peters, Ayan Ahmed Noor, Lackwinder P. Singh, Faizullah K. Kakar, Peter M. Hansen, and Gilbert Burnham, "A balanced scorecard for health services in Afghanistan," *Bulletin of the World Health Organization*, vol. 85, February 2007, pp. 146, 150.
399. David H. Peters, Ayan Ahmed Noor, Lackwinder P. Singh, Faizullah K. Kakar, Peter M. Hansen, and Gilbert Burnham, "A balanced scorecard for health services in Afghanistan," *Bulletin of the World Health Organization*, vol. 85, February 2007, pp. 147, 148.
400. David H. Peters, Ayan Ahmed Noor, Lackwinder P. Singh, Faizullah K. Kakar, Peter M. Hansen, and Gilbert Burnham, "A balanced scorecard for health services in Afghanistan," *Bulletin of the World Health Organization*, vol. 85, February 2007, p. 149.
401. David H. Peters, Ayan Ahmed Noor, Lackwinder P. Singh, Faizullah K. Kakar, Peter M. Hansen, and Gilbert Burnham, "A balanced scorecard for health services in Afghanistan," *Bulletin of the World Health Organization*, vol. 85, February 2007, pp. 149–150.
402. David H. Peters, Ayan Ahmed Noor, Lackwinder P. Singh, Faizullah K. Kakar, Peter M. Hansen, and Gilbert Burnham, "A balanced scorecard for health services in Afghanistan," *Bulletin of the World Health Organization*, vol. 85, February 2007, pp. 149–150.
403. USAID, "Monitoring Support Project Update," June 30, 2017.
404. USAID, "Task Order 3 for Monitoring Support Program," July 2015.
405. Monitoring support program contractor staff, SIGAR interview, July 2019; USAID, "Task Order 3 for Monitoring Support Program," July 2015, p. 18; SIGAR, *USAID's Eastern Provinces Monitoring under the Monitoring Support Project: Audit of Costs Incurred by The QED Group LLC*, SIGAR-18-74-FA/MSP, September 2018, p. 2.
406. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
407. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
408. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
409. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
410. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
411. Senior evaluation consultant, SIGAR interview, January 2019.
412. Afghan anti-corruption civil society organization, SIGAR interview, May 2019.
413. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, p. 13.
414. SIGAR, *2021 High-Risk List*, February 19, 2021, p. 59.
415. Julia Steets, Elias Sagmeister and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," *Secure Access in Volatile Environments (SAVE)*, October 2016, p. 38.
416. Julia Steets, Elias Sagmeister and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," *Secure Access in Volatile Environments (SAVE)*, October 2016, p. 38.
417. USAID implementing partner, SIGAR interview, May 2019.
418. USAID, "Contract for Afghanistan Monitoring, Evaluation, and Learning Activity (AMELA)," March 13, 2019.
419. USAID implementing partner, SIGAR interview, May 2019.
420. At more than \$2.4 billion, USAID's investment in programming that explicitly aimed to increase stability was substantial. These programs were intended to improve the lives of local populations and to extend the reach of the Afghan government to unstable areas. Even after USAID ceased stabilization programming in 2016, its longer-term development programs maintained stability-related objectives. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2020, p. 152; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 119; USAID, "The USAID/Afghanistan Plan for Transition 2015–2018," August 2, 2016, p. 3; USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, pp. 9–10; SIGAR, *Quarterly Report to the United States Congress*, October 30, 2015, p. 144; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 118.
421. According to the final MISTI report, of the 860 villages that received USAID stability assistance, 72 villages were determined to be under Taliban control. Further, there was "a large average increase" in Taliban support in five of these 72 villages and a substantial boost in Taliban local popularity in 13 of the 72 villages. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2015, p. 144; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 118.
422. USAID, "The USAID/Afghanistan Plan for Transition 2015–2018," August 2, 2016, p. 3.
423. According to the World Bank report, Taliban policy towards Afghan government-run schools evolved over time: "The armed opposition of the Taliban identified education delivery as a key field of state activity. It regarded it as an attempt to indoctrinate and impose Western ideas on children. The resulting campaign of violence in 2006–08 led to the destruction of numerous schools and the closure of even more. The evolution of the Taliban policy toward schools helps explain the reduction in attacks. Gradually, the Taliban's campaign evolved toward an attempt to influence state schools through local-level negotiations with Ministry of Education officials." World Bank, "Social Service Delivery in Violent Contexts: Achieving Results Against the Odds—A Report from Afghanistan, Pakistan, and Nepal," June 9, 2017, pp. 63–64.
424. World Bank, "Social Service Delivery in Violent Contexts: Achieving Results Against the Odds—A Report from Afghanistan, Pakistan, and Nepal," May 30, 2017, p. 65.
425. Jackson's recent work on Taliban governance and service delivery in Taliban-controlled areas includes "Life Under the Taliban Shadow Government" and "Rebel Rule of Law: Taliban Courts in the West and North-

- West of Afghanistan,” both published under the Overseas Development Institute, a London-based think tank. According to the former report, “Where the government and aid agencies provide public goods and services, the Taliban co-opt and control them. Health and education in Taliban areas are a hybrid of NGO and state-provided services, operating according to Taliban rules.” Ashley Jackson, “The Taliban’s Fight for Hearts and Minds,” *Foreign Policy*, September 12, 2018, accessed December 21, 2020; Ashley Jackson, “Life Under the Taliban Shadow Government,” Overseas Development Institute, June 2018, p. 5.
426. As one anonymous U.S. embassy official said in 2009, “Everyone knows this is going on.” Jean MacKenzie, “Funding the Afghan Taliban,” PRI, *GlobalPost*, August 7, 2009.
 427. USAID Office of Inspector General, “Review of Security Costs Charged to USAID Projects in Afghanistan (Review Report No. 5-306-10-002-S),” September 29, 2010, p. 2.
 428. Jean MacKenzie, “Funding the Afghan Taliban,” PRI, *GlobalPost*, August 7, 2009.
 429. The review focused on a small sample of three projects implemented by Development Alternatives, Inc. (DAI), a development contractor. Jean MacKenzie, “Are US taxpayers funding the Taliban?,” PRI, *GlobalPost*, September 3, 2009; USAID Office of Inspector General, “Review of Security Costs Charged to USAID Projects in Afghanistan (Review Report No. 5-306-10-002-S),” September 29, 2010, pp. 2, 3–4, 10.
 430. USAID Office of Inspector General, “Review of Security Costs Charged to USAID Projects in Afghanistan (Review Report No. 5-306-10-002-S),” September 29, 2010, p. 4.
 431. Former U.S. official Kirk Meyer, who served in Afghanistan for five years and directed a counterterrorist finance cell, spoke directly to this and other problems, stating that by the time a local contractor received money to build a project, he “didn’t have enough money to build the road, plus pay the Taliban, plus pay the corrupt officials who usually took about 20 percent of every development project. So basically, the contractors didn’t build the projects or they built them shoddily, and because the projects were in contested areas, nobody verified whether they were built.” USAID Office of Inspector General, “Review of Security Costs Charged to USAID Projects in Afghanistan (Review Report No. 5-306-10-002-S),” September 29, 2010, p. 6; SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, p. 20.
 432. Greg Huger, testimony before the House Foreign Affairs Committee, November 8, 2017, accessed December 21, 2020.
 433. Huger stated, “USAID employs stringent oversight safeguards to protect programs from waste, fraud, and abuse and to ensure that American investments in Afghanistan and Pakistan achieve their intended impact. USAID modifies or suspends projects if performance is not on track or oversight standards cannot be met.” Greg Huger, testimony before the House Foreign Affairs Committee, November 8, 2017, accessed December 21, 2020.
 434. USAID, Office of Program and Project Development, response to SIGAR quarterly report data call, January 6, 2019.
 435. USAID, “The USAID/Afghanistan Plan for Transition 2015–2018,” August 2, 2016, p. 3.
 436. Senior USAID officials, SIGAR interview, December 2017.
 437. The question was posed to USAID’s offices of infrastructure, health and nutrition, economic growth, agriculture, and education. For the offices of infrastructure, health and nutrition, economic growth, and agriculture, the question was worded: “Please provide a list, by district and province, of where USAID [technical office] implementers are currently working.” For USAID’s office of education, the question was worded: “Please provide an updated list of districts in which [Office of Education] implementing partners are currently operating, broken out by project.” SIGAR, quarterly report RFI to USAID and State, February 21, 2019; USAID, Office of Education, response to SIGAR data call, March 21, 2019.
 438. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019.
 439. In advance of SIGAR’s April 2019 *Quarterly Report to the United States Congress*, Resolute Support mission staff formally notified SIGAR that it was no longer assessing district-level insurgent or government control or influence. The last district stability data Resolute Support produced was for its October 22, 2018 assessment; SIGAR reported on that assessment in its January 2019 *Quarterly Report to the United States Congress*. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019.
 440. According to the definitions of Resolute Support’s district stability assessment, “insurgent activity” equated to “insurgent influence.” “High insurgent activity” equated to “insurgent control.” Before SIGAR’s April 2018 *Quarterly Report to the United States Congress*, U.S. Forces–Afghanistan (USFOR-A) changed the language for insurgent control or influence to “insurgent activity” and “high insurgent activity.” According to USFOR-A, this change was not due to adopting a new methodology for district-control assessments. Rather, it was part of an effort to make its district control map unclassified and publicly releasable. The original terms “insurgent control” and “insurgent influence” were used in the numerical district-control data provided to SIGAR in advance of its April 2018 quarterly report. See: SIGAR, *Quarterly Report to the United States Congress*, April 30, 2018, p. 88; SIGAR analysis of Resolute Support response to SIGAR quarterly report data call, December 20, 2018; USAID, Office of Health and Nutrition, response to SIGAR data call, March 21, 2019; USAID, Office of Education, response to SIGAR data call, March 21, 2019; USAID, Office of Agriculture, response to SIGAR data call, March 21, 2019.
 441. USAID, “Initiative for Hygiene, Sanitation, and Nutrition (IHSAN),” June 2019, p. 1; SIGAR analysis of USAID, Office of Health and Nutrition, response to SIGAR quarterly report data call, March 21, 2019 and Resolute Support, response to SIGAR quarterly report data call, December 20, 2018.

442. SIGAR analysis of USAID, Office of Education, response to SIGAR data call, March 21, 2019 and Resolute Support response to SIGAR quarterly report data call, December 20, 2018.
443. Senior USAID officials, SIGAR interview, December 2017.
444. Resolute Support's January 2018 district stability assessment did not include a category for "high insurgent activity." Therefore, if a district was categorized in all other available assessments as "high insurgent activity" or "insurgent control," and "insurgent activity" in the January 2018 assessment, the district control status was assumed not to have changed. SIGAR analysis of USAID, Office of Health and Nutrition, response to SIGAR quarterly report data call, March 21, 2019; USAID, Office of Education, response to SIGAR data call, March 21, 2019; Resolute Support response to SIGAR quarterly report data call, December 20, 2018; Resolute Support response to SIGAR quarterly report data call, September 19, 2018; Resolute Support, response to SIGAR quarterly report data call, June 22, 2018; Resolute Support, response to SIGAR quarterly report data call, March 22, 2018; and Resolute Support, response to SIGAR quarterly report data call, December 2017.
445. The most significant restriction was introduced in the 2017 Consolidated Appropriations Act, which prohibited funds appropriated by the act from being obligated for any project that "legitimizes the Taliban or other extremist organizations." The 2018 Consolidated Appropriations Act diluted the language somewhat: Rather than explicitly prohibiting funds appropriated by the act from being obligated for projects that legitimized the Taliban, the restriction applied to any projects or activities "conducted in areas where project and resource disbursement monitoring cannot be performed, unless the Secretary of State, in consultation with the Administrator of the United States Agency for International Development" certified that such projects fulfilled U.S. national security interests. In such cases, the Congress required a report describing how the projects contributed to those interests, "including how such project or activity does not legitimize the Taliban or other extremist organizations." Subsequent appropriations acts did not include any similar language. However, the risk that projects implemented in Taliban-controlled areas could work against the stability objectives they are intended to achieve speaks to the continued relevance of this monitoring issue even in the absence of congressional prohibitions. Consolidated Appropriations Act for Fiscal Year 2017, Pub. L. No. 115-31, § 7044 (2016); Consolidated Appropriations Act for Fiscal Year 2018, Pub. L. No. 115-141, § 7044 (2017); Consolidated Appropriations Act for Fiscal Year 2019, Pub. L. No. 116-6, § 7044 (2019); Further Consolidated Appropriations Act for Fiscal Year 2020, Pub. L. No. 116-44, § 7044 (2019).
446. Voice of America, "Rare UN-Taliban Agreement to Set Up 4,000 Schools in Insurgent-Held Afghan Territory," December 19, 2020; USAID, "Increasing Access to Basic Education and Gender Equality (CBE)," March 2019, p. 1.
447. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2020, p. 110.
448. Checchi and Company Consulting, Inc., "Final Report: Services under Program and Project Offices for Results Tracking Project (SUPPORT)," April 2013, pp. 1, 4.
449. Jacquelyn Williams-Bridgers, "Foreign Operations: Key Issues for Congressional Oversight, Testimony Before the Subcommittee on State, Foreign Operations, and Related Programs, Committee on Appropriations, House of Representatives," March 3, 2011, electronic p. 2; Checchi and Company Consulting, Inc., "Final Report: Services under Program and Project Offices for Results Tracking Project (SUPPORT)," April 2013, p. 2.
450. Additional functions were performance management, conducting field-based assessments, and supporting technical offices in new program and project design. USAID, "Contract for SUPPORT II," July 5, 2012, p. 9; Checchi and Company Consulting, "Services Under Program and Project Offices for Results Tracking Phase II, Final Report, July 5, 2012 To January 4, 2019," December 2019, p. 6.
451. Former USAID official, SIGAR interview, June 2019.
452. USAID implementing partner official, SIGAR interview, May 2019; implementing partner chief of party, SIGAR interview, August 2019; USAID implementing partner staff, SIGAR interview, May 2019.
453. USAID implanting partner chief of party, SIGAR interview, May 2019; implementing partner chief of party, SIGAR interview, August 2019.
454. Former USAID/Afghanistan mission director, SIGAR interview, June 2019.
455. Checchi and Company Consulting, Inc., "Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities," October 26, 2017, p. i.
456. Checchi and Company Consulting, Inc., "Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities," October 26, 2017, p. ii.
457. Checchi and Company Consulting, Inc., "Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities," October 26, 2017, p. iv.
458. USAID, "ADS Chapter 201 Program Cycle Operational Policy," June 11, 2019, p. 158; Checchi and Company Consulting, Inc., "USAID/Afghanistan Plan for Transition Strategy (2015-2018) Mid-Course Stocktaking Exercise," August 2017, pp. 1, 16.
459. Checchi and Company Consulting, Inc., "USAID/Afghanistan Plan for Transition Strategy (2015-2018) Mid-Course Stocktaking Exercise," August 2017, p. 16.
460. Checchi and Company Consulting, Inc., "USAID/Afghanistan Plan for Transition Strategy (2015-2018) Mid-Course Stocktaking Exercise," August 2017, p. 16.
461. USAID, "ADS Chapter 201 Program Cycle Operational Policy," July 23, 2020, p. 38.
462. A Country Development Cooperation Strategy supports State's Integrated Country Strategy, which describes overall U.S.-government objectives for a given country. USAID, "ADS Chapter 201: Program Cycle Operational Policy," June 11, 2019, pp. 41, 143.

463. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 26.
464. The results framework, which is subject to constant adjustment during the course of any mission, is guided by a set of development hypotheses that “explain the logic and causal relationships between the building blocks needed to achieve a long-term result.” These hypotheses are intended to be “based on theory, practice, literature, and/or experience,” and help link together each level of the results framework, using the mission’s best understanding of the given problems. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 27–28.
465. This series of outcomes reflects specific components of the results framework for USAID’s current Country Development Cooperation Strategy for Afghanistan. The text to which this endnote refers summarizes a broader causal chain involving other intermediate outcomes like increasing transportation and energy service delivery and increasing the skills of the Afghan workforce. For readability, the description of USAID’s results framework provided here was also simplified. For example, USAID has specific technical terms it applies to each level of outcome specified in a results framework. Sub-intermediate results (referred to as “Sub IRs”) are the lowest-level outcomes specified in the framework. Together, a set of Sub IRs aims to produce a higher-level Intermediate Result (IR). Related IRs are intended to produce a Development Objective (DO). DOs, in turn, contribute to an ultimate strategic goal. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 27.
466. USAID’s Automated Directives System states, “Implementers, funders, and members of the evaluation team must be given the opportunity to write an individual statement of difference regarding any significant unresolved differences of opinion, to be attached to the final evaluation report.” USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 128. The Strong Hubs for Afghan Hope and Resilience program, Promote’s Women’s in Leadership Development program, and the Road Sector Sustainability Project all submitted statements of difference in response to their respective evaluation reports.
467. U.S. Senate Foreign Relations Committee, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations,” June 2011, pp. 2, 13–14; SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, pp. 4, 11; SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 6; SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. 38, 46.
468. Checchi and Company Consulting, Inc., “Final Report: Evaluation of USAID’s Afghanistan Small and Medium Enterprise Development Project (ASMED),” July 2011, pp. 4, 31, 40.
469. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, pp. ii–iii.
470. USAID officials, SIGAR interview, February 2019; evaluation expert, SIGAR interview, May 2019; evaluation specialist, SIGAR interview, April 2019.
471. Evaluation expert, SIGAR interview, May 2019.
472. USAID officials, SIGAR interview, February 2019.
473. Implementing partner official, SIGAR interview, August 2018; senior evaluation consultant, SIGAR interview, January 2019.
474. USAID, “USAID Afghanistan suggested changes/clarifications to SIGAR Lessons Learned Draft Report LL-11 Monitoring and Evaluation of Contracting,” April 27, 2021; senior evaluation consultant, SIGAR interview, January 2019.
475. Evaluation specialist, SIGAR interview, April 2019; USAID official, SIGAR interview, February 2019.
476. Evaluation specialist, SIGAR interview, April 2019.
477. Evaluation expert, SIGAR interview, August 2018.
478. International development professional, SIGAR interview, April 2019; senior evaluation consultant, SIGAR interview, January 2019.
479. Implementing partner chief of party, SIGAR interview, August 2019.
480. Similarly, the Strong Hubs for Afghan Hope and Resilience program was both delayed and adjusted, resulting in a scheduled midterm evaluation occurring so early in program implementation that the program informed evaluators it was “practically impossible to measure the extent of SHAHAR’s capacity development.” USAID implementing partner staff, SIGAR interview, May 2019; Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, p. 2.
481. Evaluation expert, SIGAR interview, May 2019.
482. Former USAID official, SIGAR interview, June 2019.
483. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, pp. 9–10.
484. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. ii.
485. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 42, 57, 222.

486. William Hammink, "USAID in Afghanistan: What Have We Learned?," *Foreign Service Journal*, July 2018, p. 6.
487. William Hammink, "USAID in Afghanistan: What Have We Learned?" *Foreign Service Journal*, July 2018.
488. See: SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018, p. 169; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 47–48.
489. USAID, "Testimony by USAID Administrator Dr. Rajiv Shah on the President's Fiscal Year 2012 Budget Requests," before the House Subcommittee on State, Foreign Operations Appropriations of the Committee on Appropriations, March 30, 2011.
490. SIGAR, *Support for Gender Equality: Lessons from the U.S. Experience in Afghanistan*, SIGAR-21-18-LL, February 2021, pp. xiii, 95; SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program's Achievements and the Afghan Government's Ability to Sustain Them*, SIGAR-18-69-AR, September 2018, p. 15; USAID implementing partner chief of party, SIGAR interview, August 2018; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 191.
491. SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program's Achievements and the Afghan Government's Ability to Sustain Them*, SIGAR-18-69-AR, September 2018, p. 15; USAID implementing partner chief of party, SIGAR interview, August 2018.
492. Former USAID senior official, SIGAR interview, December 2018.
493. USAID, "ADS Chapter 201 Program Cycle Operational Policy," June 11, 2019, p. 50.
494. USAID, "ADS Chapter 201 Program Cycle Operational Policy," June 11, 2019, pp. 17, 51–52.
495. In the year in which it is conducted, a midcourse stocktaking exercise fulfills the requirement for a mission's portfolio review. USAID, "ADS Chapter 201 Program Cycle Operational Policy," June 11, 2019, p. 52.
496. USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, pp. 51–52.
497. There are numerous examples. A 2017 evaluation of USAID's \$142.0 million Power Transmission Expansion and Connectivity Commercialization Activities project, which aimed to turn Afghanistan's national power utility into a more commercially viable business entity, made conflicting statements about the extent to which intended results were achieved, leaving the reader confused about how effective the project was. Equivocation was also an issue in a 2018 midterm evaluation of USAID's Initiative for Hygiene, Sanitation and Nutrition Activity (IHSAN). Evaluators relied on implementing partner reporting of progress against indicator targets to inform their "positive view of IHSAN achievements" while also recommending that USAID and IHSAN "review IHSAN's indicator calculation methodology" and "consider undertaking a Data Quality Assessment"—in part because of "possible issues with its indicator calculation methodology or target levels." For additional examples, see Table 3 on pp. 62–63. Checchi and Company Consulting, Inc., "Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities," October 26, 2017, pp. i–ii, iv; Checchi and Company Consulting, Inc., "Mid-Term Performance Evaluation of Initiative for Hygiene, Sanitation and Nutrition Activity," December 2018, pp. ix–x, 8.
498. USAID, "ADS Chapter 201 Program Cycle Operational Policy," June 11, 2019, pp. 51–52; USAID Office of Inspector General, "Audit of USAID/Afghanistan's Strategy for Monitoring and Evaluating Programs Throughout Afghanistan," Audit Report No. F-306-16-001-P, December 10, 2015, pp. 2, 7.
499. U.S. Senate Foreign Relations Committee, "Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations," June 2011, pp. 2, 13–14; SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, pp. 4, 11; SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 6; SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. 38, 46.
500. USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, p. 2.
501. USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, p. 2.
502. USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, p. 2.
503. USAID, Office of Afghanistan and Pakistan Affairs, "Grant Agreement 306-AA-18," September 6, 2018, pp. 3–4; USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, pp. 2, 9.
504. Herbie Smith, "Staying in Afghanistan is not worth more U.S. lives," *Delaware Online*, April 25, 2019.
505. USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, p. 4.
506. Herbie Smith, "Staying in Afghanistan is not worth more U.S. lives (Opinion)," *Delaware Online*, April 25, 2019.
507. USAID, "Country Development Cooperation Strategy FY 2019–2023," September 2018, p. 2; Herbie Smith, "Staying in Afghanistan is not worth more U.S. lives," *Delaware Online*, April 25, 2019.
508. Herbie Smith, "Staying in Afghanistan is not worth more U.S. lives," *Delaware Online*, April 25, 2019.
509. Former USAID official, SIGAR interview, June 2019.
510. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 56–57, 75; Checchi and Company Consulting, Inc., "Final Report: Local Governance and Community Development Program (LGCD) Evaluation," January 2009, p. 22, footnote 2; former senior USAID official, SIGAR interview, December 2018.
511. Evaluation expert, SIGAR interview, August 2018.

512. Evaluation contractor, SIGAR interview, December 2018.
513. USAID implementing partner chief of party, SIGAR interview, May 2019.
514. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 56–57, 75; Checchi and Company Consulting, Inc., “Final Report: Local Governance and Community Development Program (LGCD) Evaluation,” January 2009, p. 22, footnote 2.
515. Former senior USAID official, SIGAR interview, December 2018.
516. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 56.
517. Senior evaluation consultant, SIGAR interview, January 2019; Checchi and Company Consulting, Inc., “Afghanistan Governance Sector Assessment,” November 2018, p. 13.
518. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
519. Former USAID official, SIGAR interview, January 2019.
520. Evaluation expert, SIGAR interview, May 2019.
521. USAID, “Strong Hubs for Afghan Hope and Resilience (SHAHAR),” December 2018, p. 1; Checchi and Company Consulting, “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, pp. 1–2.
522. USAID, “Strong Hubs for Afghan Hope and Resilience (SHAHAR),” December 2018, p. 1; Checchi and Company Consulting, “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, p. 25.
523. Checchi and Company Consulting, “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, p. 25.
524. Checchi and Company Consulting, “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, pp. 1–2, 25.
525. Checchi and Company Consulting, “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, pp. 1–2.
526. Checchi and Company Consulting, “Mid-Term Performance Evaluation of the Strong Hubs for Afghan Hopes and Resilience Program,” June 2016, p. 2.
527. Activities in particular are meant to achieve “clear and measurable results.” USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 82, 103–104.
528. Kathryn Bryant, letter to La Chenna Cromer, Subject: Cooperative Agreement No. AID-306-A-13-00009-00, USAID Afghanistan University Support and Workforce Development Program with Family Health International 360 (FHI 360), December 28, 2013, pp. i, 30, 43.
529. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation of the University Support and Workforce Development Program,” January 2017, p. 15.
530. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation of the University Support and Workforce Development Program,” January 2017, p. 15; USAID implementing partner staff, SIGAR interview, May 2019.
531. USAID implementing partner staff, SIGAR interview, May 2019.
532. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation of the University Support and Workforce Development Program,” January 2017, p. 15.
533. Checchi and Company Consulting, “Mid-Term Performance Evaluation of Engineering Support Program,” August 2018, p. ii.
534. Checchi and Company Consulting, “Mid-Term Performance Evaluation of Engineering Support Program,” August 2018, pp. ii, x.
535. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR 9-57-AR, September 2019, pp. 1, 4.
536. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 11.
537. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 2.
538. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, pp. 4, 22–23.
539. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, pp. 4, 22–23.
540. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 7.
541. SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, pp. 24–26.

542. SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 18.
543. SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, pp. ii, 20.
544. SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, pp. 12–13; USAID Office of Inspector General investigator, SIGAR interview, August 2018.
545. SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 14.
546. SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 17.
547. SIGAR, *USAID's Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government's Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, pp. 16–17.
548. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 110.
549. Implementing partner chief of party, SIGAR interview, April 2019.
550. Implementing partner staff, SIGAR interview, May 2019.
551. Former USAID official, SIGAR interview, January 2019.
552. USAID contractor, SIGAR interview, September 2018.
553. Senior evaluation consultant, SIGAR interview, January 2019.
554. Senior evaluation consultant, SIGAR interview, January 2019.
555. USAID implementing partner chief of party, SIGAR interview, May 2019.
556. USAID implementing partner chief of party, SIGAR interview, May 2019.
557. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, pp. 14–15.
558. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, p. 7.
559. USAID implementing partner staff, SIGAR interview, May 2019.
560. Special Immigrant Visas for lawful permanent residence may be granted to some Afghan nationals who have been employed by or otherwise been affiliated with U.S. government agencies and who may as a consequence be at risk. A similarly named but separate visa program is open to certain Afghan (and Iraqi) translators and interpreters. USAID staff in the Office of Acquisition and Assistance, SIGAR interview, April 2019.
561. USAID staff in the Bureau for Policy, Planning, and Learning, SIGAR interview, February 2019.
562. Monitoring Support Program contractor staff, SIGAR interview, July 2019.
563. USAID implementing partner, SIGAR interview, May 2019; USAID implementing partner chief of party, SIGAR interview, August 2018.
564. USAID implementing partner, SIGAR interview, May 2019.
565. International development professional, SIGAR interview, April 2019.
566. Former USAID senior official, SIGAR interview, December 2018.
567. Former USAID official, SIGAR interview, January 2019.
568. Referring to the “first few years” of his experience, one INL official who began working on Afghanistan in 2006 stated, “There was no rigorous M&E structure. Instead, there was more political-type, ad hoc reporting, like what was going on in an area.” Similarly, contractors for State’s long-running, Corrections System Support Program (on which State spent at least \$259.5 million, stated that, prior to a 2013 State inspector general audit, State was not “pushing for measurable outcomes” to assess the program’s effectiveness. Instead, prior to the audit, “The focus was on pictures and anecdotes. Everyone wanted a good photo, but nothing more.” State, INL staff, SIGAR interview, March 2019; PAE employees, SIGAR interview, May 2019; SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 13.
569. In the words of a 2008 RAND report, “There was little interest at the highest levels of the U.S. government to get into postwar planning” in the early stages of reconstruction. In the absence of such planning, there was little that M&E could actually assess. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, p. 91.
570. Afghanistan Freedom Support Act of 2002, Pub. L. No. 107-327, § 104, 205 (2002).
571. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, p. 99.
572. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, p. 99.

573. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, p. 99.
574. GAO, "Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed," GAO-04-403, June 2004, p. 33.
575. GAO, "Afghanistan Reconstruction: Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed," GAO-04-403, June 2004, pp. 54–55.
576. GAO, "Securing, Stabilizing, and Reconstructing Afghanistan: Key Issues for Congressional Oversight," GAO-07-801SP, May 2007, p. 2.
577. In a speech delivered at the U.S. Military Academy in December 2009, then-President Barack Obama articulated the new administration's Afghanistan strategy. Obama stated, "This effort must be based on performance. The days of providing a blank check are over." In the words of one longtime INL official, by around 2009 or 2010 "The Hill had less appetite for anecdotes, and wanted to see tangible results." White House, "The New Way Forward – The President's Address," December 1, 2009; State, INL staff, SIGAR interview, March 2019.
578. White House, "Fact Sheet: U.S. Global Development Policy," accessed July 30, 2020.
579. Similarly, GAO reported, "State officials stated that eradication program performance was continually reviewed and assessed but were unable to provide any documented evaluations of eradication programs." GAO, "Afghanistan Drug Control: Strategy Evolving and Progress Reported, but Interim Performance Targets and Evaluation of Justice Reform Efforts Needed," GAO-10-291, March 2010, pp. 36–37.
580. State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Corrections System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, p. 8; James A. Walsh, "Subject: INL Responses to the Draft Report, 'Audit of Bureau of International Narcotics and Law Enforcement Affairs Corrections System Support Program (CSSP) in Afghanistan' (AUD-MERO-13-XX, August 2013)," memorandum for OIG/MERO/AUD Deputy Assistant Inspector General Carol N. Gorman, August 20, 2013, p. 3, in State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Correction System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, p. 23.
581. State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Corrections System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, pp. 8, 11; James A. Walsh, "Subject: INL Responses to the Draft Report, 'Audit of Bureau of International Narcotics and Law Enforcement Affairs Corrections System Support Program (CSSP) in Afghanistan' (AUD-MERO-13-XX, August 2013)," memorandum for OIG/MERO/AUD Deputy Assistant Inspector General Carol N. Gorman, August 20, 2013, p. 3, in State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Correction System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, p. 23.
582. CNA, "Understanding Monitoring and Evaluation: Lessons from the Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 18.
583. CNA, "Understanding Monitoring and Evaluation: Lessons from the Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 32.
584. State Office of Inspector General "Review of Department of State Compliance with Program Evaluation Requirements," Report No. ISP-I-15-36, September 2015, p. 11.
585. State, Office of Inspector General, "Review of Department of State Compliance with Program Evaluation Requirements," Report No. ISP-I-15-36, September 2015, p. 4.
586. State Office of Inspector General, "Review of Department of State Compliance with Program Evaluation Requirements," Report No. ISP-I-15-36, September 2015, p. i.
587. State, "Report to Congress on Evaluation Quality, Cost and Other Matters," accessed January 2020.
588. CNA, "Understanding Monitoring and Evaluation: Lessons from the Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 18; State, 18 FAM 300 Strategic Direction and Management; 18 FAM 301 Principles and Practices; 18 FAM 301.1 Managing for Results (MfR) Framework, February 2019, p. 3.
589. CNA, "Understanding Monitoring and Evaluation: Lessons from the Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 18.
590. State, "Department of State Program and Project Design, Monitoring and Evaluation Policy," November 2017, p. 4; State, response to SIGAR data call, May 3, 2019.
591. See: SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 13–15; State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Corrections System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, pp. 1, 8–9; State Office of Inspector General, "Audit of Bureau of International Narcotics and Law Enforcement Affairs Counternarcotics Assistance to Afghanistan," AUD-MERO-15-02, November 2014, pp. 5–6.
592. James Dobbins, Michele A. Poole, Austin Long, and Benjamin Runkle, *After the War: Nation-Building from FDR to George W. Bush*, RAND Corporation, 2008, pp. 90–91; President George W. Bush, "Selected Speeches of President George W. Bush," White House, Office of the Press Secretary, pp. 85–86.
593. President George W. Bush, "Selected Speeches of President George W. Bush," White House, Office of the Press Secretary, pp. 86–87, 103.

594. Projects intended to benefit Afghan women were often implemented as components of other State programs. As such, such projects will not be separately examined in this chapter. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. ii; SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. iii; State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, pp. 9–10; U.S. Embassy Kabul, “U.S. State Department Announces \$10.7 Million for Humanitarian Demining Programs in Afghanistan,” August 2015, p. 1; State Office of Inspector General, “Management Assistance Report: Although Progress Has Been Made, Challenges Remain in Monitoring and Overseeing Antiterrorism Assistance Program Activities in Afghanistan,” Report No. AUD-MERO-18-16, November 2017, p. 6; SIGAR, *Afghan Women: Comprehensive Assessments Needed to Determine and Measure DOD, State, and USAID Progress*, SIGAR-15-24-AR, December 2014, p. 1.
595. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. i.
596. SIGAR, *Support for Afghanistan’s Justice Sector: State Department Programs Need Better Management and Stronger Oversight*, SIGAR-14-26-AR, January 2014, p. ii; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2020, p. 112.
597. Barbaricum LLC, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 1; State Office of Inspector General, “Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan,” Report No. AUD-MERO-13-37, September 2013, p. 1.
598. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 4–5.
599. In 2010, State drafted the “U.S. Government Anticorruption Strategy for Afghanistan,” emphasizing programs aimed at increasing Afghan political will and capacity to fight corruption, and calling for the reform of American program contracting practices. SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016, p. 39.
600. In response to a draft of this report, INL stated that including measures of progress in the 2013 Rule of Law and Law Enforcement Framework “could have led to uncoordinated and unaligned goals.” SIGAR’s view is that identifying specific measures of success is vital even at the level of a framework or a strategy. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 5, 22; State, INL, correspondence with SIGAR, May 10, 2021.
601. State, “Afghanistan: Integrated Country Strategy,” September 2018, p. 4.
602. For example, see: SIGAR, *Afghanistan’s Anti-Corruption Efforts: The Afghan Government Made Progress in Meeting its Anti-Corruption Strategy Benchmarks, but Serious Challenges Remain to Fighting Corruption*, SIGAR-20-06-AR, November 2019, p. i; Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” on behalf of Federal Ministry for Economic Cooperation and Development (BMZ), Germany Division for Afghanistan and Pakistan, March 2020, pp. 8, 16; SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015, p. i; SIGAR, *Quarterly Report to the United States Congress*, October 30, 2018, p. 123.
603. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015, p. i.
604. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015, p. i.
605. U.S. Department of Justice, Department of Justice/Office of the Deputy Attorney General – INL Inter-Agency Agreement (December 30, 2016) Amendment One Counter-Corruption, Major Crimes, Counternarcotics, and National Security Investigation and Prosecution Program Quarterly Progress Report, June 2018, p. 14.
606. Freedom House, “About Us,” n.d., accessed May 17, 2021; Freedom House, “Afghanistan: Freedom in the World 2020,” n.d., accessed May 17, 2021.
607. SIGAR, *Afghanistan’s Anti-Corruption Efforts: The Afghan Government Made Progress in Meeting its Anti-Corruption Strategy Benchmarks, but Serious Challenges Remain to Fighting Corruption*, SIGAR-20-06-AR, November 2019, p. ii.
608. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. iii.
609. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 21.
610. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. ix.
611. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 17.
612. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. x–xi.
613. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. 54, 137.

614. Despite the lack of emphasis in the Afghanistan ICS, narcotics control does feature as a theme in State-generated strategies for various central Asian countries. State, “Afghanistan: Integrated Country Strategy,” September 2018, p. 6.
615. SIGAR, *Quarterly Report to the United States Congress*, January 30, 2020, p. 21.
616. SIGAR, *Review Letter: Waste, Fraud, and Abuse Uncovered by SIGAR*, SIGAR-18-52-SP, July 2018, p. 3.
617. As Inspector General Sopko explained during a congressional hearing in January 2020, “You’ll notice three variables that we, as IGs look at: inputs, outputs, and outcomes. We look at the outcome that [an] administration told Congress they were supposed to [achieve]. So, [for] counter-narcotics, it was to lessen the amount of opium. . . . You know, end that scourge. Well, it’s been a total waste. None of our programs have led to any reduction in opium in Afghanistan. As a matter of fact, opium is the largest export of Afghanistan.” SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, pp. i, 4; John F. Sopko, “Examining the Trump Administration’s Afghanistan Strategy,” testimony before the Committee on Homeland Security and Government Affairs Subcommittee on National Security, 116th Congress, January 28, 2020, electronic p. 32.
618. UN Office on Drugs and Crime, “Afghanistan Opium Survey 2020 Cultivation and Production: Executive Summary,” April 2020, p. 8.
619. UN Office on Drugs and Crime, “Afghanistan Opium Survey 2020 Cultivation and Production: Executive Summary,” April 2020, p. 8.
620. State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, pp. 3, 9.
621. SIGAR, *Afghan Refugees and Returnees: Corruption and Lack of Afghan Ministerial Capacity Have Prevented Implementation of a Long-term Refugee Strategy*, SIGAR-15-83-AR, August 2015, p. 2; State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, p. 10; State, PRM, correspondence with SIGAR, May 17, 2021.
622. State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, p. 5; State, PRM, correspondence with SIGAR, May 17, 2021.
623. State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, p. 6; State, PRM, correspondence with SIGAR, May 17, 2021.
624. State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, p. 6; State, PRM, correspondence with SIGAR, May 17, 2021.
625. State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, pp. 5–6, 10; State, PRM, correspondence with SIGAR, May 17, 2021.
626. SIGAR, *Afghan Refugees and Returnees: Corruption and Lack of Afghan Ministerial Capacity Have Prevented Implementation of a Long-term Refugee Strategy*, SIGAR-15-83-AR, August 2015, p. ii; State, PRM, correspondence with SIGAR, May 17, 2021.
627. SIGAR, *Afghan Refugees and Returnees: Corruption and Lack of Afghan Ministerial Capacity Have Prevented Implementation of a Long-term Refugee Strategy*, SIGAR-15-83-AR, August 2015, p. ii; State, PRM, correspondence with SIGAR, May 17, 2021.
628. State, PRM, correspondence with SIGAR, May 17, 2020.
629. State’s Office of Inspector General concurred with this assessment, stating that the Afghan government did not have the capacity to implement refugee assistance programs due to systemic corruption, demand for services far exceeding supply of resources, and weak administrative and judicial institutions. SIGAR, *Afghan Refugees and Returnees: Corruption and Lack of Afghan Ministerial Capacity Have Prevented Implementation of a Long-term Refugee Strategy*, SIGAR-15-83-AR, August 2015, p. i; State Office of Inspector General, “The Bureau of Population, Refugees and Migration’s Reintegration Assistance Program for Refugees Returning to Afghanistan,” Report No. MERO-I-11-10, July 2011, p. 5; State, PRM, correspondence with SIGAR, May 17, 2021.
630. State, PRM, response to SIGAR quarterly report data call, September 18, 2020; State, PRM, response to SIGAR vetting, October 6, 2020.
631. International Organization for Migration, “Return of Undocumented Afghans Weekly Situation Report,” January 6, 2021, p. 1.
632. State Office of Inspector General, “Humanitarian Mine Action Programs in Afghanistan,” Report No. ISP-I-10-11, November 2009, pp. 1, 4–5; State Office of Inspector General, “Audit of the Conventional Weapons Destruction Program in Afghanistan, Iraq, and Lebanon,” Report No. AUD-MERO-17-49, August 2017, p. 1.
633. SIGAR, *Quarterly Report to the United States Congress*, January 30, 2020, pp. 89–90.
634. UN, “Rising Landmine blast toll in Afghanistan highlights long-term care needs of survivors,” *UN News*, February 6, 2019; Government of Afghanistan, Directorate of Mine Action Coordination, “MAPA Annual Report for 1398,” September 23, 2020, pp. 5–6.

635. State, "U.S. Foreign Assistance Reform: Achieving Results & Sustainability in Support of Transformational Diplomacy," June 2006; State, "Standard Foreign Assistance Indicators."
636. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 61, electronic p. 108.
637. State, response to SIGAR data call, May 28, 2019.
638. State, response to SIGAR data call, May 28, 2019.
639. State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," pp. 8–9.
640. State, response to SIGAR data call, May 28, 2019; SIGAR analysis.
641. State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," p. 9.
642. State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," p. 9.
643. State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," p. 9.
644. State, INL, "TAF LALE PMP Revised Final from Implementer," February 9, 2017.
645. State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," p. 9.
646. SIGAR analysis of State, INL, "Legal Aid through Legal Education Performance Measurement Plan (PMP) FY2018," pp. 6, 8–9.
647. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 14–15.
648. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 15.
649. State, "Securing U.S. National Security Objectives in Afghanistan: 2018 Successes," December 20, 2018.
650. Kabul Bank Receivership, "Kabul Bank Recovery and Loan Portfolio," September 16, 2017; Kabul Bank Receivership, "Kabul Bank Recovery and Loan Portfolio," September 23, 2018.
651. Calculation includes interest from the stolen funds owed to the Afghan government by the debtors. The government later reversed more than \$52 million of interest that had previously been waived. Applying the same recovery rate from September 2017 to September 2018 to the new amount, it would have taken 139.25 years to recover the stolen funds. SIGAR analysis of Government of Afghanistan, KBR, "Kabul Bank Recovery and Loan Portfolio Dates 16 Sep 2017," September 16, 2017; Government of Afghanistan, KBR, "Kabul Bank Recovery and Loan Portfolio Dates 23 Sep 2018," September 23, 2018; Government of Afghanistan, KBR, "Kabul Bank Recovery and Loan Portfolio Dates 20 Aug 2020," August 20, 2020.
652. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2018.
653. State Office of Inspector General, "Audit of Bureau of International Narcotics and Law Enforcement Affairs Counternarcotics Assistance to Afghanistan," AUD-MERO-15-02, November 2014, p. 5.
654. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. 88–89.
655. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. i.
656. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 89.
657. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. 86, 88–89.
658. State Office of Inspector General, "Audit of Bureau of International Narcotics and Law Enforcement Affairs Counternarcotics Assistance to Afghanistan," AUD-MERO-15-02, November 2014, pp. 5–6.
659. State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Counternarcotic Assistance to Afghanistan," Report No. MERO-A-15-02, November 2014, p. 6.
660. State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Counternarcotic Assistance to Afghanistan," Report No. MERO-A-15-02, November 2014, pp. 5–6.
661. State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY 2017," October 2018, p. 4.
662. State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY 2017," October 2018, p. 5.
663. State, INL, correspondence with SIGAR, May 10, 2021; State Department officials, SIGAR interview, April 2019.
664. INL's official position is that "we should spend the money to make sure someone is changing the lightbulbs." State Department officials, SIGAR interview, April 2019; State, INL, correspondence with SIGAR, May 10, 2021.
665. State Department officials, SIGAR interview, March 2019.
666. State officials in the Bureau of South and Central Asian Affairs, SIGAR interview, February 2019; ME&A, "Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation," April 2018, pp. 34–35.
667. ME&A, "Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation," April 2018, pp. 34–35; State officials in the Bureau of South and Central Asian Affairs, SIGAR interview, February 2019.
668. ME&A, "Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation," April 2018, pp. 22–23.
669. State, INL officials, SIGAR interview, March 2019.
670. State, INL officials, SIGAR interview, March 2019.

671. State, INL officials, SIGAR interview, March 2019.
672. State, INL, correspondence with SIGAR, May 2021.
673. State officials in the Bureau of South and Central Asian Affairs, SIGAR interview, February 2019.
674. Defense researcher, SIGAR interview, December 2018.
675. John F. Sopko, "Oversight in Conflict: Detecting Fraud in Challenging Environments," speech before the Association of Certified Fraud Examiners, Law Enforcement & Government Anti-Fraud Summit, November 7, 2019, p. 10.
676. State, INL, correspondence with SIGAR, May 10, 2021.
677. Julia Steets, Elias Sagmeister and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," Secure Access in Volatile Environments (SAVE), final report, October 2016, p. 38.
678. Julia Steets, Elias Sagmeister and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," Secure Access in Volatile Environments (SAVE), final report, October 2016, p. 38.
679. Julia Steets, Elias Sagmeister and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," Secure Access in Volatile Environments (SAVE), final report, October 2016, p. 38.
680. In response to a draft of this callout box, INL reported that, as of May 2021, it was in the midst of an ongoing, bureau-level reorganization effort that aims to improve M&E. The effort includes external subject-matter-expert contracts and the establishment of a knowledge management office.
681. State, INL, correspondence with SIGAR, May 10, 2021.
682. State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY2017," October 2018, pp. 4, 9–10.
683. State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY2017," October 2018, p. 5.
684. Human Rights Watch describes Japan as "a prosperous liberal democracy with the third largest economy and a vibrant civil society." However, in its 2020 *World Report*, it also noted "Japan's long overlooked 'hostage' justice system, in which criminal suspects are held for long periods in harsh conditions to coerce a confession." According to Human Rights Watch, that issue "received renewed attention after the high-profile arrest of former Renault and Nissan head Carlos Ghosn in November 2018 for alleged financial misconduct." The detention of Ghosn invited scrutiny from the UN. A source for a report produced by a UN working group that investigated the incident alleged "that Mr. Ghosn was exposed to a systemic pattern in which suspects are detained for long periods of time in difficult conditions that force them to confess," according to the working group's report. The group concluded that Japanese authorities had arbitrarily deprived Ghosn of his liberty and that the deprivation contravened several articles of the UN's Universal Declaration of Human Rights. According to a 2015 *Economist* article, "reformers who have surveyed some of the world's penal hellholes say that Japan's jails rank among the cruelest—for the psychological toll they take on inmates." Human Rights Watch, "World Report 2020: Rights Trends in Japan," n.d., accessed January 7, 2021; Human Rights Watch, "Call to Eliminate Japan's 'Hostage Justice' System by Japanese Legal Professionals," April 10, 2019, accessed January 7, 2021; UN, Human Rights Council, Working Group on Arbitrary Detention, "Opinions adopted by the Working Group on Arbitrary Detention at its eighty-eighth session, 24–28 August 2020, Opinion No. 59/2020 concerning Carlos Ghosn (Japan)," November 20, 2020, pp. 6, 15; *Economist*, "Silent Screams," December 3, 2015; Freedom House, "Japan," 2020, accessed April 2020.
685. SIGAR, *Afghanistan's Justice Sector Case Management System: Seized or Forfeited Assets Were Not Tracked and Nationwide Implementation is Not Complete*, SIGAR-20-20-SP, January 2020, electronic p. 2.
686. State Department officials, SIGAR interview, April 30, 2019.
687. State, response to SIGAR data call, May 22, 2019, May 30, 2019.
688. See: SIGAR, *Afghanistan's Anti-Corruption Efforts: The Afghan Government Made Progress in Meeting its Anti-Corruption Strategy Benchmarks, but Serious Challenges Remain to Fighting Corruption*, SIGAR-20-06-AR, November 2019, pp. ii, 24; SIGAR, *Afghanistan's Justice Sector Case Management System: Seized or Forfeited Assets Were Not Tracked and Nationwide Implementation is Not Complete*, SIGAR-20-20-SP, January 2020, electronic p. 2.
689. State, INL, correspondence with SIGAR, May 10, 2021.
690. Freedom House, "China," 2020, accessed April 2020.
691. According to INL, it is a "fair point that no matter how much capacity you provide to anti-corruption entities, you will only get so far without the political will to go after powerful figures." However, INL claimed that this point was "not a great fit" with respect to the Justice Sector Support Program because the program had worked to build support within the Afghan government to fully adopt the case management system. It may be the case that JSSP has developed the necessary political will within the Afghan government to implement the case management system. However, the political will to implement the system is distinct from the political will required to address the corruption and impunity of powerful individuals. If the latter does not materialize, the utility of the case management system will be limited. SIGAR, *Afghanistan's Anti-Corruption Efforts: The Afghan Government Made Progress in Meeting its Anti-Corruption Strategy Benchmarks, but Serious Challenges Remain to Fighting Corruption*, SIGAR-20-06-AR, November 2019, pp. i, iii, 24; State, INL, correspondence with SIGAR, May 10, 2021.
692. State, INL, correspondence with SIGAR, May 10, 2021.
693. SIGAR, *2021 High-Risk List*, February 19, 2021, p. 55.

694. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, electronic p. 2.
695. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, electronic, p. 2.
696. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, electronic p. 2, p. 8.
697. USAID, "AUAF Performance Management Plan," May 23, 2014.
698. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, electronic p. 1, pp. 14–15; USAID, "AUAF Performance Management Plan," May 23, 2014.
699. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 3.
700. AUAF records obtained by SIGAR and USAID Office of Inspector General investigators indicated that the university was not sustainable: It had lost more than \$63 million since 2012, it depended on U.S. aid for 86 percent of its funding, and as of February 2018 it had money for only another month's expenses. After SIGAR reported these findings in its April 30, 2019, quarterly report, AUAF board chairman David Sedney said that investigations of the university insufficiently focused on AUAF's achievements. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019, p. 40; Rod Nordland, "U.S.-Funded Afghan College Is Under Scrutiny for Missing Millions, Officials Say," *New York Times*, May 30, 2019.
701. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 8.
702. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, p. 6; SIGAR analysis of USAID, "AUAF Performance Management Plan," May 23, 2014.
703. State officials in the Bureau of South and Central Asian Affairs, SIGAR interview, February 2019.
704. In response to a draft of this report, INL stated, "This is why INL uses a mix of quantitative and qualitative reporting." State, INL, correspondence with SIGAR, May 10, 2021.
705. State Department officials, SIGAR interview, April 2019.
706. SIGAR, *The American University of Afghanistan: State and USAID Have Taken Action to Address Concerns with the Management, Processes, and Systems at the University*, SIGAR-20-33-AR, April 2019, electronic p. 2.
707. SIGAR, "Seven years after initial request, the Departments of Defense and State and the U.S. Agency for International Development have yet to list their 10 most and least successful reconstruction projects asked for by Inspector General John F. Sopko," press release, January 22, 2020, accessed December 8, 2020.
708. In reference to the request, Inspector General Sopko stated in April 2019 testimony before the Congress, "And you remember, Mr. Chairman, I got into an actual, I wouldn't say fight, a serious discussion with a senior official from USAID about racking and stacking programs in USAID. And he turned on me saying, the audacity of Mr. Sopko to want me to list which programs are working and which ones are failing. It's almost like picking your favorite child. I just had to stop for a second. I mean I never thought development aid was sort of the Sophie's choice that we had to pick. I mean, doesn't it make sense as a businessman?" SIGAR, "Seven years after initial request, the Departments of Defense and State and the U.S. Agency for International Development have yet to list their 10 most and least successful reconstruction projects asked for by Inspector General John F. Sopko," press release, January 22, 2020, accessed December 8, 2020; John F. Sopko, "2019 High-Risk List," testimony before House Oversight and Reform Subcommittee on National Security Hearing, 116th Congress, April 3, 2019, electronic p. 21.
709. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR's Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. 4.
710. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 1.
711. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 1.
712. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. i.
713. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. i.
714. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 6, 19.
715. State, "U.S. Government Rule of Law Strategy for Afghanistan," September 2009, pp. 1–2.
716. State, "U.S. Government Rule of Law Strategy for Afghanistan," September 2009, p. 11.

717. In response to a draft of this report, INL stated, “An overall State Department strategy should not necessarily include M&E components or performance measures, as those are developed at the Bureau-level for offices in charge of implementation.” One recommendation of SIGAR’s 2015 audit of U.S. rule of law assistance was that a then-draft U.S. rule of law strategy for Afghanistan include “a plan for measuring the performance of the strategy and specific performance metrics to evaluate U.S. progress in meeting the strategy’s objectives.” SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 5, 22; State, INL, correspondence with SIGAR, May 10, 2021.
718. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. i.
719. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 14.
720. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 15.
721. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 1.
722. State Office of Inspector General, “Inspection of the Bureau of South and Central Asian Affairs,” Report No. ISP-I-18-11, February 2018, p. 5.
723. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, pp. 4–5.
724. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, pp. 5, 7.
725. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 7.
726. State, Office of the Director of U.S. Foreign Assistance, “A Meta-Evaluation of Foreign Assistance Evaluations,” June 1, 2011, p. 3; ME&A, “Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation,” April 2018, pp. v–vi.
727. ME&A, “Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation,” April 2018, p. vii.
728. State, response to SIGAR data call, April 22, 2019, April 29, 2019.
729. PAE employees, SIGAR interview, May 2019.
730. ME&A, “Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation,” April 2018, pp. 30, 36.
731. State, response to SIGAR data call, May 21, 2019.
732. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 1.
733. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 6.
734. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, p. 13; Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, electronic p. 1.
735. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 2015, electronic p. 1, p. 14; Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, electronic p. 1.
736. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 1.
737. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 3–4.
738. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 18.
739. SIGAR analysis of Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 3–4.
740. INL claimed that this paragraph “is negated by the paragraph that follows. Since there weren’t baseline measures, it’s unclear whether 60 percent was an increase, decrease, or stable response. It does not imply that training did or did not work.” INL also stated, “Further independent reporting from UNAMA found no instances of reported torture in Afghan civilian prisons, while torture rates remained high in [the Afghan National Directorate of Security and Afghan National Police] facilities, indicating that some of these actions did have an impact on human rights in Afghanistan.” This report is not making a determination as to whether the training did or did not work. Rather, we point out that the evaluation failed to answer the question of whether the Correction System Support Program’s training was effective—a significant weakness. Regardless of the absence of baseline data or a control group, there was nothing stopping the evaluators from making an informed judgement about effectiveness based on the totality of the evidence available to

- them. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 4; State, INL, correspondence with SIGAR, May 10, 2021.
741. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 1.
 742. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 1.
 743. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 2.
 744. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 2.
 745. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 2.
 746. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 2.
 747. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 3.
 748. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, p. 4.
 749. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 36–37.
 750. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 36–37.
 751. Barbaricum, “Final Evaluation Report: INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 6–7.
 752. The number of evaluations conducted by State is very low compared to that of USAID. To ensure our analysis was not anecdotal, we examined all evaluations of State reconstruction programming available to us. Those included: midterm and final evaluations of the Justice Training Transition Program (2014 and 2016), a final evaluation of the Corrections System Support Program (2015), and a midterm evaluation of the Community-Based Agriculture and Rural Development–West project (2019).
 753. ME&A, “Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation,” April 2018, p. vii.
 754. ME&A, “Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation,” April 2018, p. vii.
 755. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, pp. 8, 63.
 756. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 22.
 757. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 23.
 758. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 24.
 759. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 24.
 760. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 24.
 761. SIGAR analysis of Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 24.
 762. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, pp. 46–47.
 763. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 47.
 764. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 47.
 765. The evaluators noted that the program had “failed to even attempt to measure the results of [its] efforts [on the justice sector as a whole] in any comprehensive manner.” In the absence of such measurement, it was “very difficult to identify the direct and indirect, medium and long-term results of such projects, nor to analyze how they can positively contribute to the goal of strengthening the Afghan justice system and restoring public confidence in it.” The evaluators were careful not to characterize their work as an impact evaluation. Nevertheless, they explored the potential impact of donor work in the Afghan justice sector at several points in the evaluation. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, pp. 8, 17, 21, 44, 46–47.
 766. The evaluation reached the favorable conclusion that CBARD–West “is a well-designed, well-run livelihoods project that is very well received by beneficiaries and is already making a positive impact on their lives.” But the primary purpose of the project is to “prove the viability of high-value agricultural-based interventions . . . as a sustainable alternative to illicit crop cultivation” and it was “too early to discover whether

- the project has resulted in any sustainable reduction in opium production.” The evaluation conceded that Afghanistan was a “very difficult” environment and mentioned at a high level some of the contextual challenges described in this highlight. Still, it is difficult to understand why the evaluation did not directly address the high likelihood of failure given the deeply entrenched obstacles to successful counternarcotics programming in Afghanistan, particularly when, according to the evaluation, “preliminary indications [from the UN Office on Drugs and Crime’s April 2018 Opium Survey] are that the poppy area has fallen in both treatment and control communities.” For more on those obstacles, see: GAO, “Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004; GAO, “Afghanistan Drug Control: Despite Improved Efforts, Deteriorating Security Threatens Success of U.S. Goals,” GAO-07-78, November 2006; GAO, “Afghanistan Drug Control: Strategy Evolving and Progress Reported, but Interim Performance Targets and Evaluation of Justice Reform Efforts Needed,” GAO-10-291, March 2010; SIGAR, *Counternarcotics: Lessons Learned from the U. S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018; SIGAR, *Update on the Amount of Waste, Fraud and Abuse Uncovered through SIGAR’s Oversight Work*, SIGAR-18-60, July 2018. Steve Goss, “Mid-Term Evaluation of the Community-Based Agriculture and Rural Development–West Project,” prepared for UNDP Afghanistan and INL, February 2019, pp. 5, 9, 14–15, 62; SIGAR analysis.
767. SIGAR, *Quarterly Report to the United States Congress*, January 30, 2020, p. 138; Steve Goss, “Mid-Term Evaluation of the Community-Based Agriculture and Rural Development- West Project,” prepared for UNDP Afghanistan and INL, February 2019, p. 16.
 768. Steve Goss, “Mid-Term Evaluation of the Community-Based Agriculture and Rural Development- West Project,” prepared for UNDP Afghanistan and INL, February 2019, pp. 21, 88–104.
 769. Steve Goss, “Mid-Term Evaluation of the Community-Based Agriculture and Rural Development- West Project,” prepared for UNDP Afghanistan and INL, February 2019, pp. 9, 74.
 770. SIGAR’s statement referred to the entire CBARD program, which has multiple regional projects. The evaluation described here was of CBARD-West, the component of the program implemented in Farah and Badghis Provinces. SIGAR, *Counternarcotics: Lessons Learned from the U. S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 126.
 771. More broadly, SIGAR found that no counterdrug program undertaken by the United States, its coalition partners, or the Afghan government resulted in lasting reductions in poppy cultivation. Shortly after the lessons learned report was released, SIGAR informed the Congress that almost \$7.3 billion—or, nearly the entirety of U.S. counternarcotics spending in Afghanistan—appeared to have done very little to stem the production and exportation of illicit drugs. This figure included spending on all programs with a “substantial counternarcotics focus” from 2002 to 2017. In 2020, SIGAR cautioned that an additional \$1.6 billion in counternarcotics spending had been wasted. The additional \$1.6 billion counted as additional waste were allocated to programs with a counternarcotics component, as opposed to a “substantial counternarcotics focus.” SIGAR, *Counternarcotics: Lessons Learned from the U. S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, pp. viii, 114; SIGAR, *Update on the Amount of Waste, Fraud and Abuse Uncovered through SIGAR’s Oversight Work*, SIGAR-18-60, July 2018, electronic pp. 1–2, p. 2.
 772. GAO, “Deteriorating Security and Limited Resources Have Impeded Progress; Improvements in U.S. Strategy Needed,” GAO-04-403, June 2004, p. 39.
 773. GAO, “Afghanistan Drug Control: Despite Improved Efforts, Deteriorating Security Threatens Success of U.S. Goals,” GAO-07-78, November 2006, p. 34.
 774. GAO, “Afghanistan Drug Control: Strategy Evolving and Progress Reported, but Interim Performance Targets and Evaluation of Justice Reform Efforts Needed,” GAO-10-291, March 2010, p. 12.
 775. As SIGAR reported in 2018, “The exponential rise in opium poppy cultivation and drug production is rooted in far-reaching, persistent challenges in Afghanistan—namely, lack of security, a poor economy, weak governing institutions, and failures of the wider reconstruction effort.” For more information on these topics, see: SIGAR, *Corruption in Conflict: Lessons from the U.S. Experience in Afghanistan*, SIGAR-16-58-LL, September 2016; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017; SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, April 2018. SIGAR, *Counternarcotics: Lessons Learned from the U. S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. xii.
 776. SIGAR, *Counternarcotics: Lessons Learned from the U. S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. xii.
 777. UN Office on Drugs and Crime, “Afghanistan opium survey 2018: Challenges to sustainable development, peace and security,” July 2019, pp. 3, 8.
 778. Steve Goss, “Mid-Term Evaluation of the Community-Based Agriculture and Rural Development- West Project,” prepared for UNDP Afghanistan and INL, February 2019, p. 5.
 779. SIGAR’s January 2020 report presented the data as survival rates, with 60 percent of saplings in Farah and 38 percent in Badghis surviving. Numbers have been rounded. SIGAR, *Quarterly Report to the United States Congress*, January 30, 2020, p. 139; State, INL, “Quarterly Reporting Template for US/INL Funded Projects: Community-Based Agriculture and Rural Development–West (CBARD-West),” November 3, 2019, p. 2.

780. State, INL, "Quarterly Reporting Template for US/INL Funded Projects: Community-Based Agriculture and Rural Development–West (CBARD–West)," November 3, 2019, p. 2.
781. As SIGAR reported in April 2012, "Since 2002, Afghanistan has faced repeated natural disasters, causing substantial humanitarian crises. Floods, landslides, earthquakes, avalanches, and drought affect an average of 400,000 Afghans every year, according to USAID. Since 2000, Afghanistan has suffered eight serious droughts. In 2011, a drought affected approximately 2.4 million people in 14 provinces in the northern, northeastern, and western parts of the country." In April of 2016, before CBARD began implementation in Badghis and Farah, SIGAR reported on a deteriorating security environment including record-high civilian casualties the previous year and unusual violence through the normally quiet winter. The outgoing Resolute Support commander, General John F. Campbell, warned in early February that 2016 was at risk of being "no better, and possibly worse, than 2015." Evidently, UNDP recognized the need for, but did not develop, a formal management plan for responding to natural disasters and armed conflict. Whatever informal plan existed at the time of the sapling deaths "was not successfully implemented," according to the UNDP. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2012, p. 126; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2016, pp. 5–6; State, INL, "Quarterly Reporting Template for US/INL Funded Projects: Community-Based Agriculture and Rural Development–West (CBARD–West)," November 3, 2019, p. 17.
782. State, INL, "INL Quarterly Reporting Template for US/INL Funded Projects: Community-Based Agriculture and Rural Development–West (CBARD–West)," November 3, 2019, p. 2.
783. Steve Goss, "Mid-Term Evaluation of the Community-Based Agriculture and Rural Development– West Project," prepared for UNDP Afghanistan and INL, February 2019, p. 5; SIGAR analysis.
784. Asking difficult and uncomfortable questions does not seem to be a strength of evaluations. Yet doing so may reveal that programming should never have been implemented in the first place. Such a conclusion would represent a helpful data point given that State's evaluation policy requires bureaus and independent offices to use evaluations "for making decisions about policies [and] strategies." In reference to SIGAR's 2013 request to U.S. agencies to provide a list of their 10 most and 10 least successful reconstruction programs and projects, Inspector General Sopko stated in January 2020 congressional testimony: "And I go back to why some of you may have wondered why did I attach all of those letters from 2013 when I asked the [Secretary of Defense, the Secretary of State, and the] USAID administrator, what are your 10 best successes and what were your 10 worst failures and why. I firmly believe if they had honestly answered those questions, we wouldn't be here today. Because what [it] would have done is . . . force them to answer the question, why are we spending \$9 billion on narcotics if it's a failure?" John F. Sopko, testimony before the House Committee on Foreign Affairs, 116th Cong., January 15, 2020, electronic p. 41; State, "Diplomacy in Action: Department of State Evaluation Policy," January 29, 2015.
785. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 58.
786. State Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, p. 4; SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68 Audit Report, July 2015, p. i.
787. State Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, p. 8.
788. State Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, pp. 8, 10.
789. State, Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, p. 10.
790. State Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, p. 8.
791. As State's Inspector General commented in a 2013 audit report, the Corrections System Support Program's initial performance management plan included "elements such as a program goal, objectives, inputs, outputs, outcome measures, and a methodology for evaluating performance." According to State's "Program Design and Performance Management Toolkit," regular progress reviews are intended to "assess . . . results"—a process that is functionally equivalent to "evaluating performance." State Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, p. 8; State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 58.
792. According to State's inspector general, INL had not conducted any periodic reviews of CSSP, as of July 2013. The earliest CSSP periodic review document provided to SIGAR by State was a June 2014 "mid-year program check-in." State Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," September 2013, p. 8; State, INL, "CSSP Mid-Year Program Check-in," June 2014.
793. INL conducted both "mid-year program check-ins" and annual reviews. For example, see: State, INL, "CSSP Mid-Year Program Check-in," June 2014; State, "2014 Annual Program Review Corrections System Support Program (CSSP)," December 2014; State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY2017," June 2017, p. 16.
794. State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY2017," June 2017, p. 16.

795. State, "Department of State Program and Project Design, Monitoring, and Evaluation Policy," November 2017, p. 4.
796. State, "Department of State Program and Project Design, Monitoring, and Evaluation Policy," November 2017, p. 2.
797. State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY2017," October 2018, pp. 6–9.
798. In response to a draft of this report, INL stated, "If SIGAR has access to the data that would show INL exactly what to do in an intra-conflict reconstruction environment to ensure gains in rule of law and corrections development, it would behoove them to share it with INL so we could develop our programs in line with this hypothetical research that would allow for perfect causal linkages in scenarios never before encountered by the USG. [INL's Office of Afghanistan and Pakistan] developed the logic models in line with historical INL programming." SIGAR sympathizes with the difficulty of implementing programming in an "intra-conflict environment." However, results frameworks are not rigorous merely because one declares them to be. A truly "rigorous methodology" would fully account for significant barriers to success outside the control of individual programs, such as overoptimism or the possibility that objectives are unrealistic. With respect to the Corrections System Support Program and other rule of law programming, as SIGAR reported in its 2019 audit of Afghanistan's anti-corruption efforts, transformative change will remain elusive until the political will to address pervasive corruption materializes. Obstacles exogenous to the program are not the only complicating factors. A major component of CSSP is providing "mentoring, training, and professional assistance to the Ministry of Interior's Generate Directorate of Prisons and Detention Centers and the Ministry of Justice's Juvenile Rehabilitation Directorate"—in other words, capacity building. As a meta-review of evaluations of development assistance to Afghanistan commissioned by the German government noted, "Taken together, the 148 evaluation reports [examined for the review] suggest that capacity-building measures were mostly not successful." SIGAR, *Afghanistan's Anti-Corruption Efforts: The Afghan Government Made Progress in Meeting its Anti-Corruption Strategy Benchmarks, but Serious Challenges Remain to Fighting Corruption*, SIGAR-20-06-AR, November 2019, electronic p. 2; State, INL, "Corrections System Support Program (CSSP) Performance Measurement Plan (PMP) FY2017," June 2017, p. 6; State, INL, correspondence with SIGAR, May 10, 2021.
799. State, INL, "2014 Annual Program Review Corrections System Support Program (CSSP)," December 2014, p. 3; State, INL, correspondence with SIGAR, May 10, 2021.
800. State, INL, "2014 Annual Program Review Corrections System Support Program (CSSP)," December 2014, p. 3.
801. State, INL, "2014 Annual Program Review Corrections System Support Program (CSSP)," December 2014, p. 3.
802. State, INL, "2015 Annual Program Review Corrections System Support Program (CSSP)," June 2015, p. 3.
803. SIGAR analysis of State, "2015 Annual Program Review Corrections System Support Program (CSSP)," June 2015, p. 3.
804. INL attributed the change to an adjustment to "the office-wide [program review] template" and noted that the then-new template focused on how the program contributed to overall strategic goals, including fostering stability and confidence. Drawbacks with this method are described later in this section. State, "February 2017 Annual Program Review Corrections System Support Program (CSSP)," February 2017, p. 4; State, "2019 Annual Program Review Corrections System Support Program (CSSP)," January 2019, p. 4; State, INL correspondence with SIGAR, May 10, 2021; State, correspondence with SIGAR, May 10, 2021.
805. State, Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, pp. 8–9.
806. State, Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, p. 9.
807. State, Office of Inspector General, "Audit of the Bureau of International Affairs Correction System Support Program in Afghanistan," AUD-MERO-13-37, September 2013, p. 9.
808. State, INL correspondence with SIGAR, May 10, 2021.
809. State, *2014 Annual Program Review Corrections System Support Program (CSSP)*, December 2014, p. 3.
810. State, "2015 Annual Program Review Corrections System Support Program (CSSP)," June 2015, p. 4.
811. In response to a draft of this report and commenting on this specific sentence, INL stated, "This is not an objective statement and shows a lack of ability to see the correlation between this activity and strategic contribution." INL offered that the act of "exploring funding" included "talks with Afghan government officials, international organizations, and other partners to generate the political will required for an eventual program to be successful." However, "exploring funding" or participating in "talks" that involved "exploring funding" could be a problematic representation of strategic contribution. One could hold "talks" that do *not* generate political will. Or, one could hold such "talks," generate the political will to implement an ineffective program, implement the program, achieve no results, and still claim to have contributed to strategy through the "talks." State, INL correspondence with SIGAR, May 10, 2021; SIGAR analysis.
812. SIGAR analysis of State, "February 2017 Annual Program Review Corrections System Support Program (CSSP)," February 2017, pp. 5–6; State, "2019 Annual Program Review Corrections System Support Program (CSSP)," January 2019, pp. 5–6.

813. State, "February 2017 Annual Program Review Corrections System Support Program (CSSP)," February 2017, p. 5; State, "2019 Annual Program Review Corrections System Support Program (CSSP)," January 2019, p. 5.
814. Responding to a draft of this report, INL stated, "The goal of the evaluation was to measure the program's impact on the corrections system. Any scope larger than that would have yielded results even more equivocal, as there is no real way to show that one corrections program would contribute to a nationwide change in the perception of stability." However, one question evaluators sought to answer was "Have CSSP interventions improved conditions for and treatment of inmates *as well as the fairness and responsiveness of the overall justice system?*" [emphasis added]. Barbaricum, "Final Evaluation Report: INL Corrections System Support Program in Afghanistan," July 23, 2015, pp. 1, 3–4, 20; State, INL, correspondence with SIGAR, May 10, 2021.
815. Responding to a draft of this report, INL stated, "This paragraph demonstrates a distinct misunderstanding of the capability of anyone (State, third-party, or private company) to measure impact by implying that it is at all possible to show how one small program contributes to an incredibly broad US strategy." Again, CSSP's periodic program reviews included a specific prompt titled "Contribution to Strategic Goals." INL seems to be implying that it should be allowed to claim contribution to strategy while not being held accountable for whether that is actually the case. State, INL, correspondence with SIGAR, May 10, 2021; State, INL, "2014 Annual Program Review Corrections System Support Program (CSSP)," December 2014, electronic p. 3; State, "2015 Annual Program Review Corrections System Support Program (CSSP)," June 2015, p. 4.
816. ME&A, "Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation," April 2018, pp. viii, 30.
817. State Department officials, SIGAR interview, April 2019; State, INL, correspondence with SIGAR, May 10, 2021.
818. State Department officials, SIGAR interview, April 2019.
819. State Bureau of Population, Refugees, and Migration (PRM) and PRM implementing partners, SIGAR interview, May 2019.
820. State, 18 FAM 301.4-2.b(1), "Department of State Program and Project Design, Monitoring, and Evaluation," April 2018.
821. State, response to SIGAR data call, May 3, 2019.
822. State, "Department of State Program and Project Design, Monitoring, and Evaluation Policy," November 2017, p. 2.
823. State officials in the Bureau of South and Central Asian Affairs, SIGAR interview, February 2019.
824. State officials in the Bureau of South and Central Asian Affairs, SIGAR interview, February 2019.
825. ME&A, "Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation," April 2018, p. 36; State Office of Inspector General, "Review of Department of State Compliance with Program Evaluation Requirements," Report No. ISP-I-15-36, September 2015, p. 1.
826. Evaluation expert, SIGAR interview, August 2018.
827. State Office of Inspector General, "Management Assistance Report: Although Progress Has Been Made, Challenges Remain in Monitoring and Overseeing Antiterrorism Assistance Program Activities in Afghanistan," Report No. AUD-MERO-18-16, November 2017, p. 9.
828. State Office of Inspector General, "Management Assistance Report: Although Progress Has Been Made, Challenges Remain in Monitoring and Overseeing Antiterrorism Assistance Program Activities in Afghanistan," Report No. AUD-MERO-18-16, November 2017, pp. 5, 9–10.
829. Commission on Wartime Contracting in Iraq and Afghanistan, "Transforming Wartime Contracting: Controlling costs, reducing risks," August 2011, p. 123.
830. Commission on Wartime Contracting in Iraq and Afghanistan, "Transforming Wartime Contracting: Controlling costs, reducing risks," August 2011, p. 123; State Office of Inspector General, "Summary Guide to Overseas Contingency Operations Oversight Lessons Learned," Report No. OIG-OCO-16-01, September 2016 p. 15.
831. Commission on Wartime Contracting in Iraq and Afghanistan, "Transforming Wartime Contracting: Controlling Costs, Reducing Risks," August 2011, p. 122.
832. Commission on Wartime Contracting in Iraq and Afghanistan, "Transforming Wartime Contracting: Controlling Costs, Reducing Risks," August 2011, pp. 122–123.
833. State Department officials, SIGAR interview, April 2019.
834. State Office of Inspector General, "Management Assistance Report: Contract Terms and Guidance for Approving Student Training Expenses Relating to the Justice and Corrections Programs in Afghanistan Require Attention," Report No. AUD-MERO-18-14, October 2017, p. 4.
835. State Department officials, SIGAR interview, April 2019; State Office of Inspector General, "Audit of the Bureau of International Narcotics and Law Enforcement Affairs Invoice Review Process for Contracts in Afghanistan," Report No. AUD-MERO-18-30, February 2018, p. 10.
836. State Department officials, SIGAR interview, April 2019.
837. State Department officials, ISN officials, CNA interview, transmitted to SIGAR on or about December 11, 2017.

838. As used here, the term “portfolio” refers to a topical, or sector-level area such as rule of law. In addition to the work involved in managing multiple portfolios, the security situation militates against on-site assessments of progress, which could also affect INL’s ability to evaluate outcomes. State, INL officials, SIGAR interview, December 2019.
839. State Office of Inspector General, “Audit of the Bureau of International Narcotics and Law Enforcement Affairs Invoice Review Process for Contracts in Afghanistan,” Report No. AUD-MERO-18-30, February 2018, pp. 13–14.
840. State Office of Inspector General, “Management Assistance Report: Although Progress Has Been Made, Challenges Remain in Monitoring and Overseeing Antiterrorism Assistance Program Activities in Afghanistan,” Report No. AUD-MERO-18-16, November 2017, pp. 16–17.
841. Defense researcher, SIGAR interview, February 2019.
842. Marla C. Haims, Melina Moore, Harold D. Green, Jr. and Cynthia Clapp-Wincek, “Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects,” RAND Corporation, 2011, p. 1; Jake Sherman, “Survey of the United States Government’s arrangements for monitoring and evaluating support to security sector reform,” Saferworld, January 2009, p. 7; DOD, “Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise,” January 13, 2017, p. 1.
843. DOD, “Instruction 5132.14: Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise,” January 13, 2017, pp. 3, 14–15; SIGAR analysis.
844. DOD, “Instruction 5132.14: Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise,” January 13, 2017, pp. 14–15.
845. DOD, “Instruction 5132.14: Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise,” January 13, 2017, p. 15.
846. DOD, “Instruction 5132.14: Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise,” January 13, 2017, p. 16.
847. DOD, ODASD(SC), correspondence with SIGAR, May 13, 2021; SIGAR analysis.
848. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, pp. 72–73; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51.
849. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 1; CSTC-A officials, SIGAR interview, March 2020.
850. 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, p. 1.
851. 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, p. 2.
852. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 99.
853. The assessment’s core claim was that Afghanistan’s projected 2025 GDP was \$107.8 billion with task-force projects, compared to \$53.3 billion without those projects. However, as the assessment itself stated, “the mission of the Task Force for Business and Stability Operations . . . is to promote stability in Afghanistan.” Stability was to be achieved *through* the various economic initiatives implemented by TFBSO. Thus, although the assessment’s economic impact projections were somewhat useful, in not assessing whether TFBSO’s economic initiatives actually produced stability effects, the assessment did not actually measure whether TFBSO accomplished its end goal. SIGAR analysis of Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, pp. 13–14.
854. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. 1.
855. For example, see: DOD, USACE, “AIF Status Sheets,” August 2, 2018; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. ii.
856. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. 10.
857. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. i.
858. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 9.
859. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 31.

860. DOD, "Joint Publication 3-0: Joint Operations," October 22, 2018, p. II-8.
861. Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*: vol. 64, no. 4, article 9, pp. 5–8.
862. CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 39; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 31; SIGAR analysis; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
863. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. xii, 51; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
864. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. x.
865. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. x.
866. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. x–xi.
867. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. x–xi.
868. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xi.
869. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xi–xii.
870. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. xii, 100–102.
871. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xii.
872. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. xii, 144.
873. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xiii.
874. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xiii.
875. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xiii.
876. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. xiv; SIGAR, *Quarterly Report to the United States Congress*, July 30, 2020, p. 65.
877. White House, Office of the Press Secretary, "Remarks by President Biden on the Way Forward in Afghanistan," transcript, April 14, 2021; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 29.
878. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, pp. 55–56.
879. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 73; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51. Though not fully explored in this report, data availability may have been an additional reason for shifting assessment paradigms. Changing policy priorities affected U.S. troop levels, which in turn arguably had consequences for data collection and verification. For example, see: John F. Sopko, "Assessing the Capabilities and Effectiveness of the Afghan National Defense and Security Forces," testimony before the Senate Committee on Armed Services, 114th Cong., February 12, 2016, p. 5; Sean Kimmons, "Advising at the corps and below, US Soldiers ensure Afghans are ready to fight," *Army News Service*, October 8, 2019, p. 4.
880. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51.
881. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 7; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51.
882. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51.
883. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 72. While this Quarterly Report does not provide a start date for ANET, we believe the program generally started around January 2015 concurrently with the arrival of NATO advisors to Afghanistan. Issie Lapowsky, "Meet the Nerds Coding Their Way Through the Afghanistan War," *WIRED*, May 27, 2017.
884. For instance, the CUAT system was more complex than the CM rating system, while the RASR was designed to improve upon the CUAT by measuring the ability of ANDSF units to utilize their staffing, equipping, and training numbers. Additionally, while the prior assessment mechanisms measured the ANA down to the kandak (battalion) level, the MAAR assessed the ANDSF only at the headquarters level. The words used to

- describe each rating level also tended to change from system to system. For example, the lowest rating level in the MAAR was “In development.” By contrast, the lowest rating level of the CM was “May be capable of conducting portions of missions, reliant on support.” SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. 53–54; SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 6; SIGAR, *High Risk List*, January 2017, p. 15.
885. For example, the top three rating levels of the RASR (“Fully Capable,” “Capable,” and “Partially Capable”) are identical to the second, third and fourth ratings of the MAAR. One distinguishing feature of the MAAR, however, was that “Sustainable” was the highest achievable level, above “Fully Capable,” explicitly reflecting sustainability as a priority. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. 51, 53; DOD, “United States Plan for Sustaining the Afghanistan National Security Forces,” June 2008, pp. 6–7; SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 6.
886. For example, the CM rating measurement tools emphasized quantitative inputs, which masked qualitative measures of readiness. To correct for this, the CUAT framework included both quantitative and qualitative components. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 52; SIGAR, *Actions Needed to Improve the Reliability of Afghan Security Force Assessments*, SIGAR-10-11-AR, June 29, 2010, pp. 4, 7; SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 3.
887. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 51; SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 7.
888. DOD, “United States Plan for Sustaining the Afghanistan National Security Forces,” June 2008, pp. 6–7.
889. Military officials, SIGAR interview, October 2009; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 53.
890. Military officials, SIGAR interview, February 2010; military officials, SIGAR interview, October 2009; military official, SIGAR interview, March 2010.
891. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, September 2017, pp. 51–52.
892. SIGAR, *Actions Needed to Improve the Reliability of Afghan Security Force Assessments*, June 29, 2010, p. 6.
893. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. 51–52.
894. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 1; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 52.
895. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 53; SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 3.
896. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 7; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 53.
897. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, pp. 52–53.
898. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 7.
899. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 7.
900. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 8.
901. SIGAR, *Afghan National Security Forces: Actions Needed to Improve Plans for Sustaining Capability Assessment Efforts*, SIGAR-14-33-AR, February 2014, p. 8.
902. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2014, p. 94; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 53.
903. SIGAR, *High-Risk List*, January 2017, p. 15.
904. SIGAR, *High-Risk List*, January 2017, p. 15.

905. DOD, *Enhancing Security and Stability in Afghanistan*, June 2016, pp. 12, 21; SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 54.
906. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 54.
907. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 72; U.S. Digital Service, *Report to Congress*, July 2017, p. 11; DOD Office of Inspector General, "Operation Freedom's Sentinel: Lead Inspector General Report to the United States Congress, January 1, 2020–March 31, 2020," May 15, 2020, p. 23.
908. DOD Office of Inspector General, "Operation Freedom's Sentinel: Lead Inspector General Report to the United States Congress, January 1, 2020–March 31, 2020," May 15, 2020, pp. 23–24.
909. DOD Office of Inspector General, "Operation Freedom's Sentinel: Lead Inspector General Report to the United States Congress, January 1, 2020–March 31, 2020," May 15, 2020, pp. 23–24.
910. DOD Office of Inspector General, "Operation Freedom's Sentinel: Lead Inspector General Report to the United States Congress, January 1, 2020–March 31, 2020," May 15, 2020, p. 23.
911. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 72.
912. While the effectiveness of the revamped ANET system remains to be seen, there is a risk that it falls into a pattern in Afghanistan reconstruction of new, seemingly innovative solutions that ultimately do little to fix more fundamental underlying problems. This has happened to ANET once before, when coders from the Defense Digital Service modernized the software in 2017, eliminating bugs and increasing usability. The project garnered significant publicity. A 2017 *Wired* article titled "Meet the Nerds Coding Their Way through the Afghanistan War" described a "a scrappy band of military neophytes who deployed to a war zone, traversed mountains of red tape, and finally planted their flag of intuitive tech on the front lines." Although improving the ANET interface did make the software more efficient to use, it did little to fix the underlying challenge of accurately assessing capacity. This pattern extends beyond ANET. CSTC-A officials, SIGAR interview, March 2020; U.S. Digital Service, "Report to Congress," July 2017, p. 11; Issie Lapowsky, "Meet the Nerds Coding Their Way through the Afghanistan War," *WIRED*, May 27, 2017, pp. 16–17.
913. For example, in a February 2012 interview with NPR, Congressman Jim McGovern of Massachusetts said that Lt. Col. Daniel Davis' article in *Armed Forces Journal* (described later in the paragraph to which this endnote corresponds), "[confirmed] some of the concerns that [he] had over the years." Congressman McGovern added, "Since the war in Afghanistan began . . . we've been misled about the wars in Iraq and Afghanistan time and time again. And I think a lot of us are getting sick and tired of the whitewash." In testimony before the Congress, Special Inspector General Sopko has characterized ANDSF assessment mechanisms as "unreliable and inconsistent." NPR, "What's the Truth about the War in Afghanistan?," February 9, 2012; John F. Sopko, "Actions Needed to Improve U.S. Security-Sector Assistance Efforts in Afghanistan," testimony before the Subcommittee on National Security, Committee on Oversight and Government Reform, U.S. House of Representatives, 115th Congress, November 1, 2017.
914. General David Petraeus, "The Situation in Afghanistan," testimony before the Senate Committee on Armed Services, 112th Congress, March 15, 2011, pp. 2, 16.
915. Daniel Davis, "Truth, lies and Afghanistan: How military leaders have let us down," *Armed Forces Journal*, February 1, 2012, p. 2.
916. Daniel Davis, "Truth, lies and Afghanistan: How military leaders have let us down," *Armed Forces Journal*, February 1, 2012, p. 1.
917. Daniel Davis, "Truth, lies and Afghanistan: How military leaders have let us down," *Armed Forces Journal*, February 1, 2012, pp. 3–4.
918. CNA, "Independent Assessment of the Afghan National Security Forces," January 2014, p. i.
919. CNA, "Independent Assessment of the Afghan National Defense and Security Forces," November 2015, p. i; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
920. CNA, "Independent Assessment of the Afghan National Defense and Security Forces," November 2015, p. i; CNA, "Independent Assessment of the Afghan National Security Forces," January 2014, p. 1.
921. CNA, "Independent Assessment of the Afghan National Security Forces," January 2014, p. 3.
922. In 2014, CNA concluded that the force size should be 373,400, and in 2015, CNA concluded that it should be 383,800. CNA, "Independent Assessment of the Afghan National Security Forces," January 2014, p. 26; CNA, "Independent Assessment of the Afghan National Defense and Security Forces," November 2015, pp. iv–v.
923. CNA, "Independent Assessment of the Afghan National Security Forces," January 2014, p. 2; John F. Sopko, "Assessing the Capabilities and Effectiveness of the Afghan National Defense and Security Forces," testimony Before the Subcommittee on Oversight and Investigations, Committee on Armed Services U.S. House of Representatives, February 12, 2016, p. 3.
924. CNA, "Independent Assessment of the Afghan National Defense and Security Forces," November 2015, p. v.
925. SIGAR, *Reconstructing the Afghan National Defense and Security Forces: Lessons from the U.S. Experience in Afghanistan*, SIGAR-17-62-LL, September 2017, p. 188.
926. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 3.

927. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 3.
928. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 15; SIGAR analysis.
929. DOD, "Performance Work Statement (PWS) W560MY-15-C-0005MoD/ANA Advisors and Mentors," April 6, 2017, electronic pp. 19, 23.
930. DOD, "Performance Work Statement (PWS) W560MY-15-C-0005MoD/ANA Advisors and Mentors," April 6, 2017, electronic pp. 21–22; SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 3.
931. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 3.
932. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2017, p. 150.
933. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2017, p. 150.
934. Although DOD has described the Afghanistan Compact as "essentially a list of tasks that the Afghans were expected to do," Congress understood it as a way to "measure security-sector progress." DOD, "Enhancing Security and Stability in Afghanistan," June 2020, p. 48.
935. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2017, p. 150.
936. According to DOD, "about a year after the compact was inaugurated, the MOD and MOI largely completed the checklist and it became apparent that doing so did not translate into changes in institutional capacity. Thus, the list of compact related tasks expanded to more than 1,300, and the impact on outcomes remained ambiguous." DOD, "Enhancing Security and Stability in Afghanistan," June 2020, p. 49.
937. DOD, "Enhancing Security and Stability in Afghanistan," June 2020, p. 49.
938. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
939. Federal Acquisition Regulation, Part 37: "Service Contracting," Subpart 37.6: "Performance-based Acquisition," §37.601: "General," subsection (b)(2), §37.603: "Performance standards," subsection (a).
940. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 8.
941. ASI Government, "A COR's Guide to Statements of Work, Performance Work Statements, and Statements of Objectives," January 9, 2011, p. 1; SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 8.
942. DOD, "Performance Work Statement (PWS) W560MY-15-C-0005MoD/ANA Advisors and Mentors," April 6, 2017, p. 7-2.
943. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 9.
944. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. i.
945. Legacy awards required contractors to provide training and mentoring to Afghan intelligence officers, with success measured by the "capabilities and independence" of intelligence-training sites. "Intelligence Training Sites" were mostly exactly what the name sounds like: locations where the contractor developed Afghanistan's intelligence model and trained intelligence officers. However, the contractor was also responsible for institutionalizing the sites in a way that the Afghan government would eventually be able to take them over. Multiple contracts were awarded. The figure "\$457.7 million" refers to the total value of all contracts combined. Specific contractor requirements included developing doctrine and courses and hiring mentors and trainers. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, pp. i–ii, 2–3, 5–6.
946. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, pp. 2–3.
947. DOD also minimized the value of the R&D contracting vehicle by not waiting until each Legacy contract had ended before initiating new contracts under the program. As a result, these new contracts were awarded before lessons learned from older contracts could be fully integrated into the new awards. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, electronic p. 2, pp. 3, 10.

948. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. 3.
949. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. 3.
950. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. i.
951. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-Audit Report, July 2017, p. ii.
952. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. i.
953. SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. i.
954. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, pp. 2, 9.
955. For example, the December 2006 version of the U.S. Army's counterinsurgency field manual stated, "In times of turmoil, political, social, security, and economic benefits can often entice people to support one side or the other." SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 91–92; U.S. Army, "FM 3-24: Counterinsurgency," December 2006, p. 1-9.
956. FM 3-24 added, "COIN programs for political, social, and economic well-being are essential to developing the local capacity that commands popular support." U.S. Army, "FM 3-24: Counterinsurgency," December 2006, p. 2-1–2-2.
957. Karl Eikenberry, "The Limits of Counterinsurgency Doctrine in Afghanistan: The Other Side of the COIN," *Foreign Affairs*, September/October 2013.
958. Daniel P. Bolger, *Why We Lost: A General's Inside Account of the Iraq and Afghanistan Wars* (New York, NY: First Mariner Books, 2014), p. 237.
959. For example, see: Craig A. Collier, "Observations from a Year in the Sunni Triangle," *Small Wars Journal*, April 14, 2008, electronic p. 1; Gian P. Gentile, "Gaining the Initiative in Afghanistan," *Small Wars Journal*, September 2, 2009, electronic pp. 1–2; Stathis N. Kalyvas, "Counterinsurgency Manual Review," *Perspectives on Politics*, vol. 6, no. 2, June 2008, pp. 351–352.
960. Stathis N. Kalyvas, "Counterinsurgency Manual Review," *Perspectives on Politics*, vol. 6, no. 2, June 2008, pp. 351–352.
961. Criticism of counterinsurgency was not limited to Senate reports and often centered around FM 3-24's apparent reliance on historical anecdotes, and around the implication that counterinsurgency victories could be achieved at acceptable cost. As military historian Douglas Porch wrote in 2013, "Claims in doctrine for success in small wars, at least at a reasonable strategic, financial, and moral cost, have relied on mythologized versions of the past too often supported by shoddy research and flawed, selective analysis of cases." Such critiques are often qualified by their acknowledgement that aspects of counterinsurgency theory may be valid under particular circumstances—though they frequently also emphasize that some of the central assumptions of FM 3-24 were unproven. For example, regarding the use of development aid to combat insurgencies, Porch says, "Economic development may prove an antidote to insurgency—surging demand for rubber and tin with the Korean War created a wave of trickle-down prosperity in Malaya that helped to diminish the allure of insurgency there, while also funding counterinsurgency operations. However, the success of strategies of occupation anchored in developmental theory that seek to purchase loyalty, bolster legitimacy and good governance with economic or infrastructure improvements remain unproven." U.S. Senate Committee on Foreign Relations, "Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations," 112th Congress, 1st session, June 8, 2011, pp. 8–9; Douglas Porch, *Counterinsurgency: Exposing the Myths of the New Way of War* (Cambridge, UK: Cambridge University Press, 2013), pp. xii, 330.
962. Although the report focused on funding appropriated to State and USAID and did not cover aid provided by the U.S. military, it raised questions about the underlying intervention logic for all stabilization programs, noting that the U.S. military's Commander's Emergency Response Program—the subject of this section of the chapter—warranted "closer scrutiny." U.S. Senate Committee on Foreign Relations, "Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations," 112th Congress, 1st session, June 8, 2011, pp. 1, 8–10.
963. U.S. Senate Committee on Foreign Relations, "Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations," 112th Congress, 1st session, June 8, 2011, p. 9.

964. 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. 2.
965. 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.
966. U.S. Senate Committee on Foreign Relations, "Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations," 112th Congress, 1st session, June 8, 2011, p. 31; 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. 13.
967. Christoph Zürcher, "What Do We (Not) Know About Development Aid and Violence? A Systematic Review," *World Development*, vol. 98, p. 511.
968. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 97–98, 101.
969. Mark Martins, "No Small Change of Soldiering: The Commander's Emergency Response Program (CERP) in Iraq and Afghanistan," *The Army Lawyer*, Department of the Army Pamphlet 27-50-369, February 2004, pp. 1, 3.
970. Quoted in 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. 55.
971. Quoted in 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. 2.
972. SIGAR has previously described CERP as part of the U.S. military's attempt to rebuild Afghanistan from the bottom up. A second, "bottom-up" approach to counterinsurgency in Afghanistan was DOD's Village Stability Operations (VSO) program. We focus on CERP here because of its explicit connection to contracting. Implemented by U.S. Special Operations Forces from 2010 to 2014, VSO aimed to stabilize villages believed to have strategic value. The program sought to connect these villages to the Afghan central government via district and provincial governments, respectively. VSO is another example of a DOD initiative for which impact was often assumed through intuition—or, at best, through art rather than science. For example, describing VSO's purported successes, former ISAF commander General David Petraeus told SIGAR, "Blocks were turning from red to amber to green. And if you do this week after week after week, you can feel whether it's working or not. I also visited many sites and was comfortable that, while very difficult, the concept was working." SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 91, 107, 118; USFOR-A, "Money as a Weapon System–Afghanistan," March 2012, pp. ii, 119–120.
973. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 95; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
974. CERP was authorized for a variety of purposes, including force protection, and for condolence and battle damage payments. USFOR-A, "Money as a Weapon System–Afghanistan," May 15, 2009, pp. 145–146; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 91.
975. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 91–92.
976. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 94; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
977. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 92; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
978. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 92.
979. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 92–93, 96–97; ISAF Joint Command, "CERP Strategy in Afghanistan (DRAFT)," August 29, 2010, p. 2.
980. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 92–93.
981. 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. 89.
982. In January 2009 the FMR was amended to restrict the use of "bulk CERP" for large-scale projects. Projects in Afghanistan over \$2 million required CENTCOM approval. The FMR added that "small-scale" projects were those costing less than \$500,000, and that not many projects should cost more than that. The increasing emphasis on smaller scale projects stretches back to at least 2005 and 2006 when the Congress set clearer guidelines on project size. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 94, 260.
983. Emergency Supplemental Appropriations Act for Defense and for the Reconstruction of Iraq and Afghanistan, 2004, Pub. L. No. 108-106, §1110 (2003).
984. According to State, "the primary goal of U.S. humanitarian assistance is to save lives and alleviate suffering." As defined in SIGAR's enabling legislation, the term "reconstruction" refers to building or rebuilding "physical infrastructure," establishing or reestablishing "a political or societal institution," or providing "products

- or services to the people of Afghanistan.” State, “Refugee and Humanitarian Assistance,” n.d., accessed May 26, 2021; National Defense Authorization Act for Fiscal Year 2008, Pub. L. No. 110-181 (2008), §1229.
985. Mark Martins, “No Small Change of Soldiering: The Commander’s Emergency Response Program (CERP) in Iraq and Afghanistan,” *The Army Lawyer*, Department of the Army Pamphlet 27-50-369, February 2004, p. 2.
 986. Dov Zakheim, “Senate Appropriations Subcommittee on Defense Holds Hearing on Defense Appropriations,” testimony before the Senate Appropriations Subcommittee on Defense, March 1, 2004, pp. 7, 34.
 987. Mark Martins, “The Commander’s Emergency Response Program,” *Joint Forces Quarterly*, issue 37, second quarter (2005), p. 49.
 988. 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. 48.
 989. 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. 47.
 990. Stuart W. Bowen, Jr. and Craig Collier, “Reconstruction Leaders’ Perceptions of CERP in Iraq: Report Overview,” *PRISM*, vol. 4, no. 1, 2012, p. 120.
 991. Stuart W. Bowen, Jr. and Craig Collier, “Reconstruction Leaders’ Perceptions of CERP in Iraq: Report Overview,” *PRISM*, vol. 4, no. 1, 2012, p. 122.
 992. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
 993. 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, pp. 48–49.
 994. Doug Struck, “U.S. Using Cash as a Defensive Weapon,” *Washington Post*, 26 July, 2004, A14; Dan Murphy, “The GF’s Weapon of Choice in Iraq: Dollars,” *Christian Science Monitor*, 29 January 2004, p. 1.
 995. USFOR-A, “Money as a Weapons System–Afghanistan,” May 15, 2009, pp. 145–147; Stuart W. Bowen, Jr. and Craig Collier, “Reconstruction Leaders’ Perceptions of CERP in Iraq: Report Overview,” *PRISM*, vol. 4, no. 1, 2012, p. 120.
 996. USFOR-A, “Money as a Weapons System–Afghanistan,” May 15, 2009, pp. 145–147. The series of MAAWS publications were superseded in 2017 by the publication of U.S. Army, “The Commanders’ Emergency Response Program, May 2017, p. v.
 997. Center for Army Lessons Learned, “Commander’s Guide to Money as a Weapon System: Tactics, Techniques, and Procedures,” April 2009, p. i.
 998. 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. 5.
 999. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 94.
 1000. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 94–95; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
 1001. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 92.
 1002. The inherent flexibility of CERP and the authority of local commanders was reflected in the 2009 FMR, which stipulated that CERP was “designed to enable local commanders in Iraq and Afghanistan to respond to urgent humanitarian relief and reconstruction requirements within their areas of responsibility by carrying out programs that will immediately assist the indigenous population. As used here, reconstruction does not limit efforts to restore previous conditions/structure[s] in Afghanistan. Also, as used here, urgent is defined as any chronic or acute inadequacy of an essential good or service that, in the judgment of a local commander, calls for immediate action.” SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 92, 259.
 1003. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 96; USFOR-A, “Money as a Weapons System – Afghanistan,” May 15, 2009, pp. 87–88, 159–160; USFOR-A, “Money as a Weapons System – Afghanistan,” December 2009, pp. 4, 92–94.
 1004. U.S. Army, “FM 3-24: Counterinsurgency,” December 2006, p. 1-27.
 1005. U.S. Army, “FM 3-24: Counterinsurgency,” December 2006, p. 1-27.
 1006. USFOR-A, “Money as a Weapons System – Afghanistan,” May 15, 2009, pp. 87–88.
 1007. USFOR-A, “Money as a Weapons System – Afghanistan,” May 15, 2009, p. 87.
 1008. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 97.
 1009. DOD, U.S. Army Audit Agency, “Commander’s Emergency Response Program, U.S. Forces – Afghanistan,” Audit Report A-2011-0020-ALL, November 16, 2010, p. 5.
 1010. DOD, U.S. Army Audit Agency, “Commander’s Emergency Response Program, U.S. Forces – Afghanistan,” Audit Report A-2011-0020-ALL, November 16, 2010, p. 18.
 1011. USFOR-A, “Money as a Weapons System – Afghanistan,” February 2011, pp. 71–72.
 1012. USFOR-A, “Money as a Weapons System – Afghanistan,” March 2012, pp. x, 120–123.
 1013. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 99.
 1014. 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. xviii.
 1015. 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. xviii.
 1016. As SIGAR’s lessons learned report on U.S. stabilization programming details, numerous SIGAR interviews with personnel who attempted to develop metrics revealed they struggled to devise methods to measure

- CERP projects' impact on the counterinsurgency fight. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 97–98.
1017. As a former senior USFOR-A official in charge of CERP implementation told SIGAR in an interview for SIGAR's report on stabilization programming, "We did a good job of vetting projects and making sure all the boxes were checked, but we had no infrastructure in place to assess the effectiveness of projects. We only cared about questions like, 'Does the project meet legal [and] ethical requirements? Did the Army Corps of Engineers and other required entities sign off on it?'" Former senior USFOR-A official, SIGAR interview, July 2016.
 1018. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 98.
 1019. USFOR-A, "Money as a Weapons System – Afghanistan," March 2012, p. 134.
 1020. USFOR-A, "Money as a Weapons System – Afghanistan," March 2012, p. 134.
 1021. USFOR-A, "Money as a Weapons System – Afghanistan," March 2012, electronic pp. 3–4.
 1022. Further, use of the multiplier metric seemed to contradict separate MAAWS-A guidance to "invest time in developing precise economic measures of effectiveness." USFOR-A, "Money as a Weapons System," March 2012, electronic p. 9, p. 134; SIGAR analysis.
 1023. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 100–101.
 1024. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 101.
 1025. The insistence that CERP was effective has been suggested as one reason why the program received such substantial scrutiny. As a RAND impact assessment of CERP commissioned by DOD stated, "Given the argued importance of CERP, at least from the perspective of senior military leaders, in providing force protection for deployed forces and achieving the goals of counterinsurgency, it has received substantial scrutiny from military and academic scholars." Of course, the military was not alone in advocating for CERP, as many civilian officials echoed the sentiment that the program was a potent counterinsurgency tool. In 2009 testimony to Congress, then-Secretary of Defense Robert Gates said that CERP was the "single most effective program to enable commanders to address local populations' needs and get potential insurgents in Iraq and Afghanistan off the streets and into jobs" (quoted in Egel et al, p. 55). Views of CERP's efficacy within the military may have remained fairly rigid. Although interviewees selected for the qualitative component of RAND's impact evaluation of CERP in Afghanistan were not randomly selected, the perspectives they offered are largely consistent with the thinking at the time of CERP's widespread use in Afghanistan that CERP was generally effective. As the RAND report stated, "The commanders we interviewed were . . . unanimous in their support for CERP." 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, pp. 45, 55; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 101.
 1026. General James Mattis, "Senate Armed Services Committee Holds Hearing on the Nomination of Gen. James Mattis to U.S. Central Command," testimony before the Senate Armed Services Committee, July 27, 2010, pp. 29–30.
 1027. USFOR-A, "Money as a Weapons System – Afghanistan," March 2012, p. viii.
 1028. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 97.
 1029. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 100.
 1030. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
 1031. The period of 2008 through 2011 saw by far the highest levels of CERP disbursements in Afghanistan. Five of the six studies of CERP examined in the systematic review were published after 2011. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 95; Christoph Zürcher, "What Do We (Not) Know About Development Aid and Violence? A Systematic Review," *World Development*, vol. 98, pp. 506, 510.
 1032. For example, a 2013 paper authored by Berman et al. used "attack data . . . based on 'significant activity' (SIGACT) reports by Coalition forces that capture a wide variety of information about ' . . . executed enemy attacks targeted against coalition, Iraqi Security Forces (ISF), civilians, Iraqi infrastructure and government organizations'" that provided "the location and date of attack incidents between February 2004 and February 2009." For development spending, Berman et al. utilized "data from the U.S. Army Corps of Engineers Gulf Region Division's Iraq Reconstruction Management System (IRMS)" that included "the start date, end date, project description, funding source, and amount spent for 62,966 aid projects awarded through December 2008." Eli Berman, Joseph Felter, Jacob Shapiro, and Erin Troland, "Modest, Secure, and Informed: Successful Development in Conflict Zones," National Bureau of Economic Research working paper 18674, 2013, pp. 10, 12.
 1033. 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, pp. 75–76; Christoph Zürcher, "What Do We (Not) Know About Development Aid and Violence? A Systematic Review," *World Development*, vol. 98, pp. 509, 511.
 1034. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 99; Christoph Zürcher, "What Do We (Not) Know About Development Aid and Violence? A Systematic Review," *World Development*, vol. 98, p. 510.
 1035. The effects of CERP independent of counterinsurgency operations linked to CERP spending have generally proven difficult to disentangle. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*,

- SIGAR-18-48-LL, May 2018, p. 99; 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. 162.
1036. Christoph Zürcher, “What Do We (Not) Know About Development Aid and Violence? A Systematic Review,” *World Development*, vol. 98, pp. 506, 510.
1037. As the systematic review stated, “Likewise, for aid which is given by the military, only one out of six studies finds an unqualified violence-dampening effect (Berman et al., 2011). A subsequent study by the same authors qualifies that result in important ways. Berman et al. (2013) find that the violence dampening effect actually occurs only in interaction with increased troop presence.” Christoph Zürcher, “What Do We (Not) Know About Development Aid and Violence? A Systematic Review,” *World Development*, vol. 98, p. 510–511.
1038. An earlier study jointly funded by DOD and DHS examined the effectiveness of CERP in Iraq. DOD, OUSD-P, “Comments from the Office of the Under Secretary of Defense for Policy,” provided in response to a draft of SIGAR audit SIGAR-18-45-AR on the Commander’s Emergency Response Program, April 27, 2018, p. 2; 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. iii; DOD, correspondence with SIGAR, April 27, 2018.
1039. 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. xi; SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 92.
1040. 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, pp. 11, 100, 102.
1041. 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, pp. xi–xii.
1042. 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. xii.
1043. More specifically, according to the study, “both economic and security conditions drive internal migration in developing countries,” meaning that “these data function as a proxy measure for overall changes in economic and security conditions.” 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, pp. 163–164, 225–226.
1044. 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, pp. xii, 167.
1045. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 100.
1046. SIGAR, *Increased Visibility, Monitoring, and Planning Needed for Commander’s Emergency Response Program*, SIGAR-09-5-AR, p. 3.
1047. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1048. As one expert on the data bluntly summarized, “SIGACTs are never accurate.” The expert added, “A lot of the fields in the SIGACT reports are listed as ‘unknown,’ so each report has varying levels of information.” Others are less categorical in their critiques, pointing, for example, to significant omissions in the data. Ben Connable, “Embracing the Fog of War: Assessment and Metrics in Counterinsurgency,” RAND Corporation, 2012, pp. 161–162.
1049. Eli Berman, Jacob N. Shapiro, and Joseph H. Felter, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, vol. 119, no. 4 (August 2011), pp. 808–809.
1050. Ben Connable, “Embracing the Fog of War: Assessment and Metrics in Counterinsurgency,” RAND Corporation, 2012, p. 163.
1051. Renard Sexton, “Aid as a Tool against Insurgency: Evidence from Contested and Controlled Territory in Afghanistan,” *American Political Science Review*, vol. 110, no. 4, (November 2016), pp. 736, 737, 742; Tiffany Chou, “Does Development Assistance Reduce Violence? Evidence from Afghanistan,” *The Economics of Peace and Security Journal*, vol. 7, no. 2 (2012), p. 7; Eli Berman, Jacob N. Shapiro, and Joseph H. Felter, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, vol. 119, no. 4 (August 2011), p. 792.
1052. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, pp. 100, 128; Greg Adams, “Honing the Proper Edge: CERP and the Two-sided Potential of Military-led Development in Afghanistan,” *The Economics of Peace and Security Journal*, vol. 10, no. 2 (2015), p. 55; SIGAR analysis.
1053. “Consistent” here refers to the possibility of a consistent tendency for units operating in areas of high violence to underreport incidents. To be clear, there would still be random variation in reporting thresholds. Eli Berman, Jacob N. Shapiro, and Joseph H. Felter, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, vol. 119, no. 4 (August 2011), pp. 808–809.
1054. Eli Berman, Jacob N. Shapiro, and Joseph H. Felter, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, vol. 119, no. 4 (August 2011), pp. 801–802.
1055. Eli Berman, Jacob N. Shapiro, and Joseph H. Felter, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, vol. 119, no. 4 (August 2011), pp. 806, 810.

1056. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 128; Renard Sexton, "Aid as a Tool against Insurgency: Evidence from Contested and Controlled Territory in Afghanistan," *American Political Science Review*, vol. 110, no. 4 (November 2016), p. 735.
1057. 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, pp. xii, 12.
1058. The inclusion of dependent variables other than violence provides a way of examining intermediate counterinsurgency outcomes. However, as intermediate outcomes like economic development or improved governance were ultimately assumed to impact popular support for the insurgency, one would expect that, perhaps at a later point in the presumed causal chain, violence levels would be the most appropriate long-term outcome metric. 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. xv.
1059. Previous research on CERP suggested that smaller CERP projects (under \$50,000) were violence-reducing. A prominent explanation for the violence-reducing effect found by some researchers is that CERP projects encourage local communities to share information with a host-nation government and its allies. 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, pp. 198–199; Christoph Zürcher, "What Do We (Not) Know About Development Aid and Violence? A Systematic Review," *World Development*, vol. 98, p. 511.
1060. According to RAND, "This dual result for the relationship between CERP activity and attacks involving coalition forces is consistent with previous analyses exploring this relationship (compared with Gorkowski, 2009; Berman, Shapiro, and Felter, 2011; Chou, 2012; Berman et al., 2013; Clark and Jackson, 2013; Jackson and Clark, 2015)." 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, pp. 88, 198–199.
1061. Previous research on CERP focused largely on violence as an outcome measure, rather than economic effects or non-violence-related coalition-focused outcomes. 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, pp. xv, 198.
1062. 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. 179.
1063. 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. 202.
1064. 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, pp. xi, 202.
1065. 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. 197.
1066. As one former U.S. official put it more directly and starkly, "Economic development efforts have not been just wasteful but have been institutionalized within COIN in unique ways which need to be reexamined." 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. 5; former U.S. government official, SIGAR interview, August 2015.
1067. 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.
1068. 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, pp. 9, 11, 13.
1069. U.S. Senate Committee on Appropriations, Department of Defense Appropriations Bill, 2011, S. Rep. No. 111-295, at 207 (2010), 111th Congress, 2nd Session, September 16, 2010, pp. 6, 207.
1070. U.S. Senate Committee on Appropriations, Department of Defense Appropriations Bill, 2011, S. Rep. No. 111-295, at 207 (2010), 111th Congress, 2nd Session, September 16, 2010, p. 207.
1071. GAO, "Afghanistan Development: Agencies Could Benefit from a Shared and More Comprehensive Database on U.S. Efforts," GAO-13-34, November 2012, p. 8.
1072. DOD's AIF funding requests for FYs 2012, 2013, and 2014 were all premised on the notion that AIF projects were needed to support the U.S. COIN strategy in Afghanistan. In response to a draft of this report, DOD stated that agencies' joint statement in responses to SIGAR's 2017 audit of FY 2017 AIF projects provides useful, additional context regarding the origins of AIF. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12, July 2012, p. 2; Department of Defense and Full-Year Continuing Appropriations Act, 2011, Public Law 112-10, April 15, 2011, electronic p. 140; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, ii; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1073. Department of Defense and Full-Year Continuing Appropriations Act, 2011, Public Law 112-10, April 15, 2011, electronic pp. 140–141.
1074. Department of Defense and Full-Year Continuing Appropriations Act, 2011, Public Law 112-10, April 15, 2011, electronic p. 140; SIGAR analysis of DOD, response to SIGAR quarterly report data call, March 18, 2020.

1075. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1076. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12 July 2012, electronic p. 2.
1077. The major exception to large power-infrastructure construction under AIF was the Kandahar Bridging Solution, which provided temporary power to two industrial parks in Kandahar City. The intent was to better meet electricity demand until transmission lines and substations could be constructed. It too had counterinsurgency goals. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2019, p. 149; DOD, USFOR-A JENG, “Fiscal Year 2011–2014 Afghanistan Infrastructure Fund (AIF) Program Status Report,” December 21, 2015. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1078. Department of Defense and Full-Year Continuing Appropriations Act, 2011, Pub. L. No. 112-10, 125 Stat. 38 (2011); U.S. Government Interagency Counterinsurgency Initiative, U.S. Government Counterinsurgency Guide, January 2009; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, i, p. 1.
1079. The SIGAR audit on USACE’s third-party monitoring program technically reviewed ANDSF construction projects funded by ASFF rather than AIF-funded projects. However, we extrapolate the findings to AIF. SIGAR, *USACE’s Local National Quality Assurance Program: USACE Used Qualified Personnel to Monitor Construction in Afghanistan and Is Taking Steps to Improve Contractor Reporting*, SIGAR-19-60-AR, September 2019, i; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, ii.
1080. For example, in an audit of FY 2011 AIF projects completed in July 2012, SIGAR reported that five of seven FY 2011 AIF projects were six to 15 months behind schedule. In an October 2017 follow-up audit, SIGAR reported that three FY 2011 AIF power sector projects were up to five years behind schedule. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12 July 2012, electronic p. 2; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. 13.
1081. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. 13.
1082. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, pp. 147–148; DOD, USFOR-A JENG, “Fiscal Year 2011–2014 Afghanistan Infrastructure Fund (AIF) Program Status Report,” December 21, 2015; SIGAR, *Quarterly Report to the United States Congress*, July 30, 2020, p. 146.
1083. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12 July 2012, p. 7.
1084. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12 July 2012, electronic p. 2.
1085. As DOD told SIGAR in response to a draft of SIGAR’s October 2017 follow-on audit of FY 2011 AIF projects, “the primary impact of increasing electrical capacity cannot be assessed until the projects are complete and electricity is available.” DOD, OUSD-P, “Department of Defense Comments on SIGAR Draft Audit Report,” October 18, 2017, p. 1; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 147; SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 2019, p. 26; USAID, “Power Transmission Expansion and Connectivity Project,” May 7, 2019; SIGAR, *Quarterly Report to the United States Congress*, July 30, 2020, p. 145.
1086. For the agencies’ perspectives on the audit’s conclusions, see the responses beginning on p. 28 of SIGAR’s 2017 audit. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. ii.
1087. DOD, OUSD-P, “Department of Defense Comments on SIGAR Draft Audit Report,” July 13, 2012, p. 6.
1088. DOD, OUSD-P, “Department of Defense Comments on SIGAR Draft Audit Report,” July 13, 2012, p. 4.
1089. DOD, OUSD-P, “Department of Defense Comments on SIGAR Draft Audit Report,” October 18, 2017, p. 1.
1090. Particularly following the apparent success of the 2007 surge in Iraq, belief in counterinsurgency doctrine became widespread within the U.S. military. This became a frequent topic of discussion among counterinsurgency critics. For example, in a 2009 interview with PBS, history professor and retired Colonel Andrew Bacevich described counterinsurgency as a “fad” that had been widely embraced by the American officer corps. Reflecting a similar sentiment in a 2014 book about the wars in Iraq and Afghanistan, historian and retired Lt. Gen. Daniel Bolger went so far as to describe FM 3-24 as the “holy writ.” PBS, “Interview: Col. Andrew Bacevich (Ret.),” Frontline interview, October 13, 2009; Daniel P. Bolger, *Why We Lost: A General’s Inside Account of the Iraq and Afghanistan Wars* (New York, NY: First Mariner Books, 2014), p. 237.
1091. In response to SIGAR’s 2017 audit of AIF projects, DOD objected to SIGAR’s use of the phrase “intuitive process,” stating, “the report highlights comments made by project managers in the U.S. Forces-Afghanistan’s

- (USFOR-A) engineering office, personnel with no role in measuring counterinsurgency impacts.” However, the supposed intuitiveness of AIF’s COIN impact is evident in statements made by other DOD officials. For example, in response to SIGAR’s July 2012 Afghanistan Infrastructure Fund (AIF) report, the Deputy Assistant Secretary of Defense for Afghanistan, Pakistan, and Central Asia stated, “The COIN impact of some of these projects is already apparent.” It is also worth noting that although U.S. Army Corps of Engineers officials told SIGAR that they only “build things,” those “things” are built for a purpose. According to the Corps’ Transatlantic Division, which oversees activities in the Middle East and Central Asia, the Corps not only executes construction and engineering operations in support of NATO forces in Afghanistan, but also “provides sustainable development projects for the Afghan people that employ the local populace, build skilled human capital and promote the future stability of [Afghanistan].” SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, pp. 10, 29; DOD, USFOR-A JENG and USACE, SIGAR interview, May 2015; DOD, USACE, Transatlantic Division, “Transatlantic Division Public Website,” accessed September 1, 2020; DOD, USACE, Transatlantic Division, “Transatlantic Division Public Website – Careers – Deployments,” accessed September 1, 2020.
1092. USFOR-A JENG and USACE officials, SIGAR interview, May 2015; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2018, p. 179.
 1093. Responding to a draft of this report, DOD reiterated the view that assessment of effects could be “intuitive”: “When businesses and factories re-open or operate for longer hours due to additional electricity, you don’t need sophisticated studies to determine if its providing a benefit. You can assess it in more detail if needed, but some things (like electricity) tend to offer more obvious benefits.” The comment included a reference to FM 3-24, the U.S. Army’s Counterinsurgency Field Manual. USFOR-A JENG and USACE officials, SIGAR interview, May 2015; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
 1094. A legal analysis for the Bridging Solution accompanying the U.S. Central Command memorandum added, “Presently, the residents of Kandahar City have limited electricity. With no electricity, the factories cannot function and employ residents.” DOD, U.S. CENTCOM, Office of the Commander, “Memorandum for Under Secretary of Defense for Policy (USD(P)) and Under Secretary of Defense (Comptroller) (USD),” June 5, 2010, p. 1; DOD, USFOR-A(S) SJA, “Memorandum for Chief of Staff, United States Forces, Afghanistan South, Kandahar, Afghanistan,” May 22, 2010, p. 1.
 1095. In the course of SIGAR’s 2017 audit of FY 2011 AIF projects, DOD pointed to Asia Foundation survey data that, according to DOD, showed that “reconstruction efforts were directly related to people’s perception of how Afghanistan is doing. Based on more than 9,500 interviews with Afghans from across all 34 Afghan provinces, the 2015 survey asked participants if they think Afghanistan is going in the right direction or the wrong direction, and why. Of the respondents who think the country is going in the right direction, reconstruction/rebuilding was selected as the top reason for the third straight year. In fact, reconstruction/rebuilding has been the either the first or second reason mentioned by participants every year since 2007. Respondents are not asked about specific projects, but the results are an indication that reconstruction efforts are viewed positively by the Afghan people.” SIGAR agrees that reconstruction can have positive effects. However, this response does not address whether the views of survey respondents who believed Afghanistan was moving in the *wrong* direction could be swayed by projects similar to those implemented with AIF funds. This omission is particularly salient given that the 2015 Asia Foundation survey showed that the percentage of respondents who believed that the situation in Afghanistan was “going in the right direction” had declined in two consecutive years. DOD, OSD-P, correspondence with SIGAR, December 23, 2015; The Asia Foundation, “A Survey of the Afghan People,” November 10, 2015, pp. 16, 97.
 1096. USFOR-A JENG and USACE officials, SIGAR interview, May 2015; Eli Berman, Michael Callen, Joseph H. Felter and Jacob N. Shapiro, “Do Working Men Rebel: Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines,” *Journal of Conflict Resolution*, March 22, 2011, vol. 55, no. 4, p. 496; Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 3; 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.
 1097. Regarding counterinsurgency theory more broadly, some academics have levied scathing criticism. For example, Yale political scientist Stathis Kalyvas wrote that FM 3-24 “betrays zero impact by [current] political science research.” Quoted in: Gian Gentile, *Wrong Turn: America’s Deadly Embrace of Counter-Insurgency* (New York, NY: The New Press, 2013), p. 26; 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, pp. 1, 9.
 1098. Eli Berman, Michael Callen, Joseph H. Felter and Jacob N. Shapiro, “Do Working Men Rebel: Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines,” *Journal of Conflict Resolution*, March 22, 2011, vol. 55, no. 4, p. 496.
 1099. As the authors stated, “Our findings are consistent with two hypotheses concerning counterinsurgency: (1) as local economic conditions deteriorate, government forces and their allies are able to buy more intelligence on insurgents (i.e., the price of information falls), and (2) actions taken to enhance security—establishing checkpoints, building walls, and the like—damage the economy.” Related to the first hypothesis, the authors argue here and elsewhere that development dollars can be used to fund small-scale projects that

- incentivize noncombatants to share information that can be exploited to combat insurgents. For a brief description of this “information-sharing” model of counterinsurgency, see: Eli Berman, Jacob N. Shapiro, and Joseph H. Felter, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, August 2011, vol. 119, no. 4, p. 768; Eli Berman, Michael Callen, Joseph H. Felter and Jacob N. Shapiro, “Do Working Men Rebel: Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines,” *Journal of Conflict Resolution*, March 22, 2011, vol. 55, no. 4, pp. 498, 519.
1100. Eli Berman, Michael Callen, Joseph H. Felter and Jacob N. Shapiro, “Do Working Men Rebel: Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines,” *Journal of Conflict Resolution*, March 22, 2011, vol. 55, no. 4, p. 519.
 1101. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 3.
 1102. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, pp. 3, 12.
 1103. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 26.
 1104. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. i.
 1105. Total obligations include those for contracts, grants, task orders and interagency agreements, the latter of which are not covered in this report. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 1.
 1106. Initial TFBSO director Paul Brinkley, for example, had previously worked in Silicon Valley. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 3; Rebecca Zimmerman, Daniel Egel, and Ilana Blum, “Task Force for Business and Stability Operations Lessons from Afghanistan,” RAND Corporation, 2016, pp. 72–73.
 1107. Former senior TFBSO official, SIGAR interview, November 2015.
 1108. As used here, “results” refers to contract deliverables. Of the \$316.3 million TFBSO obligated for direct program costs, SIGAR determined that approximately 26 percent was obligated to contracts that met few to none of TFBSO’s deliverables, 52 percent was obligated to contracts that partially met deliverables or met them after significant delay, and 22 percent was obligated to contracts that mostly or fully met contract deliverables. TFBSO has been covered extensively by SIGAR in other reports. For example, see SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, pp. 152–160; and SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, pp. 23–24.
 1109. Contracting officer’s representatives also did not always submit required reports on time. SIGAR, “DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects,” SIGAR-18-19-AR, January 2018, p. 21.
 1110. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 13.
 1111. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 13.
 1112. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 32.
 1113. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 32.
 1114. A 2018 SIGAR audit of TFBSO observed that developing monitoring and evaluation plans for all projects might improve a future, task-force-like entity’s ability to achieve results. S. Rebecca Zimmerman, Daniel Egel, and Ilana Blum, “Task Force for Business and Stability Operations Lessons from Afghanistan,” RAND Corporation, 2016, pp. 90–91; SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. ii.
 1115. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 9.
 1116. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, pp. 6, 10.
 1117. While the task force was not the only entity involved in Afghanistan reconstruction to rely on such “success stories,” it “appears to have relied on these ad hoc stories and outputs virtually exclusively,” in the words of a DOD-commissioned lessons learned report on TFBSO produced by RAND. S. Rebecca Zimmerman, Daniel Egel, and Ilana Blum, “Task Force for Business and Stability Operations Lessons from Afghanistan,” RAND Corporation, 2016, p. 90.
 1118. TFBSO ended its operations in Afghanistan in December 2014, the same month in which the DOD-commissioned economic impact of TFBSO was completed. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. 1; Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact

- Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, pp. 1, 15–16.
1119. The core claim of the assessment was that Afghanistan’s projected 2025 GDP was \$107.8 billion with task-force projects, compared to \$53.3 billion without those projects. Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, p. 14.
 1120. Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, p. 14.
 1121. The projects referenced here are those involving the exploration and production of Afghan minerals. Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, p. 72.
 1122. One contract—for Herat Province’s Shaida copper mine—remains active and could produce economic benefits in the future. Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, pp. 189–190; SIGAR, *Quarterly Report to the United States Congress*, January 30, 2020, p. 123; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 138; SIGAR, *Quarterly Report to the United States Congress*, October 30, 2018, p. 145.
 1123. Overoptimism is not limited to programmatic assessment. Broader macroeconomic analysis has also been shaky at times. For example, the IMF has continually revised downward its GDP-growth-rate projections for Afghanistan. See: SIGAR, *Quarterly Report to the United States Congress*, January 30, 2019, pp. 148–149.
 1124. Again, the core claim of the assessment was that Afghanistan’s projected 2025 GDP was \$107.8 billion with task-force projects, compared to \$53.3 billion without those projects. This is a projection of economic impact and does not address stability impacts. Vestige Consulting, LLC, and Acertas, LLC, “Economic Impact Assessment Task Force for Business & Stability Operations (TFBSO) in Afghanistan,” December 29, 2014, pp. 13, 14.
 1125. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, pp. 7, 9; SIGAR analysis.
 1126. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, pp. 4–5.
 1127. “Joint Explanatory Statement of the Committees on Armed Services of the U.S. Senate and House of Representatives on H.R. 6523, Ike Skelton National Defense Authorization Act for Fiscal Year 2011,” December 22, 2010, p. 126.
 1128. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. ii.
 1129. DOD, OSD-P, correspondence with SIGAR, October 18, 2017, p. 1.
 1130. Secretary of Defense Robert M. Gates and Secretary of State Hillary Rodham Clinton, correspondence with the Honorable Daniel Inouye, Chairman, Committee on Appropriations, November 19, 2010, p. 1.
 1131. SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. 4; DOD, OUSD(C), “United States Department of Defense Fiscal Year 2012 Budget Request, Overview,” February 2011, p. 6-7; DOD, OUSD(C), “United States Department of Defense Fiscal Year 2013 Budget Request, Overview,” February 2012, p. 6-6; DOD, OUSD(C), “United States Department of Defense Fiscal Year 2014 Budget Request, Overview, Addendum A: Overseas Contingency Operations,” May 2013, p. 7.
 1132. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12, July 2012, p. 22.
 1133. DOD, HQDA, “FM 3-24: Counterinsurgency,” December 2006, p. 5-28.
 1134. General James Mattis, “Defense Authorization: Overview,” testimony before the Senate Armed Services Committee, March 1, 2011, *Congressional Quarterly* transcript, p. 13.
 1135. General David Petraeus, “Senate Armed Service Committee Holds Hearing on the Situation in Afghanistan,” testimony before the Senate Armed Services Committee, March 15, 2011, p. 98.
 1136. DOD, OSD-P, correspondence with SIGAR, October 18, 2017.
 1137. However, if projects *are* significantly delayed and/or at risk of not being sustained upon completion, it would be appropriate to determine, mid-course, the projected costs and benefits of continued investment.
 1138. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2018, p. 150.
 1139. DOD, OSD-P, correspondence with SIGAR, December 23, 2015.
 1140. DOD, OSD-P, correspondence with SIGAR, July 13, 2012.
 1141. USAID, correspondence with SIGAR, October 17, 2017.
 1142. USAID, response to SIGAR quarterly reporting vetting, April 7, 2021.
 1143. USAID, Office of Infrastructure, response to SIGAR quarterly report data call, June 17, 2021.
 1144. USAID’s baseline survey covered a very small area—two 20-kilometer zones surrounding substations constructed in Ghazni and Wardak Provinces. Data collection was limited to secure areas within those zones, raising questions about generalizability and limiting the sample size. Although USAID’s midline survey will

- gather additional baseline data on the areas surrounding five additional substations, aspects of the baseline survey raise questions about analytical quality. For example, the dimensions of impact analysis that are arguably most relevant for assessing counterinsurgency outcomes are those for governance and security. Although the baseline survey was not, in itself, an impact evaluation, the survey did run regressions that tested the association between a connection to the electrical grid and governance and security indicators. Some results selected for the survey's discussion of findings related to governance and security did not meet the threshold for statistical significance defined by the survey, making them odd choices for discussion. In one instance, a statistically significant but counterintuitive result in one of the survey's regression models appeared to be ignored in favor of a more intuitive result that was not statistically significant. Additionally, the issue of practical significance is not directly addressed in the survey's discussion of key findings related to governance and security. The future utility of USAID's quasi-experimental impact evaluation will depend on the clear and comprehensive presentation of its results. Management Systems International, "Electricity Access and Socioeconomic Baseline Survey for Energy Infrastructure Activities," May 2020, pp. iii, x–xi, 2, 28, 50–51, 128, 130, 132–133; USAID, OI, correspondence with SIGAR, June 21, 2021; SIGAR analysis.
1145. In response to a draft of this report, DOD stated, "I suppose we could have used scarce time and resources to study the impacts of the construction phase (jobs and such), but what would the results tell us? If few jobs were created and/or there was no evidence of benefits during construction, would we cancel the projects and walk away? No. If many jobs were created and there was tremendous evidence of positive benefits prior to completion, would we use that as a basis for planning more projects? No. Therefore, it's not clear what intermediate assessments would accomplish." SIGAR does not share this view. Agencies must hold themselves accountable to claims made to the Congress and taxpayer that serve as justification for funding. In this case, a key claim was that AIF projects would contribute to counterinsurgency objectives. That claim has not been verified. Additionally, the rhetorical questions offered by DOD omit the reality that reconstruction projects can be *destabilizing*. SIGAR has made no determination as to whether this was or will be the case for AIF-funded projects. But intermediate assessment puts agencies in a better position to identify destabilizing effects and correct course. Finally, given the centrality of counterinsurgency theory to DOD's economic development efforts in Afghanistan and the still-evolving discussion around whether economic development can be leveraged to achieve stability or security effects—and under what circumstances—intermediate assessment could have provided agencies with valuable information to refine their assumptions and beliefs around what does and not does work. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. vi; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. ii; DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1146. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, p. I-13.
1147. As one researcher told SIGAR, "the first closest existing analogue that DOD has for M&E would be independent assessment" (discussed earlier in this chapter), and "the second would be operation assessment." Defense researcher, SIGAR interview, February 2019.
1148. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, pp. I-13, I-14, xxvi; DOD, "Joint Publication 3-0: Joint Operations," October 22, 2018, p. II-9.
1149. Andrew Williams, James Bexfield, Fabrizio Farina, and Johannes de Nijs, "Innovation in Operations Assessment: Recent Developments in Measuring Results in Conflict Environments," NATO Headquarters Supreme Allied Commander Transformation, 2013, p. 64.
1150. ISAF, "Commander's Initial Assessment, August 30, 2009.
1151. COMISAF, "COMISAF Assessment," January 25, 2011, p. 1.
1152. CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, pp. 44–45.
1153. As JP 5-0 states, an "assessment plan should focus on the information and intelligence that directly support the commander's decision making." DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, p. xxvi; ISAF Headquarters, "2009 COMISAF's Initial Assessment," August 30, 2009, p. 1-3; CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 44.
1154. The challenge of shifting assessment paradigms brought about by changes in command was one insight emerging from a Military Operations Research Society (MORS) conference hosted by U.S. Central Command on November 5–8, 2012. According to the conference's report on proceedings, "An important point . . . was that commanders want different inputs, which was illustrated when General John Allen took over ISAF from General David Petraeus and asked for a completely different assessment process." The conference is described in a separate CNA report (produced for SIGAR but unpublished) as a useful summary of the "broad outlines" of the debate surrounding how operation assessments should be conducted. The RAND report referenced here also noted, "trying to find the 'right' metrics is an exercise in futility" as "there are no inherently 'good' or 'bad' metrics in COIN." MORS, "MORS Special Meeting: Assessments of Multinational Operations from Analysis to Doctrine and Policy," November 2012, p. 8; Ben Connable, "Embracing the Fog of War: Assessment and Metrics in Counterinsurgency," RAND Corporation, 2012, p. xxi, 46; CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 44.

1155. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, p. 1; SIGAR, *Private Sector Development and Economic Growth: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-38-LL, September 2016, p. 150.
1156. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, electronic p. 2, p. 2.
1157. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, p. 13.
1158. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, p. 13.
1159. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, pp. ii, 13.
1160. SIGAR, *Afghan First Initiative Has Placed Work with Afghan Companies, but Is Affected by Inconsistent Contract Solicitation and Vetting, and Employment Data is Limited*, SIGAR Audit-12-6, January 2012, p. 11.
1161. Schroden points to a “failure cycle” beginning with deficiencies in assessment doctrine that, exacerbated by lack of training for assessment practitioners, result in poor processes and products which, because they do not meet expectations, lead to commander disinterest in the assessment process altogether. Although he at times elevates other explanatory factors, Ben Connable of RAND arrives at a similar conclusion: “The lack of a comprehensive methodology, the inability to account for inconsistencies in centralized COIN assessment, and the onerous demands of a coalition bureaucracy have resulted in an assessment process that arguably does not meet most of the standards for quality assessment.” Chapter Eight of Connable’s monograph on COIN assessment and metrics is titled, “Why Does Centralized Assessment Fail in Counterinsurgency?” Ben Connable, “Embracing the Fog of War: Assessment and Metrics in Counterinsurgency,” RAND Corporation, 2012, p. 154–155, 171; Jonathon Schroden, “Why Operations Assessments Fail: It’s Not Just the Metrics,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 89.
1162. Other research affirms that the issues described in these two articles were core problems underlying operations assessment in Afghanistan. For example, echoing Schroden’s point around doctrinal deficiencies, Connable notes that between 2008 and 2011, commanders, policy makers and assessment staff “had very little doctrine to fall back on.” He also points to the challenge of finding the right mix of qualitative and quantitative information: “[Commanders and staff officers] fell back on heavily quantitative processes at first and then struggled to figure out how to capture contextual narrative and popular sentiment.” Citing initially to Downes-Martin’s observation that “a color-coded map hides information,” Connable goes on to provide his own analysis of the methodological flaws and logical failures in such maps. Finally, regarding the tendency of assessments to be overly optimistic, he observes, “It is clear that Vietnam-era analysts were under considerable pressure to deliver positive reports. Research into the Iraq and Afghanistan campaigns revealed similar pressures on analysts in both theaters.” He adds, “Observation and interviews for this study showed that the majority of this observed and reported pressure in Iraq and Afghanistan was implicit and probably nowhere near as egregious as during the Vietnam campaign. This kind of pressure is typically nuanced, and it is sometimes unintentional or self-generated.” Ben Connable, “Embracing the Fog of War: Assessment and Metrics in Counterinsurgency,” RAND Corporation, 2012, pp. 154, 191, 201–202; Jonathon Schroden, “Why Operations Assessments Fail: It’s Not Just the Metrics,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, pp. 92, 94–95; Stephen Downes-Martin, “Operations Assessment in Afghanistan Is Broken—What Is to Be Done?,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, pp. 103, 108, 109–110, 111–112, 115.
1163. Regarding rigor, Schroden points out conflicting guidance provided by individual doctrinal publications such as “U.S. Army Field Manual 5-0.” On the one hand, the then-current version of FM 5-0 stated, “detailed analysis is to be avoided, that ‘committing valuable time and energy to developing excessive and time-consuming assessment schemes squanders resources better devoted to other operations process activities.’” On the other, the FM said, “establishing cause and effect is sometimes difficult, but crucial to effective assessment. Commanders and staffs are well advised to devote the time, effort, and energy needed to properly uncover connections between causes and effects.” Jonathon Schroden, “Why Operations Assessments Fail: It’s Not Just the Metrics,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, pp. 92, 94.
1164. Jonathon Schroden, “Why Operations Assessments Fail: It’s Not Just the Metrics,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 95.
1165. Examples of quantitative flaws included trend analyses of polling data that failed to account for margins of error. In other words, movement of polling data in a certain direction was sometimes claimed even when changes fell within the polling margin of error and there was in fact no statistically observable trend. Stephen Downes-Martin, “Operations Assessment in Afghanistan Is Broken—What Is to Be Done?,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, pp. 103, 109–110, 115.
1166. Additionally, if data collection requirements are too onerous, they may be impossible to meet without compromising basic data integrity. According to Connable, “some [military officers] . . . make up data to avoid irreconcilable friction and to allow them to focus on what they see as more important tasks.” Ben Connable, “Embracing the Fog of War: Assessment and Metrics in Counterinsurgency,” RAND Corporation, 2012, p.

- 168; Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 108.
1167. Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*, Autumn 2011, vol. 64, no. 4, pp. 111–112.
1168. Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 108.
1169. Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 108.
1170. Former senior military official, SIGAR, interview, November 2015.
1171. For Stewart's scathing account of overoptimism in the military, see: Rory Stewart and Gerald Knaus, *Can Intervention Work*, (New York, NY: W.W. Norton & Company, 2011), pp. 50–58. Stewart adds, "Every year from 2005, more and more money was spent, more and more troops were deployed. Each year was to prove decisive. None was. . . . These attitudes were not confined to the center: they echoed down the hierarchy in every province." John Glaser and John Mueller of the CATO Institute similarly describe this phenomenon in an August 2019 policy analysis arguing for an end to the war in Afghanistan. Regarding military overoptimism, Glaser and Mueller write, "Overly optimistic portrayals are partly a result of institutional habits and a view about civil-military relations that calls for focusing on tactical and operational facts on the ground while leaving broader strategic and political assessments of the war to elected leaders. Some military leaders publicly misrepresented the course of the war to avoid the hit to troop morale they expected would result from more honest and critical presentations. Others felt strongly that negotiations with the Taliban should only occur from a 'position of strength,' which they believed was always just around the corner. But sometimes the deception was more flagrant: Media reports revealed in 2011 that commanders tasked with briefing congressional delegations in Afghanistan deliberately misled members of Congress about the progress of the war." Such observations were the premise of Lt. Col. Daniel Davis' report that sought to dispel "rosy official statements." As Inspector General Sopko said in January 2020 testimony before the Congress, "We have incentivized lying to Congress and by that, I mean the whole incentive is to show success and to ignore the failure and when there's too much failure, classify it or don't report it." For more, see John Glaser and John Mueller, "Overcoming Inertia Why It's Time to End the War in Afghanistan," CATO Institute, Policy Analysis no. 878, August 13, 2019, pp. 5–6; John F. Sopko, "U.S. Lessons Learned in Afghanistan," testimony before the Committee on Foreign Affairs, 116th Congress, January 15, 2020, electronic p. 42.
1172. Barack Obama, "Against Going to War in Iraq: Remarks by Illinois State Senator Barack Obama," Organization for Action (OFA), OFA website, accessed on April 20, 2017; Barack Obama, "Obama's Remarks on Iraq and Afghanistan," *New York Times*, July 15, 2008; Athena Jones, "Obama defends 'just' war in Afghanistan," *NBC News*, December 10, 2009.
1173. Rajiv Chandrasekaran, *Little America: The War Within the War for Afghanistan* (New York, NY: Alfred A. Knopf, 2012), p. 53.
1174. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 26.
1175. The assessment was leaked to Bob Woodward of the *Washington Post*, whose September 21, 2009, article on the topic was titled, "McChrystal: More Forces or 'Mission Failure.'" Regarding Afghan government corruption, Woodward wrote, "The assessment offers an unsparing critique of the failings of the Afghan government, contending that official corruption is as much of a threat as the insurgency to the mission of the International Security Assistance Force." Bob Woodward, "McChrystal: More Forces or 'Mission Failure,'" *Washington Post*, September 21, 2009.
1176. Months before the assessment, McChrystal described his "strong belief in the importance of a holistic counterinsurgency campaign" in his nomination hearing before the U.S. Senate on June 2, 2009. ISAF, "Commander's Initial Assessment, August 30, 2009, p. 1-3; Lt. Gen. Stanley A. McChrystal, testimony before the U.S. Senate Committee on Armed Services, June 2, 2009; ISAF, "Commander's Initial Assessment," August 30, 2009, p. i.
1177. ISAF, "Commander's Initial Assessment, August 30, 2009, pp. 1-4-2-1.
1178. ISAF, "Commander's Initial Assessment, August 30, 2009, p. 1-4.
1179. ISAF, "Commander's Initial Assessment, August 30, 2009, p. 2-4.
1180. ISAF, "Commander's Initial Assessment, August 30, 2009, pp. 2-10, 2-18.
1181. ISAF, "Commander's Initial Assessment, August 30, 2009, p. 2-1.
1182. ISAF, "Commander's Initial Assessment, August 30, 2009, pp. 1-1, 1-3.
1183. ISAF, "Commander's Initial Assessment, August 30, 2009, pp. 1-1, 1-3.
1184. ISAF, "Commander's Initial Assessment, August 30, 2009, p. 1-1.
1185. Karl W. Eikenberry, "The Limits of Counterinsurgency Doctrine in Afghanistan," *Foreign Affairs*, vol. 92, no. 5 (September/October 2013), p. 61.
1186. Karl W. Eikenberry, "The Limits of Counterinsurgency Doctrine in Afghanistan," *Foreign Affairs*, vol. 92, no. 5 (September/October 2013), p. 61.
1187. Former U.S. government official, SIGAR interview, August 2015.
1188. Task and Purpose's reporting relied on an audio recording of a talk McChrystal gave to a small group as part of his December 2018 book tour. McChrystal declined to expand further on his comments in a subsequent phone call with *Task and Purpose*. Paul Szoldra, "General McChrystal Told Pompeo to 'Muddle Along' in Afghanistan, Leaked Audio Reveals," *Task & Purpose*, December 6, 2018.

1189. Former U.S. government official, SIGAR interview, August 2015.
1190. Rajiv Chandrasekaran, *Little America: The War Within the War for Afghanistan* (New York, NY: Alfred A. Knopf, 2012), p. 59.
1191. Rajiv Chandrasekaran, *Little America: The War Within the War for Afghanistan* (New York, NY: Alfred A. Knopf, 2012), p. 59.
1192. Rajiv Chandrasekaran, *Little America: The War Within the War for Afghanistan* (New York, NY: Alfred A. Knopf, 2012), p. 59.
1193. SIGAR has had numerous internal debates over the years about how to interpret and present the litany of security-oriented metrics described by Chandrasekaran.
1194. ISAF, “Commander’s Initial Assessment, August 30, 2009, p. 1-1.
1195. ISAF, “Commander’s Initial Assessment, August 30, 2009, p. 2-20.
1196. Jonathon Schroden, “Why Operations Assessments Fail: It’s Not Just the Metrics,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 90.
1197. For example, see: Ben Connable, “Embracing the Fog of War: Assessments and Metrics in Counterinsurgency,” RAND Corporation, 2012, pp. 4-5.
1198. Recognizing the importance of district control as a proxy for reconstruction success, SIGAR publicly reported Resolute Support district control data from 2016 to 2018. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2016, pp. 95-96; U.S. Department of the Army and Marine Corps, “Insurgencies and Countering Insurgencies, Field Manual no. 3-24/Marine Corps Warfighting Publication no. 3-33.5,” May 2014, p. 5-2.
1199. Ben Connable, “Embracing the Fog of War: Assessments and Metrics in Counterinsurgency,” RAND Corporation, 2012, pp. 4-5.
1200. While the data was available, SIGAR reported on population and territorial control trends. For example, see: SIGAR, *Quarterly Report to the United States Congress*, April 30, 2018, p. 87.
1201. DOD Office of Inspector General, “Report on Operation Freedom’s Sentinel Report to the United States Congress January 1, 2018–March 31, 2018,” May 21, 2018, pp. 16-17.
1202. DOD Office of Inspector General, “Report on Operation Freedom’s Sentinel Report to the United States Congress January 1, 2018–March 31, 2018,” May 21, 2018, p. 17.
1203. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019, p. ii.
1204. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019, p. ii.
1205. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019, p. ii.
1206. For example, see the April 3, 2019 exchange on this issue between Inspector General Sopko and Congressman Stephen Lynch, chair of the Subcommittee on National Security, House Committee on Oversight and Reform, as well as IG Sopko’s January 15, 2020 testimony on this issue before the House Foreign Affairs Committee. John F. Sopko, “Special Inspector General for Afghanistan Reconstruction’s 2019 High-Risk List,” testimony before the House Oversight and Reform Subcommittee on National Security Hearing, April 3, 2019; John F. Sopko, “House Foreign Affairs Committee Holds Hearing on Lessons Learned in Afghanistan,” testimony before the House Foreign Affairs Committee, January 15, 2020.
1207. DOD, “Report on Enhancing Security and Stability in Afghanistan,” June 2020, p. 45.
1208. Some reports used “insurgent activity” and “high insurgent activity” as euphemisms for “insurgent influence” and “insurgent control.” According to at least two reports, this was only a shift in terminology and did not reflect a change in methodology. However, it is notable that “insurgent activity” is not a direct opposite to “government influence/control”; nor is it semantically equivalent to “insurgent control.” Both reports claimed that the terminology change was necessary in order to make the information publically releasable. Furthermore, “influence” is a challenging term, with no clear objective distinction between an influenced district and a contested one. *The Long War Journal* eliminated the “influence” rating and rolled the category into “contested,” resulting in a three-point scale. *Long War Journal* argued that this method was more analytically sound, as the Afghan government and the Taliban could both influence different elements of the same district, resulting in a “contested” rating. The journal also replaces the generic “insurgent” with “Taliban.” SIGAR, *Quarterly Report to the United States Congress*, April 30, 2018, p. 88; SIGAR, *Quarterly Report to the United States Congress*, October 30, 2018, p. 70; Anthony Cordesman, “Afghanistan’s Uncertain Conflict Metrics 2017–2018,” CSIS, September 12, 2018, p. 67.
1209. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2016, p. 96; DOD Office of Inspector General, “Operation Freedom’s Sentinel Report to the United States Congress January 1, 2018–March 31, 2018,” May 21, 2018, p. 16.
1210. DOD Office of Inspector General, “Operation Freedom’s Sentinel Report to the United States Congress January 1, 2018–March 31, 2018,” May 21, 2018, p. 16; State, “Integrated Country Strategy: Afghanistan,” September 27, 2018, pp. 3-5.
1211. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2016, p. 96.
1212. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2018, p. 88.
1213. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2016, p. 96.
1214. DOD Office of Inspector General, “Operation Freedom’s Sentinel Report to the United States Congress January 1, 2018–March 31, 2018,” May 21, 2018, p. 17.

1215. Conversely, American forces and Afghan commandos have historically been able to access and commit acts of violence in Taliban-controlled districts. The fact that both parties enter the other's controlled territory and commit violence may raise questions about the broad concept of control. DOD Office of Inspector General, "Operation Freedom's Sentinel Report to the United States Congress January 1, 2018–March 31, 2018," May 21, 2018, p. 17.
1216. Migration may also reflect harsh economic conditions and data is typically not nuanced enough to differentiate between wage-earners traveling temporarily for work and whole families relocating permanently or semi-permanently because of security issues. DOD Office of Inspector General, "Operation Freedom's Sentinel Report to the United States Congress January 1, 2018–March 31, 2018," May 21, 2018, p. 18; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2018, p. 88.
1217. DOD Office of Inspector General, "Operation Freedom's Sentinel Report to the United States Congress January 1, 2018–March 31, 2018," May 21, 2018, p. 18.
1218. DOD, "Report on Enhancing Security and Stability in Afghanistan," June 2020, p. 45.
1219. See the April 3, 2019, exchange on this issue between Inspector General Sopko and Congressman Stephen Lynch, Chair of the Subcommittee on National Security, House Committee on Oversight and Reform and IG Sopko's January 15, 2020 testimony on this issue before the House Foreign Affairs Committee. SIGAR, transcript of John F. Sopko, "Special Inspector General for Afghanistan Reconstruction's 2019 High-Risk List," testimony before the House Oversight and Reform Subcommittee on National Security Hearing, April 3, 2019; John Sopko, "House Foreign Affairs Committee Holds Hearing on Lessons Learned in Afghanistan," testimony before the House Foreign Affairs Committee, *Congressional Quarterly* transcripts, January 15, 2020.
1220. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2018, p. 88.
1221. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2019, p. 74.
1222. Luke Kelly, "Lessons on the effectiveness of risk management units in reducing fiduciary risk," report commissioned by the UK Department for International Development, pp. 5–6; Marc Jacquand and Shelley Ranii, "UN Development System Risk Management in Fragile States," New York University Center on International Cooperation, May 2014, accessed July 6, 2020, p. 6; Independent Commission for Aid Impact, "About Us," accessed July 14, 2020.
1223. Nicole Ball, Sue Emmott, Maja Greenwood, Najib Murshed and Pablo Uribe, "Afghanistan: Lessons Identified 2001–2014: Part II: Development Cooperation in Afghanistan," Landell Mills Development Consultants, 2016, p. 14.
1224. Julia Steets, Elias Sagmeister, and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," final report of the Secure Access in Volatile Environments Project, October 2016, pp. 9, 23.
1225. According to Stoddard and Jillani (2016), among other purposes, increased data sharing can help illuminate the extent of the humanitarian footprint in a given country (such as which agencies are operating where, with how many people and how much money), allowing for better measurement of the impact of insecurity on humanitarian aid. Stoddard and Jillani point to several reasons why agencies may be unwilling to share data, including "conflicting pressures to keep their specific locations and activities quiet for the security of their staff and programmes on the one hand, and on the other hand to exaggerate the extent of their presence for funding public relations purposes, demonstrating to donors and the general public that they are capable of going where needed." Humanitarian Outcomes, "Secure Access in Volatile Environments: Overview of M&E practices in Afghanistan and discussion of priorities for further research, SAVE," February 2015, PowerPoint, p. 14; Julia Steets, Elias Sagmeister, and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," final report of the Secure Access in Volatile Environments Project, October 2016, pp. 10, 24, 31; Abby Stoddard, Shoaib Jillani, John Caccavale, Peyton Cooke, David Guillemois, and Vassily Klimentov, "The effects of insecurity on humanitarian coverage," final report of the Secure Access in Volatile Environments Programme, November 2016, p. 13.
1226. For example, overestimating the degree to which insecure areas could be accessed and the capacity to collect data. Humanitarian Outcomes, "Secure Access in Volatile Environments: Overview of M&E practices in Afghanistan and discussion of priorities for further research, SAVE," February 2015, PowerPoint, p. 17; Julia Steets, Elias Sagmeister, and Lotte Ruppert, "Eyes and Ears on the Ground: Monitoring aid in insecure environments," final report of the Secure Access in Volatile Environments Project, October 2016, p. 38.
1227. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. 6.
1228. On-budget assistance refers to donor funds that are aligned with Afghan government plans, included in Afghan government budget documents, and included in the budget approved by the parliament and managed by the Afghan treasury system. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2020, p. 110; SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR April 2018, pp. 1, 3.
1229. World Bank, Independent Evaluation Group, "Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group," 2016, p. viii.

1230. World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d., p. 3.
1231. World Bank, "The World Bank's Approach to Grievance Redress in Projects," February 28, 2013, p. 2; World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d., p. 2.
1232. The role that International Relief and Development Inc. (now known as Blumont) began to play in building the capacity of Afghan ministries implementing ARTF-supported projects shows how the line between monitoring and implementation is sometimes blurred. World Bank, "Third Party Monitoring Programme for the Afghanistan Reconstruction Trust Fund: A Review, March 2014, accessed July 2, 2020, pp. 4, 24; World Bank, "Environmental and Social Framework for IPF Operations ESS1: Assessment and Management of Environmental and Social Risks and Impact," June 2018, p. 11; SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How it Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. ii.
1233. World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d.; World Bank, Independent Evaluation Group, "Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group," 2016, p. viii.
1234. World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d.
1235. World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d.
1236. World Bank, "July 10, 2019 Heads of Cooperation Discussion In-depth Fiduciary Review of EQUIP II," draft brief, July 10, 2019, p. 1; SIGAR, *Quarterly Report to the United States Congress*, October 30, 2019, pp. 115, 152–153.
1237. SIGAR, *Quarterly Report to the United States Congress*, October 30, 2019, p. 115.
1238. World Bank, "July 10, 2019 Heads of Cooperation Discussion In-depth Fiduciary Review of EQUIP II," draft brief, July 10, 2019, p. 2; SIGAR analysis.
1239. World Bank, "July 10, 2019 Heads of Cooperation Discussion In-depth Fiduciary Review of EQUIP II," draft brief, July 10, 2019, p. 2.
1240. World Bank, "July 10, 2019 Heads of Cooperation Discussion In-depth Fiduciary Review of EQUIP II," draft brief, July 10, 2019, p. 1.
1241. World Bank, "July 10, 2019 Heads of Cooperation Discussion In-depth Fiduciary Review of EQUIP II," draft brief, July 10, 2019, p. 2; SIGAR analysis.
1242. World Bank, "July 10, 2019 Heads of Cooperation Discussion In-depth Fiduciary Review of EQUIP II," draft brief, July 10, 2019, p. 2.
1243. World Bank, "Afghanistan Reconstruction Trust Fund Strategy Group Ad Hoc Meeting on Fiduciary Risk in the ARTF Portfolio," Meeting Summary, July 24, 2019, p. 1.
1244. Concerns surrounding the results of the review mirrored those previously expressed by SIGAR. As SIGAR reported in a 2018 audit of the Afghanistan Reconstruction Trust Fund, "Donors, including USAID and U.S. taxpayers, will continue to provide billions of dollars to the ARTF without adequate assurance that the money being spent is actually contributing support to the ongoing operation of the Afghan government, addressing fiscal gaps in Afghanistan's national budget, and furthering the international community's overall reconstruction efforts." SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, pp. 22–23.
1245. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. ii.
1246. Third-party monitor expatriate staff are based primarily in Kabul. World Bank, "Third Party Monitoring Programme for the Afghanistan Reconstruction Trust Fund: A Review," March 2014, pp. 4, 6, 27, 32.
1247. Here, the clause "a misperception that could discourage donors from also using feedback from project-internal systems to assess progress" is a SIGAR interpretation of the review, which is not specific on this point. What is clearer is the review's conclusion that some donors view the supervisory agent as the primary monitoring mechanism for ARTF-funded development projects. According to the review, "This perception risks undermining the impact of the comprehensive monitoring and evaluation systems the projects have in place." World Bank, "Third Party Monitoring Programme for the Afghanistan Reconstruction Trust Fund: A Review," March 2014, p. 5.
1248. World Bank, "Third Party Monitoring Programme for the Afghanistan Reconstruction Trust Fund: A Review," March 2014, pp. 5, 7.
1249. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, pp. 1, 3.
1250. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. 3.

1251. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. i–ii.
1252. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. ii.
1253. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. iii.
1254. Marc Jacquand and Shelley Ranii, “UN Development System Risk Management in Fragile States,” New York University Center on International Cooperation, May 2014, accessed July 6, 2020, pp. 6, 12, 14.
1255. Luke Kelly, “Lessons on the effectiveness of risk management units in reducing fiduciary risk,” report commissioned by the UK Department for International Development, July 31, 2019, p. 8.
1256. Luke Kelly, “Lessons on the effectiveness of risk management units in reducing fiduciary risk,” report commissioned by the UK Department for International Development, July 31, 2019, p. 9.
1257. Luke Kelly, “Lessons on the effectiveness of risk management units in reducing fiduciary risk,” report commissioned by the UK Department for International Development, July 31, 2019, p. 9.
1258. Luke Kelly, “Lessons on the effectiveness of risk management units in reducing fiduciary risk,” report commissioned by the UK Department for International Development, July 31, 2019, p. 9.
1259. United Nations, Risk Management Unit Afghanistan, Technology, Monitoring and Evaluation: The Use of Technology by UN Agencies in Afghanistan,” April 2016, pp. 7, 11.
1260. Australian Department of Foreign Affairs and Trade, “Review of Law and Order Trust Fund for Afghanistan,” 2017, p. 7.
1261. Australian Department of Foreign Affairs and Trade, “Review of Law and Order Trust Fund for Afghanistan,” 2017, p. 7.
1262. Tolo News, “UN Sacks LOTFA Officials over Mismanagement of Funds,” June 11, 2012, accessed June 25, 2020; Australian Department of Foreign Affairs and Trade, “Review of Law and Order Trust Fund for Afghanistan,” 2017, p. 13.
1263. Kwanpadh Suddhi-Dhamakit and Helge Rieper, “Transformative Change of moving M&E from ‘ME’ to ‘MorE’,” National Evaluation Capacities Conference, 2019.
1264. Kwanpadh Suddhi-Dhamakit and Helge Rieper, “Transformative Change of moving M&E from ‘ME’ to ‘MorE’,” National Evaluation Capacities Conference, 2019, p. 1.
1265. UNDP official, correspondence with SIGAR, July 14, 2020.
1266. Kwanpadh Suddhi-Dhamakit and Helge Rieper, “Transformative Change of moving M&E from ‘ME’ to ‘MorE’,” National Evaluation Capacities Conference, 2019.
1267. United Nations World Food Programme, “Third Party Monitoring Guidelines,” June 2014, pp. 5–6; SIGAR, *Emergency Food Assistance to Afghanistan: Incomplete Reporting and Limited Site Visits Hindered USAID’s Oversight of Millions of Dollars of Food Assistance*, SIGAR-20-10-AR, November 2019, p. 2.
1268. SIGAR, *Emergency Food Assistance to Afghanistan: Incomplete Reporting and Limited Site Visits Hindered USAID’s Oversight of Millions of Dollars of Food Assistance*, SIGAR-20-10-AR, November 2019, p. 6.
1269. World Food Programme, “Operation Evaluation: Afghanistan-Protracted Relief and Recovery Operation: 200447, Assistance to Address Food Insecurity and Undernutrition, A mid-term evaluation of WFP’s Operation (2014–2016),” March 2016.
1270. Samuel Hall Consulting, “Food for Training Programme: Implementation Study,” report commissioned by the World Food Programme, 2012, p. 16.
1271. SIGAR, *Emergency Food Assistance to Afghanistan: Incomplete Reporting and Limited Site Visits Hindered USAID’s Oversight of Millions of Dollars of Food Assistance*, SIGAR-20-10-AR, November 2019, p. 6.
1272. Lotte Ruppert, Elias Sagmeister and Julia Steets, “Listening to communities in insecure environments: Lessons from community feedback mechanisms in Afghanistan, Somalia and Syria,” Secure Access in Volatile Environments Project, October 2016; Tearfund, “Monitoring and accountability practices for remotely managed projects implemented in volatile operating environments,” 2012, pp. 8, 41.
1273. Integrity Watch Afghanistan, “Monitoring Handbook: Aide memoire for Community Monitors,” n.d., electronic p. 6; Integrity Watch Afghanistan, “CBM–Infrastructures,” accessed July 10, 2020.
1274. Integrity Watch Afghanistan, “CBM–Infrastructures,” accessed July 10, 2020; Integrity Watch Afghanistan, “CBM–Schools,” July 10, 2020.
1275. The National Solidarity Program, Afghanistan’s largest development program, was intended to deliver services and small infrastructure while building social cohesion. The two main elements were the creation of local community development councils and the provision of small block grants to be administered through those councils. Activities are conducted by NGO facilitating partners. Andrew Beath, Fotini Christia, and Ruben Enikolopov, “Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan,” World Bank Policy Research Working Paper 6129, July 2012, p. 7.
1276. Humanitarian Outcomes, “Secure Access in Volatile Environments: Overview of M&E practices in Afghanistan and discussion of priorities for further research,” SAVE, February 2015, PowerPoint.

1277. According to the World Bank, “a results framework represents the underlying logic that explains how the development objective of a project is to be achieved.” It will typically contain the project’s objective, indicators to measure outcomes linked to the objectives, immediate results to track progress towards outcomes, and an M&E framework that allows measurement of project progress. World Bank, “The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework,” n.d.; German BMZ, “The Monitoring Framework for German Development Cooperation with Afghanistan,” n.d.; Organisation for Economic Cooperation and Development, “Interactive summary charts by aid (ODA) recipients – Afghanistan,” n.d.
1278. German BMZ, “The Monitoring Framework for German Development Cooperation with Afghanistan,” n.d.
1279. German BMZ, “The Monitoring Framework for German Development Cooperation with Afghanistan,” n.d., pp. 4–5; Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 11.
1280. German Institute for Development Evaluation (DEval), “A Review of Evaluative Work of German Development Cooperation in Afghanistan,” 2014, pp. vi–vii.
1281. German Institute for Development Evaluation (DEval), “Evaluations of German Development Cooperation with Afghanistan,” DEval Policy Brief 1, 2015, p. 2.
1282. German Institute for Development Evaluation (DEval), “Evaluations of German Development Cooperation with Afghanistan,” DEval Policy Brief 1 /2015, 2015, p. 2; Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 11.
1283. German Institute for Development Evaluation (DEval), “A Review of Evaluative Work of German Development Cooperation in Afghanistan,” 2014, p. viii.
1284. German Institute for Development Evaluation (DEval), “A Review of Evaluative Work of German Development Cooperation in Afghanistan,” 2014, pp. vii–viii; Jan R. Böhnke, Jan Koehler, and Christoph Zürcher, “Assessing the Impact of Development Cooperation in North East Afghanistan, 2007 – 2013: Final Report,” German Federal Ministry for Economic Cooperation and Development, January 2015, accessed July 2, 2020.
1285. Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 6.
1286. The meta-analysis also drew data from Organisation for Economic Cooperation and Development countries, NGOs, and international organizations such as the Asian Development Bank. Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, pp. 6–7, 9.
1287. Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 7.
1288. Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, p. 8.
1289. Christoph Zürcher, “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018,” March 2020, pp. 8, 16.
1290. There is at least one other notable German attempt to capture impact. Every two years the German Federal Ministry for Economic Cooperation and Development commissions a set of extra questions to be asked in select districts in the northern provinces as part of The Asia Foundation’s annual “Survey of the Afghan People.” The purpose of the additional questions is to create data that “allows plausible assumptions to be drawn on the impact of German assistance.” German Institute for Development Evaluation (DEval), “A Review of Evaluative Work of German Development Cooperation in Afghanistan,” 2014, pp. vii–viii; German BMZ, “The Monitoring Framework for German Development Cooperation with Afghanistan,” n.d., p. 4.
1291. UK Department for International Development staff, SIGAR interview, May 2019; Kirsten Bording Collins, “What is a LogFrame?”, n.d., American University; Organisation for Economic Cooperation and Development, “Interactive summary charts by aid (ODA) recipients – Afghanistan,” n.d.
1292. UK Department of International Development staff, SIGAR interview, May 2019.
1293. Independent Commission for Aid Impact, “About Us,” accessed July 14, 2020.
1294. Independent Commission for Aid Impact, “DFID’s Bilateral Support to Growth and Livelihoods in Afghanistan,” Report Number 31, March 2014, p. 1.
1295. Independent Commission for Aid Impact, “DFID’s Bilateral Support to Growth and Livelihoods in Afghanistan,” Report Number 31, March 2014, p. 1.
1296. Independent Commission for Aid Impact, “DFID’s Bilateral Support to Growth and Livelihoods in Afghanistan,” Report Number 31, March 2014, p. 1.
1297. United Kingdom Stabilisation Unit, “Monitoring and Evaluation of Conflict and Stabilisation Interventions,” October 2014, p. 25.
1298. United Kingdom Stabilisation Unit, “Monitoring and Evaluation of Conflict and Stabilisation Interventions,” October 2014, p. 25.
1299. United Kingdom Stabilisation Unit, “Monitoring and Evaluation of Conflict and Stabilisation Interventions,” October 2014, p. 26.
1300. United Kingdom Stabilisation Unit, “Monitoring and Evaluation of Conflict and Stabilisation Interventions,” October 2014, pp. 7, 25, 27.
1301. United Kingdom Stabilisation Unit, “Monitoring and Evaluation of Conflict and Stabilisation Interventions,” October 2014, p. 28.

1302. German Institute for Development Evaluation (DEval), "A Review of Evaluative Work of German Development Cooperation in Afghanistan," 2014, p. vii.
1303. German Institute for Development Evaluation (DEval), "A Review of Evaluative Work of German Development Cooperation in Afghanistan," 2014, p. vii.
1304. German Institute for Development Evaluation (DEval), "A Review of Evaluative Work of German Development Cooperation in Afghanistan," 2014, p. viii.
1305. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 19, 2018, p. i; World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d., electronic p. 1.
1306. World Bank, "The Afghanistan Reconstruction Trust Fund, Guidance Notes Series: Monitoring and Evaluation Framework," n.d., electronic pp. 1–2.
1307. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 19, 2018, pp. 10–12.
1308. SIGAR, *Afghanistan Reconstruction Trust Fund: The World Bank Needs to Improve How It Monitors Implementation, Shares Information, and Determines the Impact of Donor Contributions*, SIGAR-18-42-AR, April 2018, p. ii.
1309. German BMZ, "BMZ response to the DEval report: 'A Review of Evaluative Work of German Development Cooperation in Afghanistan,'" n.d.
1310. Independent Commission for Aid Impact, "DFID's Bilateral Support to Growth and Livelihoods in Afghanistan," Report Number 31, March 2014.
1311. Australian Department of Foreign Affairs and Trade, "Review of Law and Order Trust Fund for Afghanistan," 2017, pp. 4, 10.
1312. Elias Sagmeister and Julia Steets, "The use of third-party monitoring in insecure contexts: lessons from Afghanistan, Somalia, and Syria," Secure Access in Volatile Environments Project, October 2016, p. 10.
1313. Elias Sagmeister and Julia Steets, "The use of third-party monitoring in insecure contexts: lessons from Afghanistan, Somalia, and Syria," Secure Access in Volatile Environments Project, October 2016, p. 16.
1314. A monitor reporting on the community they live in may be unwilling or unable to report wrongdoing due to potential risks to their safety. Elias Sagmeister and Julia Steets, "The use of third-party monitoring in insecure contexts: lessons from Afghanistan, Somalia, and Syria," Secure Access in Volatile Environments Project, October 2016, pp. 17–18.
1315. Elias Sagmeister and Julia Steets, "The use of third-party monitoring in insecure contexts: lessons from Afghanistan, Somalia, and Syria," Secure Access in Volatile Environments Project, October 2016, p. 19.
1316. In social science research, randomized controlled trials are "evaluation designs in which potential participants are randomly selected to receive a program or become part of a control group that does not receive the program" to determine impact and link it to that program. David Bernstein, "Some Not So Random Thoughts About Random Control Trials," American Evaluation Association Public Library, October 28, 2013.
1317. Andrew Beath, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan," World Bank Policy Research Working Paper 6129, July 2012, p. 3.
1318. Andrew Beath, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan," World Bank Policy Research Working Paper 6129, July 2012, p. 11.
1319. Andrew Beath, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan," World Bank Policy Research Working Paper 6129, July 2012, p. 23.
1320. Andrew Beath, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan," World Bank Policy Research Working Paper 6129, July 2012, p. 22.
1321. Andrew Beath, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan," World Bank Policy Research Working Paper 6129, July 2012, p. 22.
1322. Richard Fontaine and John Nagl, "Contractors in American Conflicts: Adapting to a New Reality," Center for a New American Security, December 2009, p. 12.
1323. See: 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; SIGAR, *Afghanistan's Health Care Sector: USAID's Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017; SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015; SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program's Achievements and the Afghan Government's Ability to Sustain Them*, SIGAR-18-69-AR, September 7, 2018; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency*

- Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 31, 2017; SIGAR, *Afghanistan's Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, May 1, 2019.
1324. USAID, "ADS Chapter 201 Program Cycle Operational Policy," July 23, 2020, p. 128; USAID, "Criteria to Ensure the Quality of the Evaluation Report: A Mandatory Reference for ADS Chapter 201," September 7, 2016, p. 1.
1325. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 58.
1326. For instance, the 2014 annual program review of State's Corrections System Support Program claimed that the program "contributed to stability and confidence in the Afghan government." But the scant evidence the review provided for this claim included such tepid statements as, "CSSP now has a full slate of senior advisors resident in Kabul and working daily with [General Directorate for Prisons and Detention Centers] leadership" and the directorate had "formed an assessment team to provide headquarters staff with audit reports from facilities around the country." State, INL, "2014 Annual Program Review Corrections System Support Program (CSSP)," December 2014, p. 3.
1327. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, p. VI-3.
1328. Jonathon Schroden, "Why Operations Assessments Fail: It's Not Just the Metrics," *Naval War College Review*, vol. 64, no. 4 (Autumn 2011), p. 90.
1329. For examples of reconstruction programs or programmatic areas that did not achieve their intended effects because of unrealistic or otherwise faulty assumptions, as well as questions raised about those assumptions, see Laurel E. Miller, "Afghanistan in Review: Oversight of U.S. Spending in Afghanistan," testimony before the Committee on Homeland Security and Government Affairs Subcommittee on Federal Spending Oversight and Emergency Management, 115th Congress, May 9, 2018, pp. 3–4; Christoph Zürcher, "What Do We (Not) Know About Development Aid and Violence? A Systematic Review," *World Development*, vol. 98 (2017), pp. 506, 509, 511; U.S. Senate Foreign Relations Committee, "Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report Prepared for the Use of the Committee on Foreign Relations, United States Senate," June 2011, pp. 8–9; SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR's Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. 4.
1330. Stabilization operations planner for Regional Command–East, SIGAR interview, July 2016.
1331. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 114.
1332. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 114.
1333. SIGAR, *Counternarcotics: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-52-LL, June 2018, p. 114.
1334. ME&A, "Evaluation: Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation," April 2018, p. 36.
1335. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR's Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. 4.
1336. Justin Sandefur, "Here's the Best Thing the U.S. Has Done in Afghanistan," *Atlantic*, October 10, 2013; Christoph Zürcher, "Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018," pp. 10, 27; Laurel E. Miller, "Afghanistan in Review: Oversight of U.S. Spending in Afghanistan," testimony before the Committee on Homeland Security and Government Affairs Subcommittee on Federal Spending Oversight and Emergency Management, 115th Congress, May 9, 2018, p. 3; USAID, "USAID Afghanistan: Education," February 20, 2019; World Bank, "World Bank Report: Afghanistan Makes Strong, Sustained Health Gains Despite Continuing Insecurity," March 6, 2018.
1337. SIGAR, *Quarterly Report to the United States Congress*, July 30, 2017, p. 182.
1338. As Inspector General Sopko said in January 2020 testimony before the Congress in reference to claims regarding life expectancy, "We talk about some of the statements made by AID about the great success on life expectancy. It was statistically impossible to double the life expectancy in the time given. I think it's a combination of hubris and mendacity that anybody can do that. I mean, the next thing you know is we are going to be walking on water on an aid program." John F. Sopko, testimony before the House Committee on Foreign Affairs, 116th Cong., January 15, 2020, p. 15. Also, SIGAR, *Afghanistan's Health Care Sector: USAID's Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017, p. i; Rod Nordland, Ash Ngu and Fahim Abed, "How the U.S. Government Misleads the Public on Afghanistan," *New York Times*, September 8, 2018; Natalie Carvalho, Julia Hussein, Sue Goldie, and Stewart Britten, "Maternal mortality reported trends in Afghanistan: too good to be true?," British and Irish Agencies Afghanistan Group, December 2015, pp. 1–2; John F. Sopko, "U.S. Lessons Learned in Afghanistan," testimony before the House Committee on Foreign Affairs, January 15, 2020, electronic p. 15.
1339. SIGAR, *Quarterly Report to the United States Congress*, January 30, 2016, p. 119.
1340. Defense researcher, SIGAR interview, February 2019.
1341. Former USAID official, SIGAR interview, January 2019; Matt Millham, "Despite lingering reliance on U.S. aid, Afghanistan's Nawa district a model of 'transition'," *Stars and Stripes*, May 28, 2011; Frances Z. Brown,

- “Afghanistan’s need for reform: We have seen the enemy, and it is our anecdotes,” *Foreign Policy*, October 2012.
1342. Former USAID official, SIGAR interview, January 2019; SIGAR, *SIGAR has Identified up to \$15.5 Billion in Waste, Fraud, Abuse, and Failed Reconstruction Efforts Since its Inception in 2008 Through December 31, 2017*, SIGAR-18-60-SP, p. 3.
1343. USAID, “Plan for Transition 2015–2018,” August, 2016, p. 3.
1344. USAID’s current strategy for Afghanistan states that all of its development objectives “are designed to contribute to the stability, and ultimately, increased self-reliance of Afghanistan because they address the security and development challenges that have made Afghanistan a safe haven for terrorists and violent extremist organizations and have driven the conflict with the Taliban.” USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, p. 10.
1345. For example, in 2011 the GAO concluded that DOD lacked enough oversight personnel to oversee the thousands of contracts being administered in Afghanistan. Similarly, a 2017 State Office of Inspector General report noted that staff at the U.S. embassy in Kabul could not perform all their program oversight duties due to their workload. GAO, “Actions Needed to Address Contract Oversight and Vetting of Non-U.S. Vendors in Afghanistan,” GAO-11-771T, June 30, 2011, p. 8; State Office of Inspector General, “Management Assistance Report: Contract Management- Lessons Learned from Embassy Kabul, Afghanistan, Operations and Maintenance Contract,” Report No. AUD-MERO-17-04, October 2017, p. 5.
1346. SIGAR, *Stabilization: Lessons from the U.S. Afghanistan Experience*, SIGAR-18-48-LL, May 2018, p. 52.
1347. U.S. Senate Committee on Foreign Relations, “Evaluating U.S. Foreign Assistance to Afghanistan: A Majority Staff Report,” June 8, 2011, p. 16.
1348. DOD, CSTC-A, “Train, Advise, & Assist Contract Review,” PowerPoint presentation produced for CSTC-A CMR, April 18, 2019, p. 1.
1349. In September 2017, one permanent INL contracting officer was responsible for monitoring both INLs Justice Sector Support Program and the Corrections System Support Program, cumulatively worth hundreds of millions dollars. SIGAR, interview with State officials, April 2019; State Office of Inspector General, “Audit of the Bureau of International Narcotics and Law Enforcement Affairs Invoice Review Process for Contracts in Afghanistan,” Report No. AUD-MERO-18-30, February 2018, p. 10.
1350. This figure includes only those instances of waste, fraud, and abuse identified by SIGAR and may therefore undercount the amount. Other oversight entities like GAO and inspectors general at DOD, State, and USAID have also identified numerous instances of waste, fraud, and abuse in Afghanistan reconstruction. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, pp. i, 11; SIGAR, *Review Letter: Waste, Fraud, and Abuse Uncovered by SIGAR*, SIGAR-18-60-SP, July 17, 2018, electronic p. 8.
1351. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. 4.
1352. 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.
1353. SIGAR, *Quarterly Report to the United States Congress*, July 30, 2017, p. 182.
1354. SIGAR, *Afghanistan’s Health Care Sector: USAID’s Use of Unreliable Data Presents Challenges in Assessing Program Performance and the Extent of Progress*, SIGAR-17-22-AR, January 19, 2017, p. i.
1355. SIGAR, *Rule of Law in Afghanistan: U.S. Agencies Lack a Strategy and Cannot Fully Determine the Effectiveness of Programs Costing More Than \$1 Billion*, SIGAR-15-68-AR, July 1, 2015, p. ii; SIGAR, *Promoting Gender Equity in National Priority Programs (Promote): USAID Needs to Assess This \$216 Million Program’s Achievements and the Afghan Government’s Ability to Sustain Them*, SIGAR-18-69-AR, September 7, 2018, electronic p. 1, p. 1; SIGAR, *Afghanistan’s Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, May 1, 2019, p. i; SIGAR, *USAID’s Power Transmission Expansion and Connectivity Project: The Project is Behind Schedule, and Questions Remain about the Afghan Government’s Ability to Use and Maintain the New Power Infrastructure*, SIGAR-19-57-AR, September 4, 2019, electronic p. 3.
1356. For one of SIGAR’s more direct critiques, see: SIGAR, correspondence with Secretary of State John F. Kerry, Secretary of Defense Chuck Hagel, and USAID Administration Rajiv Shah, July 5, 2013.
1357. SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 31.
1358. In reference to the request, Inspector General Sopko stated in April 2019 testimony before the Congress, “And you remember, Mr. Chairman, I got into an actual, I wouldn’t say fight, a serious discussion with a senior official from USAID about racking and stacking programs in USAID. And he turned on me saying, the audacity of Mr. Sopko to want me to list which programs are working and which ones are failing. It’s almost like picking your favorite child. I just had to stop for a second. I mean I never thought development aid was sort of the Sophie’s choice that we had to pick. I mean, doesn’t it make sense as a businessman?” SIGAR, “Seven years after initial request, the Departments of Defense and State and the U.S. Agency for International Development have yet to list their 10 most and least successful reconstruction projects asked for by Inspector General John F. Sopko,” January 22, 2020, accessed December 8, 2020; John F. Sopko, “SIGAR’s

- 2019 High-Risk List,” testimony before the House Oversight and Reform Subcommittee on National Security Hearing, 116th Congress, April 3, 2019, electronic p. 21; SIGAR, correspondence with State, USAID, and DOD, March 25, 2013, pp. 1–3.
1359. Mercy Corps, “Does Youth Employment Build Stability? Evidence from an Impact Evaluation of Vocational Training in Afghanistan,” January 2015, p. 3.
1360. James H. Stock and Mark W. Watson, *Introduction to Econometrics*, updated 3rd Edition, (Essex, UK: Pearson Education Limited, 2015), p. 823.
1361. James H. Stock and Mark W. Watson, *Introduction to Econometrics*, updated 3rd Edition, (Essex, UK: Pearson Education Limited, 2015), p. 52.
1362. James H. Stock and Mark W. Watson, *Introduction to Econometrics*, updated 3rd Edition, (Essex, UK: Pearson Education Limited, 2015), p. 131.
1363. James H. Stock and Mark W. Watson, *Introduction to Econometrics*, updated 3rd Edition, (Essex, UK: Pearson Education Limited, 2015), pp. 521, 823.
1364. James H. Stock and Mark W. Watson, *Introduction to Econometrics*, updated 3rd Edition, (Essex, UK: Pearson Education Limited, 2015), p. 53.
1365. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 229.
1366. One of the most prominent arguments is that of former USAID Administrator Andrew Natsios, whose views this report refers to several times. Natsios argues that an imbalance has developed between the compliance and technical sides of aid programs, threatening program integrity. In his view, a major contributor to the problem is a decline in the number of career USAID staff even as spending increased dramatically since September 11, 2001. Andrew Natsios, “The Clash of the Counter-bureaucracy and Development,” Center for Global Development, July 2010.
1367. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, pp. 12, 14–15.
1368. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, pp. 1, 5, 7.
1369. SIGAR, *USAID Spent Almost \$400 Million on an Afghan Stabilization Project despite Uncertain Results, but Has Taken Steps to Better Assess Similar Efforts*, SIGAR Audit-12-8, April 2012, p. 7.
1370. LGCD was a stabilization project. In 2018 SIGAR concluded that all U.S. stabilization efforts appeared to have failed. SIGAR, *Update on the Amount of Waste, Fraud, and Abuse Uncovered through SIGAR’s Oversight Work between January 1, 2018 and December 31, 2019*, SIGAR-21-05-SP, October 2020, p. 4.
1371. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” July 23, 2020, p. 18.
1372. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 13.
1373. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 13.
1374. Kartika Bhatia and Hafez Ghanem, “How do Education and Unemployment Affect Support for Violent Extremism? Evidence from Eight Arab Countries,” Global Economy and Development Working Paper 102, Brookings Institution, March 22, 2017, p. 1.
1375. USAID, “Country Development Cooperation Strategy FY 2019–2023,” September, 2018, p. 26.
1376. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, p. 47.
1377. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, p. 47.
1378. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, pp. 48–49.
1379. Checchi and Company Consulting, Inc., “USAID/Afghanistan Plan for Transition Strategy (2015–2018) Mid-Course Stocktaking Exercise,” August 2017, p. 1.
1380. Checchi and Company Consulting, Inc., “USAID/Afghanistan Plan for Transition Strategy (2015–2018) Mid-Course Stocktaking Exercise,” August 2017, p. 16.
1381. Checchi and Company Consulting, Inc., “USAID/Afghanistan Plan for Transition Strategy (2015–2018) Mid-Course Stocktaking Exercise,” August 2017, p. 16.
1382. Checchi and Company Consulting, Inc., “USAID/Afghanistan Plan for Transition Strategy (2015–2018) Mid-Course Stocktaking Exercise,” August 2017, p. 16.
1383. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, p. 38.
1384. Herbie Smith, “Staying in Afghanistan is not worth more U.S. lives,” *Delaware Online*, April 25, 2019.
1385. USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, p. 128; USAID, “Criteria to Ensure the Quality of the Evaluation Report: A Mandatory Reference for ADS Chapter 201,” September 7, 2016, p. 1.
1386. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities,” October 26, 2017, p. i.
1387. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities,” October 26, 2017, p. ii.
1388. Checchi and Company Consulting, Inc., “Mid-Term Performance Evaluation: Power Transmission Expansion and Connectivity (PTEC) Commercialization Activities,” October 26, 2017, p. iv.
1389. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 3.
1390. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 58.
1391. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 5.
1392. State, “2015 Annual Program Review Corrections System Support Program (CSSP),” June 2015, p. 3.

1393. SIGAR analysis of State, “2015 Annual Program Review Corrections System Support Program (CSSP),” June 2015, p. 3.
1394. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 6.
1395. A related objective was to eventually transition the training provided under the program to the Afghan government. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 63.
1396. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, p. 23.
1397. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, pp. 23–24.
1398. The evaluators noted that the program had “failed to even attempt to measure the results of [its] efforts [on the justice sector as a whole] in any comprehensive manner.” In the absence of such measurement, it was “very difficult to identify the direct and indirect, medium and long-term results of such projects, nor to analyze how they can positively contribute to the goal of strengthening the Afghan justice system and restoring public confidence in it.” The evaluators were careful not to characterize their work as an impact evaluation. Nevertheless, they explored the potential impact of donor work in the Afghan justice sector at several points in the evaluation. Conor Foley and Katerina Stolyarenko, “External Mid-Term Evaluation of Afghanistan Justice Training Transition Program (JTTP) (2013–2014),” November 2014, pp. 8, 17, 21, 44, 46–47.
1399. For example, a final evaluation of State’s Corrections System Support Program deflected in response to a key question regarding the program’s effectiveness: “Have CSSP interventions improved conditions for and treatment of inmates as well as the fairness and responsiveness of the overall justice system?” Evaluators answered this question by stating that “a total of 1,557 hours of advising were conducted between 2012 and 2014 on the topics of human rights and humane treatment,” that “nearly all . . . correctional officers and facility managers perceive that inmates are treated humanely,” and that “only 61 percent . . . of inmates agree that inmates are treated humanely.” Based on this data, a more accurate response to the evaluation question would have been “we don’t know.” Barbaricum, “Final Evaluation Report INL Corrections System Support Program in Afghanistan,” July 23, 2015, pp. 3–4; SIGAR analysis.
1400. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 3.
1401. FAR 37.602; FAR Subpart 37.601(b)(2), “Performance-Based Acquisition: General,” and Subpart 37.603(a), “Performance-Based Acquisition: Performance standards.”
1402. DOD, OSD-P, correspondence with SIGAR, October 24, 2018, electronic p. 2.
1403. SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 6.
1404. In response to a draft of this report, DOD stated, “It’s absurd to think contracted advisors should be held accountable for the capacity building outcomes that were the responsibility of RS and ISAF before it. The contractors just did what they were instructed to do by their government supervisor; they did not own the capacity building effort.” We do not suggest that contracted advisors should be held *solely* accountable. Rather, the recommendation seeks to address the problem of spending substantial amounts of money on contracts whose effects are indeterminate. With respect to DOD’s DynCorp contracts, SIGAR previously recommended that DOD “incorporate specific, measurable performance standards into its current and future ministerial advising contracts.” With respect to separate contracts intended to build the intelligence capability of the ministries of defense and interior, SIGAR recommended that DOD “review ongoing ANDSF intelligence training and mentoring contracts, and incorporate into them requirements to enable the measurement and verification of contractor performance and contract outcomes, including training and mentoring results, and impacts on the ANDSF’s human intelligence capability.” DOD concurred with both recommendations. DOD, OSD-P, correspondence with SIGAR, May 11, 2021; SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. ii; DOD, OSD-P, correspondence with SIGAR, October 24, 2018, electronic p. 2; SIGAR, *Afghanistan National Defense and Security Forces: DOD Spent \$457.7 Million on Intelligence Capacity-Building Programs, but Impact Cannot Be Fully Assessed Because of a Lack of Performance Metrics*, SIGAR-17-57-AR, July 2017, p. iii; DOD, OSD-P, correspondence with SIGAR, July 28, 2017, electronic p. 3.
1405. SIGAR, *DOD Task Force for Business and Stability Operations: \$675 Million in Spending Led to Mixed Results, Waste, and Unsustained Projects*, SIGAR-18-19-AR, January 2018, p. ii.
1406. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1407. AIF was part of a joint program. AIF projects were to be “implemented by the Secretary of State, in coordination with the Secretary of Defense, unless they agree that the project should be implemented by the Secretary of Defense,” in the words of the 2011 National Defense Authorization Act. However, all AIF funds were appropriated to DOD (25 percent of those funds were subsequently transferred to USAID). As SIGAR stated in its 2017 audit of FY 2011 AIF projects, “In practice, DOD implemented all but one of the FY 2011

- AIF projects.” Moreover, both USAID and State currently have relatively robust M&E systems that can be used to track the effects of development projects (that is, projects that fall outside the scope of DOD’s core mission). For these reasons, we direct this recommendation to DOD only. DOD, OSD-P, correspondence with SIGAR, May 11, 2021; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, p. 4.
1408. Department of Defense and Full-Year Continuing Appropriations Act, 2011, Pub. L. No. 112-10, 125 Stat. 38 (2011); U.S. Government Interagency Counterinsurgency Initiative, “U.S. Government Counterinsurgency Guide,” January 2009; SIGAR, *Afghanistan Infrastructure Fund: Agencies Have Not Assessed Whether Six Projects That Began in Fiscal Year 2011, Worth about \$400 Million, Achieved Counterinsurgency Objectives and Can Be Sustained*, SIGAR-18-10-AR, October 2017, pp. i, 1.
1409. SIGAR, *USACE’s Local National Quality Assurance Program: USACE Used Qualified Personnel to Monitor Construction in Afghanistan and Is Taking Steps to Improve Contractor Reporting*, SIGAR-19-60-AR, p. i; USFOR-A, JENG, “Afghanistan Infrastructure Fund Program FY 11 through FY 14,” September 20, 2019; SIGAR, *Afghanistan’s Energy Sector: USAID and DOD Did Not Consistently Collect and Report Performance Data on Projects Related to Kajaki Dam, and Concerns Exist Regarding Sustainability*, SIGAR-19-37-AR, i–ii; USAID, Office of Infrastructure, response to SIGAR quarterly report vetting draft, April 7, 2021; Management Systems International, “Electricity Access and Socioeconomic Baseline Survey for Energy Infrastructure Activities,” May 2020, pp. iii, x–xi, 2, 28, 50–51, 128, 130, 132–133; SIGAR analysis.
1410. In its 2012 audit of FY 2011 AIF projects, SIGAR raised the question of whether projects delays could create an expectations gap among affected Afghans. SIGAR, *Fiscal Year 2011 Afghanistan Infrastructure Fund Projects Are behind Schedule and Lack Adequate Sustainment Plans*, SIGAR Audit-12-12, July 2012, electronic p. 1.
1411. DOD, Joint Chiefs of Staff, “Joint Publication 5-0: Joint Planning,” June 16, 2017, p. VI-3; CNA, “Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs,” unpublished draft for SIGAR LLP, p. 44.
1412. Illustratively, then-ISAF Commander General Stanley McChrystal’s initial assessment of the situation in Afghanistan is replete with direct and indirect references to contracted efforts that were believed to be necessary for success. For example, key components of the strategy suggested by McChrystal were “[developing] a significantly more effective and larger ANSF” and “[prioritizing] responsive and accountable governance.” To help develop Afghan security forces, the U.S. Army awarded contracts totaling at least \$1.62 billion to DynCorp International to advise Afghanistan’s ministries of defense and interior. USAID, meanwhile, has spent more than \$3.2 billion on governance projects since 2002. ISAF, “Commander’s Initial Assessment,” August 30, 2009, p. 2-2; SIGAR, *Afghanistan National Defense and Security Forces: DOD Lacks Performance Data to Assess, Monitor, and Evaluate Advisors Assigned to the Ministries of Defense and Interior*, SIGAR-19-03-AR, October 2018, p. 3; SIGAR, *Quarterly Report to the United States Congress*, April 30, 2021, p. 131.
1413. Rory Stewart and Gerald Knaus, *Can Intervention Work*, (New York, NY: W.W. Norton & Company, 2011), p. 50.
1414. Stephen Downes-Martin, “Operations Assessment in Afghanistan Is Broken—What Is to Be Done?,” *Naval War College Review*, Autumn 2011, vol. 64, no. 4, p. 108.
1415. DOD, Joint Chiefs of Staff, “Joint Publication 5-0: Joint Planning,” December 2020, p. J-1.
1416. Commenting on this recommendation in response to a draft of this report, DOD stated, “It should not be up to the operational command to produce the primary assessment of its progress. It should be something that is assessed at the headquarters (Pentagon) level by an organization that has oversight responsibility such as Policy, but for which Policy is not staffed to do because there is no requirement either from Congress or the senior leadership to go down this road.” Although there may be value in centralizing campaign assessments within an entity like the Office of the Under Secretary of Defense for Policy, the evidence base developed for this report did not support making such a recommendation. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1417. Commenting on this recommendation in response to a draft of this report, and in reference to its comment on Recommendation 12, DOD suggested that Recommendation 14 should include additional language that would encourage the Congress to provide funds for a program of impact evaluation carried out by federally funded research and development centers. There is nothing in the wording of this recommendation that would prevent DOD from commissioning federally funded research and development centers to carry out impact evaluations. However, as stated in the previous endnote, although there may be value in centralizing campaign assessment within an entity like the Office of the Under Secretary of Defense for Policy, the evidence base developed for this report did not support making such a recommendation. DOD, OSD-P, correspondence with SIGAR, May 11, 2021.
1418. The power of the U.S. government to enter into contracts is grounded in Article II of the Constitution, which states that the President “shall take care that the laws be faithfully executed,” and grants the President the power of appointment. The President is not only constitutionally authorized, but also constitutionally bound, in the words of an 1831 Supreme Court decision, to “avail himself of every appropriate means not forbidden by law” to ensure that “the laws are faithfully executed.” As contracts were not forbidden by law, the Supreme Court decided that the President could enter into contracts. This constitutionally derived au-

- thority extended to presidentially appointed government officers, who, as agents of the President, were too “enjoined ‘to take care that the laws be faithfully executed’” by any means that were not illegal. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. i; *United States v. Tingey*, 30 U.S. (5 Pet.) 115 (1831); National Archives, “The Constitution of the United States: A Transcription,” accessed October 14, 2020.
1419. National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, § 1211 (2018).
 1420. Consolidated Appropriations Act for Fiscal Year 2017, Pub. L. No. 115-31, § 7044 (2016).
 1421. Foreign Aid Transparency and Accountability Act of 2016, Pub. L. No. 114-191, § 3 (2016).
 1422. Office of Management and Budget, “Monitoring and Evaluation Guidelines for Federal Departments and Agencies that Administer United States Foreign Assistance,” January 11, 2018.
 1423. As various parts of this report show, there is a sense in which debates about efficacy are sometimes caught between policy and best practice. As policy typically dominates, M&E is sometimes relegated to examinations of efficacy that are arbitrarily confined. For example, this particular presidential policy directive “recognizes that development is vital to national security.” The question becomes: what is M&E permitted to say about efficacy in cases where such statements are not grounded in evidence, even as they permeate the logic of interventions? In other words, whether M&E reaches its full potential hinges on the degree of freedom it is granted. White House, “Fact Sheet: U.S. Global Development Policy,” accessed July 30, 2020.
 1424. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 10.
 1425. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 10; GAO, “Recommendations of the Commission on Government Procurement: A Final Assessment,” May 31, 1979, accessed August 27, 2020.
 1426. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. i.
 1427. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 8.
 1428. Federal Acquisition Regulation, Part 37: “Service Contracting,” Subpart 37.6: “Performance-based Acquisition,” §37.602: “Performance work statement,” §37.603: “Performance standards.”
 1429. Federal Acquisition Regulation, Part 37: “Service Contracting,” Subpart 37.6: “Performance-based Acquisition,” §37.602: “Performance work statement,” §37.603: “Performance standards”; ASI Government, “A COR’s Guide to Statements of Work, Performance Work Statements, and Statements of Objectives,” January 9, 2011, p. 1.
 1430. This observation is sometimes misinterpreted. To be clear, it is often useful, if not necessary to track outputs. But where feasible, it is equally, if not more useful and necessary to track outcomes and impact.
 1431. CRS, “The Federal Acquisition Regulation (FAR): Answers to Frequently Asked Questions,” R42826, February 3, 2015, p. 9.
 1432. 48 CFR §1.102: “Statement of guiding principles for the Federal Acquisition System,” July 3, 1995, p. 6.
 1433. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 103.
 1434. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 1, 5.
 1435. Per ADS Chapter 201, there are technically four stages. For the purposes of simplification, we have grouped “Project Design and Implementation” and “Activity Design and Implementation” into a single stage. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 13.
 1436. A Country Development Cooperation Strategy supports State’s Integrated Country Strategy, which describes overall U.S.-government objectives for a given country. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 41, 143.
 1437. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 26.
 1438. The results framework, which is subject to constant adjustment during the course of any mission, is guided by a set of development hypotheses that “explain the logic and causal relationships between the building blocks needed to achieve a long-term result.” These hypotheses are intended to be “based on theory, practice, literature, and/or experience,” and help link together each level of the results framework, using the mission’s best understanding of the given problems. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 27–28.
 1439. This series of outcomes reflects specific components of the results framework for USAID’s current Country Development Cooperation Strategy for Afghanistan. For the purposes of concision, the text to which this endnote refers summarizes a broader causal chain involving other intermediate outcomes like increasing transportation and energy service delivery and increasing the skills of the Afghan workforce. For readability, the description of USAID’s results framework provided in the text was also simplified. For example, USAID has specific technical terms it applies to each level of outcome specified in a mission-level results framework. Sub-Intermediate results (referred to as “Sub IRs”) are the lowest-level outcomes specified in the results framework. Together, a set of Sub IRs aims to produce a higher-level Intermediate Result (IR). Related IRs are intended to produce a Development Objective (DO). DOs, in turn, contribute to an ultimate strategic goal. USAID, “Country Development Cooperation Strategy for Afghanistan FY 2019–2023,” September 2018, p. 27.
 1440. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 46.
 1441. As ADS Chapter 201 states, “A PMP is a Mission-wide tool for planning and managing the processes of monitoring strategic progress, project performance, programmatic assumptions, and operational context;

- evaluating performance and impact; and learning and adapting from evidence.” USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 46.
1442. Limitations of the PMP are discussed in Chapter 3.
1443. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 53.
1444. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 53.
1445. Per the ADS, activities are generally implemented via contracts, grants, “or agreement[s] with another U.S. Government agency or with the partner country government.” While we simplify the definition here to refer only to “cooperative agreements,” USAID may use other kinds of agreements depending on the nature of an activity. For example, USAID may use a Participating Agency Program Agreement (PAPA) when the technical expertise of another federal agency is required. In such cases, the partnering federal agency assumes responsibility for contracting whatever services may be necessary to implement the activity. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 54, 140, 156; USAID, “ADS Glossary of Terms,” April 18, 2018, p. 187.
1446. As implementation in Afghanistan often involves the continuation of ongoing, multi-year projects, a project appraisal document (PAD) is sometimes carried over from one strategy to another. For example, the Afghanistan Country Development Cooperation Strategy for fiscal years 2019 to 2023 indicated that USAID planned to continue a large number of projects, including supporting essential hospital services across approximately 2,400 health facilities nationwide, supporting the Afghanistan Reconstruction Trust Fund’s Recurrent Cost Window Incentive Program (which rewards the attainment of development benchmarks), and focusing on climate-smart agriculture interventions such as expanding access to irrigation. Part of the reason for continuing projects and activities is that they are often implemented over three- to five-year periods. Carrying a PAD over from one strategy to another implies an ongoing belief in a project’s theory of change. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 61; USAID, “Country Development Cooperation Strategy FY 2019–2023,” September 2018, pp. 25, 33, 38.
1447. The theory of change is depicted visually in an associated logic model, a graphic that “organizes and depicts the team’s thinking on the logical relationships between what the project will do and the changes it expects to see.” This core step of the project appraisal document already involves an opportunity for learning, since missions are supposed to review existing evaluations and analyses before constructing the theory of change. USAID expects that the theory of change will evolve during implementation. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 62, 67.
1448. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 62, 67, 72.
1449. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 78, 103.
1450. Activities in particular are meant to achieve “clear and measurable results.” USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 82, 104.
1451. FHI 360, “Initiative for Hygiene, Sanitation and Nutrition (IHSAN) Activity Monitoring Evaluation & Learning Plan (AMELP) FY 2020,” November 13, 2019, pp. 4, 9–11.
1452. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 13, 111.
1453. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 112.
1454. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 112–113.
1455. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 112.
1456. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 114.
1457. There is recognition in USAID’s policy that data must be of good quality for performance indicators to be meaningful. Consequently, ADS Chapter 201 emphasizes that “high-quality data are the cornerstone for evidence-based decision making.” USAID lists five standards of data quality, and requires missions to conduct a data quality assessment for each performance indicator to make sure that the indicator data meets those five data quality standards, and to determine the extent to which that data should be allowed to influence management decisions. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 114, 119–120.
1458. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 114.
1459. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 114.
1460. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 114.
1461. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 121.
1462. USAID policy defines three instances where evaluations are required. First, missions must conduct at least one evaluation per project, which may address any part of the project, from one particular activity to the project as a whole. Second, missions must conduct an impact evaluation of new, untested approaches that are expected to expand in scale or scope. This includes integrating impact evaluations into new interventions. If impact evaluations are not feasible, missions must explain why, and conduct a performance evaluation instead. Third, missions must conduct at least one “whole-of-project” performance evaluation during their Country Development Cooperation Strategy time frame that examines an entire project, including all of its activities. Evaluations must be carried out by independent evaluators who have no fiduciary relationship with the implementing partner, in order to avoid conflicts of interest. Although missions are allowed to carry out internal evaluations led by USAID staff or by an implementing partner, such evaluations do not count toward the requirements described here. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 121, 123–125.
1463. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 121–122.
1464. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 123.

1465. USAID, “Glossary of ADS Terms,” April 18, 2018, p. 126.
1466. USAID, “Glossary of ADS Terms,” April 18, 2018, p. 191.
1467. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 123.
1468. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 123.
1469. SIGAR conclusion based on: USAID, *ADS Chapter 201: Program Cycle Operational Policy*, partial revision dated June 11, 2019, pp. 51–52.
1470. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 51.
1471. The portfolio review during the final year of the Country Development Cooperation Strategy “must include a review of the cumulative achievements toward the development objectives and intermediate results.” USAID, *ADS Chapter 201: Program Cycle Operational Policy*, partial revision dated June 11, 2019, pp. 51–52.
1472. For example, portfolio reviews are listed as a key source of learning in USAID’s initial description of the program cycle. Missions must consolidate lessons and findings from past portfolio reviews when developing a follow-up CDCS, and regularly reflecting on portfolio reviews is recommended as a potential approach for adaptive management. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, pp. 15, 37, 52, 106.
1473. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 52.
1474. Other functions include identifying appropriate amendments to the results framework, re-engaging stakeholders, reinforcing continuity and institutional knowledge among new staff, and identifying future analytic needs. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 52.
1475. USAID, “ADS Chapter 201: Program Cycle Operational Policy,” June 11, 2019, p. 52.
1476. USAID, “How-To Note: Country Development Cooperation Strategy Mid-Course Stocktaking,” November 2017, p. 1.
1477. State, Office of Foreign Assistance, “Establishing, Monitoring, and Refining an Evaluation Policy at the Department of State,” March 29, 2018.
1478. CNA, “Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs,” unpublished draft for SIGAR LLP, p. 32.
1479. CNA, “Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs,” unpublished draft for SIGAR LLP, p. 32.
1480. State, “Diplomacy in Action: Managing for Results: Planning, Budgeting, Managing, and Measuring,” December 16, 2013.
1481. State, 18 Foreign Affairs Manual 301.1, Managing for Results Framework; State, “Diplomacy in Action: Managing for Results: Planning, Budgeting, Managing, and Measuring,” December 16, 2013.
1482. State, “Bureau of Budget and Planning – Our Mission,” n.d., accessed August 31, 2020.
1483. State, “Office of Foreign Assistance – Our Mission,” n.d., accessed August 31, 2020.
1484. CNA, “Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs,” unpublished draft for SIGAR LLP, p. 18.
1485. State, “Program Design and Performance Management Toolkit,” n.d., p. 2.
1486. State, “Diplomacy in Action: Department of State Evaluation Policy,” January 29, 2015.
1487. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, p. 1.
1488. The Managing for Results framework was ultimately codified in the 2015 18 Foreign Assistance Manual 301.1-1. Together with associated foreign affairs handbooks, the FAM governs the operations of State, the Foreign Service, and other federal agencies (in situations where they fall under State’s authority). CNA, “Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs,” unpublished draft for SIGAR LLP, p. 18; State, 18 Foreign Affairs Manual 301.1, Managing for Results Framework; State, “Foreign Affairs Manual and Handbook,” n.d., accessed August 31, 2020.
1489. State, 18 Foreign Affairs Manual 301.1, Managing for Results Framework.
1490. The latter is separately articulated in 18 FAM 301.4. State, 18 Foreign Affairs Manual 301.4, Program and Project Design, Monitoring and Evaluation; State, 18 Foreign Affairs Manual 301.1, Managing for Results Framework.
1491. State, “Department of State Program and Project Design, Monitoring, and Evaluation Policy,” November 2017, pp. 3–4.
1492. State, response to SIGAR data call, May 3, 2019; State, “Program Design and Performance Management Toolkit,” March 15, 2017, electronic p. 70.
1493. CNA, “Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs,” unpublished draft for SIGAR LLP, p. 32.
1494. State, “Program Design and Performance Management Toolkit,” March 15, 2017, electronic pp. 40, 69–70; USAID, “ADS Chapter 201 Program Cycle Operational Policy,” July 23, 2020, pp. 71, 77.
1495. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 3.
1496. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 8.
1497. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 8.
1498. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 8.
1499. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 29.
1500. State, “Program Design and Performance Management Toolkit,” March 15, 2017, pp. 65–67.
1501. State, “Program Design and Performance Management Toolkit,” March 15, 2017, p. 19.

1502. State, "Program Design and Performance Management Toolkit," March 15, 2017, pp. 19–20.
1503. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 20.
1504. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 20.
1505. State, "Program Design and Performance Management Toolkit," March 15, 2017, pp. 20, 29.
1506. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 37.
1507. State, 18 Foreign Affairs Manual 301.4, Program and Project Design, Monitoring and Evaluation, p. 2.
1508. State, Office of Foreign Assistance, "Design, Monitoring, and Evaluation Policy at State and Demystifying the PPR and Standard Indicators," September 12, 2019, p. 4.
1509. State, Office of Foreign Assistance, "Design, Monitoring, and Evaluation Policy at State and Demystifying the PPR and Standard Indicators," revised September 12, 2019, p. 10.
1510. State, Office of Foreign Assistance, "Design, Monitoring, and Evaluation Policy at State and Demystifying the PPR and Standard Indicators," September 12, 2019, p. 10; State, "Standard Foreign Assistance Indicators," n.d., accessed September 1, 2020; evaluation expert, SIGAR interview, August 2018; implementing partner staff member, SIGAR interview, May 2019.
1511. State, "Standard Foreign Assistance Indicators," n.d., accessed September 1, 2020.
1512. State, "Resources and Reports," n.d., accessed September 1, 2020.
1513. For example, see footnote 3 of State, "FY 2021 Annual Performance Plan and FY 2019 Annual Performance Report," July 28, 2020, p. 36; State, "Performance Plans and Reports," n.d., accessed September 1, 2020.
1514. At times, State has used annual performance reports to justify requests for funding from Congress. State, "Performance Plans and Reports," n.d., accessed September 1, 2020.
1515. Richard Blue, Cynthia Clapp-Wincek, and Holly Benner, "Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change," May 2009, p. 13.
1516. For example, USAID funded the five-year, \$77 million Increasing Access to Basic Education and Gender Equality – Community-Based Education (CBE) project in Afghanistan. According to USAID, "CBE is a proven means to reach children who are unable to attend formal schools due to insecurity, distance or other constraints." USAID, "Increasing Access to Basic Education and Gender Equality – Community-Based Education (CBE)," August 2016, p. 1.
1517. See the performance indicator matrix in USAID, "Afghan Children Read Annual Report Year 4: Fiscal Year 2019 October 2018–September 2019," October 31, 2019, pp. 68–73; USAID, "Afghan Children Read (ACR)," December 2018, p. 1.
1518. USAID, "Standard Indicators," June 12, 2019, accessed September 1, 2020.
1519. USAID, "Standard Indicators," June 12, 2019, accessed September 1, 2020.
1520. State, "Program Design and Performance Management Toolkit," March 15, 2017, p. 43.
1521. Defense researcher, SIGAR interview, February 2019; DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 1.
1522. For example, DOD uses operation assessments to determine the progress of military operations and campaigns in order to improve their efficacy. Separately, DOD has doctrine governing operational contract support, which describes how to integrate contracts into military operations and how to perform oversight of those contracts. M&E-like processes at DOD extend beyond reconstruction to warfighting. DOD, "Joint Planning," Joint Publication 5-0, June 16, 2017, p. VI-1; defense researcher, SIGAR interview, February 2019; DOD official, SIGAR interview, October 2019; DOD officials in the Office of the Under Secretary of Defense–Policy and Defense Pricing and Contracting, SIGAR interview, March 2019; DOD, "Operational Contract Support," Joint Publication 4-10, March 4, 2019, pp. ix, I-5, III-17.
1523. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 95; CSTC-A Resource Management, SIGAR interview, May 2019; DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. I-5.
1524. Marla C. Haims, Melina Moore, Harold D. Green, Jr. and Cynthia Clapp-Wincek, "Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects," RAND Corporation, 2011, p. 1.
1525. Marla C. Haims, Melina Moore, Harold D. Green, Jr. and Cynthia Clapp-Wincek, "Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects," RAND Corporation, 2011, electronic pp. 3, 17.
1526. Defense researcher, SIGAR interview, February 2019.
1527. National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328, § 383 (2016).
1528. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 1.
1529. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 12.
1530. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, pp. 3, 21.
1531. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 15.

1532. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 3.
1533. Defense researcher, SIGAR interview, February 2019.
1534. National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, § 1211 (2018).
1535. National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, § 1211 (2018); National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328, § 383 (2016).
1536. Defense researcher, SIGAR, interview, February 2019.
1537. DOD, ODASD(SC), correspondence with SIGAR, May 13, 2021.
1538. DOD, "Joint Planning," Joint Publication 5-0, June 16, 2017, p. xx.
1539. For example, regional commands in Afghanistan produced assessments not only for security conditions in their areas, but also for governance and development. Separate assessments of each of these lines of effort were then combined to produce an overall assessment. Stephen Downes-Martin, "Operations Assessment in Afghanistan is Broken—What Is to Be Done?," *Naval War College Review*, 2011, p. 112.
1540. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, pp. I-13–I-14.
1541. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, p. I-1.
1542. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, p. I-13.
1543. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, p. xxvi.
1544. DOD, "Joint Publication 5-0: Joint Planning," June 16, 2017, pp. I-13–I-14.
1545. For examples see: 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; CNA, "Independent Assessment of the Afghan National Security Forces," January 24, 2014.
1546. SIGAR conclusion based on SIGAR interview with defense researcher, February 2019; CNA, "Understanding Monitoring and Evaluation: Lessons from Afghanistan Reconstruction Programs," unpublished draft for SIGAR LLP, p. 6; CNA, "Independent Assessment of the Afghan National Security Forces," January 24, 2014, p. 1.
1547. For example, CNA's 2014 assessment made predictions about which geographic areas the Taliban were likely to challenge for control and the mid-term likelihood of a negotiated political settlement. CNA, "Independent Assessment of the Afghan National Security Forces," January 24, 2014, pp. 1, 21–22; CNA, "Independent Assessment of the Afghan National Defense and Security Forces," November 2015, p. i.
1548. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 21.
1549. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 21.
1550. ForeignAssistance.gov, "Summary of Department of Defense Foreign Assistance Portfolio."
1551. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. I-5.
1552. U.S. Army Corps of Engineers, SIGAR interview, May 2019.
1553. DOD, "Defense Contingency COR Handbook version 2," September 2012, pp. 193, 328.
1554. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. III-17.
1555. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. III-17.
1556. SIGAR, *Stabilization: Lessons from the U.S. Experience in Afghanistan*, SIGAR-18-48-LL, May 2018, p. 95; DOD, "DOD Dictionary of Military and Associated Terms," January 2020, p. 139.
1557. CSTC-A, SIGAR interview, May 2019.
1558. For example, according to CSTC-A officials, CSTC-A's 2019 contract management review was prompted by the arrival of then-Lt. Gen. James Rainey, who was interested in looking at the portfolio of contracts managed by CSTC-A and understanding whether there were opportunities to better control costs. CSTC-A, SIGAR interview, May 2019.
1559. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 3.
1560. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 21.
1561. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 21.
1562. DOD, "Instruction 5132.14, Assessment, Monitoring and Evaluation Policy for the Security Cooperation Enterprise," January 13, 2017, p. 21.
1563. GAO, "Foreign Assistance: Federal Monitoring and Evaluation Guidelines Incorporate Most but Not All Leading Practices," GAO-19-466, July 2019, p. 42.
1564. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, pp. III-17–III-19.
1565. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, pp. III-17–III-19.
1566. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. I-12.
1567. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. I-12.
1568. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, p. III-24.
1569. DOD, "Joint Publication 4-10: Operational Contract Support," March 4, 2019, pp. III-24–III-25.





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.

The background of the entire page is a photograph of a construction site at sunset. The sky is a vibrant orange and yellow, with silhouettes of construction workers and scaffolding visible against the bright light. The silhouettes are dark and detailed, showing the complex structure of the building under construction.

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