



USAID
FROM THE AMERICAN PEOPLE

AN INVENTORY AND REVIEW OF COUNTERING VIOLENT EXTREMISM AND INSURGENCY MONITORING SYSTEMS

DECEMBER 10, 2012

This publication was produced for review by the United States Agency for International Development. It was prepared by Lynn Carter and Phyllis Dininio, with support from Jennifer Ulman and Lindsay North, Management Systems International.

AN INVENTORY AND REVIEW OF COUNTERING VIOLENT EXTREMISM AND INSURGENCY MONITORING SYSTEMS



600 Water Street, SW, Washington, DC 20024, USA
Tel: +1.202.484.7170 | Fax: +1.202.488.0754
www.msiworldwide.com

Contracted under AID-OAA-TO-11-00051
Democracy and Governance and Peace and Security in the Asia and Middle East

DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

- Acronyms..... **i**
- Executive Summary..... **iv**
- I. Introduction..... 1**
 - Scope of Work 1
 - Approach..... 3
 - Outcome..... 3
 - Organization of the Report..... 4
- II. Program Summaries..... 5**
 - CVE Systems 5
 - COIN Systems 17
- III. Comparison of Programs 40**
 - Theories of Change..... 40
 - Indicator Selection and Organization..... 45
 - Introduction 46
 - CVE M&E Systems 50
 - COIN M&E Systems 57
 - Stabilization/COIN Indices 58**
 - Data Collection Approaches 66
 - Data Use and Learning 70
 - Recommendations..... 71**
 - Improving the State of M&E for CVE/I Programs 71
 - Addressing Gaps..... 73
- Annex A: Scope of Work..... 74**
- Annex B: Interview Guide..... 77**
- Annex C: Interviews and Materials Reviewed 80**

ACRONYMS

| | |
|---------|---|
| ABP | Afghan Border Police |
| ACAP II | Afghan Civilian Assistance Program II |
| AED | Academy for Educational Development |
| AGE | Anti-government Element |
| ALP | Afghan Local Police |
| AME | Asia-Middle East Bureau, USAID |
| ANA | Afghan National Army |
| ANP | Afghan National Police |
| ANSF | Afghan National Security Forces |
| ASI | Afghanistan Stabilization Initiative |
| AUP | Afghan Uniform Police |
| BEST | Basic Education Support and Training Project |
| CAN | Community Needs Assessment |
| CCI | Community Cohesion Initiatives |
| CELI | Consolidated and Enhanced Livelihoods Initiative |
| CLP | Community Livelihoods Project |
| COIN | Counter-Insurgency |
| CSP | Community Stabilization Program |
| CVE | Counter-Violent Extremism |
| DDP | District Development Plan |
| DOD | Department of Defense |
| DRG | Democracy, Human Rights, and Governance Bureau, USAID |
| DSF | District Stability Framework |
| DST | District Support Team |
| EDC | Education Development Center |
| FATA | Federally Administered Tribal Agencies |
| GIRoA | Government of the Islamic Republic of Afghanistan |

| | |
|---------|--|
| GIS | Geographic Information Systems |
| GOK | Government of Kenya |
| GOM | Government of Mauritania |
| GOP | Government of Pakistan |
| GOY | Government of Yemen |
| G-Youth | Garissa Youth Program, Kenya |
| HMEP | Helmand Monitoring and Evaluation Programme |
| IBTCI | International Business and Technical Consultants, Inc. |
| IDP | Internally Displaced Person |
| IGRP | Initial Governance Response Program |
| IMEC | Independent Monitoring and Evaluation Project, Pakistan |
| IP | Implementing Partner |
| IR | Intermediate Result |
| IRD | International Relief & Development |
| IRI | Interactive Radio Instruction |
| ISAF | International Security and Assistance Force, Afghanistan |
| KTI | Kenya Transition Initiative |
| LGCI | Local Government Capacity Index |
| MISTI | Measuring the Impact of Stabilization Initiatives, Afghanistan |
| MPICE | Measuring Progress in Conflict Environments |
| MSI | Management Systems International |
| OCAT | Organizational Capacity Assessment Tool |
| OCHA | Office for the Coordination of Humanitarian Affairs, USAID |
| OTI | Office of Transition Initiatives, USAID |
| PATA | Provincially Administered Tribal Areas |
| PDEV I | Peace and Development, phase I, Sahel |
| PDEV II | Peace and Development, phase II, Sahel |
| PMP | Performance Monitoring Plan |
| POC | Point of Contact |

| | |
|----------|---|
| PPL | Bureau of Policy, Planning, and Learning, USAID |
| PRT | Provincial Reconstruction Team |
| RC | Regional Command, ISAF |
| RC-East | Regional Command-East |
| RC-South | Regional Command-South |
| RCT | Randomized Control Trials |
| RGP | Responsive Governance Program |
| RP | Regional Platform |
| RPGO | Regional Peace and Governance Office, USAID/West Africa |
| SIKA | Stabilization in Key Areas |
| SO | Strategic Objective |
| SOI | Source of Instability |
| STTA | Short-Term Technical Assistance |
| Tashkil | Civil service positions in Afghanistan |
| TCAPF | Tactical Conflict Assessment and Planning Framework |
| TSM | Tactical Stability Matrix |
| USAID | United States Agency for International Development |
| USG | United States Government |
| VE | Violent Extremism or Violent Extremist |
| VE/I | Violent Extremism and Insurgency |
| YMEP | Yemen Monitoring and Evaluation Program |
| YSO | Youth-Serving Organization |

EXECUTIVE SUMMARY

USAID's Bureau for the Asia and Middle East contracted Management Systems International (MSI) to conduct an inventory of program monitoring and reporting systems that track progress in countering violent extremism and insurgency (VE/I). This inventory supports the VE/I Steering Committee's intent to capture useful information from previous and current VE/I monitoring efforts as part of a larger endeavor to produce a field handbook to assist USAID practitioners.

The sample of monitoring efforts reviewed for this inventory is small and diverse. The programs to which they were/are attached also differ greatly. Both CVE and COIN programs are included. The earliest monitoring system examined dates to 2006 and the most recent ones are still under construction. These characteristics pose a challenge for analysis and for making useful generalizations. Included in the sample are two categories of monitoring systems: 1) single-purpose systems attached to a single development program or project; and 2) multi-purpose systems that track progress in meeting program objectives and/or monitor the broader environment (the extent to which stability conditions exist). The Community Stabilization Program (CSP) in Iraq, the Eastleigh and Garissa youth programs in Kenya, and Peace through Development (PDEV) I and II in the Sahel fall into the first category of project-specific approaches. The Helmand Monitoring and Evaluation Programme (HMEP) in Afghanistan, the Monitoring Impacts of Stabilization Initiatives (MISTI) in Afghanistan, the Yemen Monitoring and Evaluation Program (YMEP), the Regional Command (RC)-South monitoring system in Afghanistan, the District Stabilization Framework (DSF) application in Afghanistan, and the FATA/Malakand Stability Index in Pakistan fall into the second category. The purpose of the FATA index is to track stability conditions only.

For each of the 11 programs, MSI reviewed project documents and conducted interviews. Documents included Performance Monitoring Plans, Results Frameworks, program reports, survey data and questionnaires, and evaluations. Telephone and in-person interviews involved USAID staff and implementers and sought to gain a more in-depth understanding of how the monitoring systems were utilized and discuss any challenges, innovations, and undocumented aspects or changes in system design. Documentation was sometimes partial or challenging to locate.

The report captures information on these monitoring systems in four sections. Section I introduces the report. Section II presents summaries, on a standard template, of each of the monitoring systems reviewed. Section III analyzes and compares in sequence the theories of change that could be discerned in these programs, the indicators used to measure common results, and data collection and utilization. Section IV presents recommendations. In addition, Annex A presents the Scope of Work, Annex B provides the interview protocol, and Annex C provides a list of documents reviewed and individuals interviewed.

Our research shows that the narrower implementation programs have results frameworks that tend to articulate a **theory of change** whereas the broader monitoring projects tend to have less explicit causal theories of change, with the notable exception of RC-South which has a formal model of change, the District Stabilization Model. For monitoring systems that are trying to gauge the broader context and not the particular outcomes of programs, like the FATA/Malakand Stability Index, a theory of change is less relevant; such systems do though suggest which conditions drive instability or are pertinent to stability and thus imply a theory of what needs to change in order to achieve stability (but not how those changes might be achieved).

In terms of substantive thrust, all the programs focus on livelihoods as a way to counter insurgency and violent extremism. The CVE (Garissa and Eastleigh youth programs, PDEV I and II), CSP, and Yemen programs focus on livelihoods for youth, whereas the other stabilization programs (with the exception of the DSF application by ASI-East in Afghanistan) do not single out that population. Most of the programs focus on governance and often participation as well. The stabilization programs place greater emphasis on setting up a functioning state and on government legitimacy than do the CVE programs. The stabilization programs

reviewed are broad and operate at national, regional, provincial and local levels but tend to prioritize work at more local levels. Those CVE programs in the sample that included governance focus exclusively on local government. The stabilization programs tend to emphasize security, which is unsurprising, whereas the CVE programs do not. By contrast, all the CVE programs focus on 1) youth engagement in community life and decision making; and 2) enhancing attitudes and behaviors of tolerance and moderation, whereas only CSP and the Yemen program seemed to do so among the stabilization programs. Stabilization programs operating in Afghanistan and Pakistan emphasize rule of law/justice and some emphasize addressing the problem of corruption. A number of CVE and stabilization programs are beginning to focus on community resilience and social cohesion.

Overall, **indicators** developed for both the CVE and COIN programs reviewed show steady improvement and increasing sophistication over time, from 2006 when the first programs in our sample started. We can for example contrast the weak monitoring system developed for the \$675 million Community Stabilization Program (CSP) in Iraq, which ran from 2006-9, with the far more sophisticated efforts apparent in the emerging PDEV II indices and the RC-South/Afghanistan, MISTI/Afghanistan, and the FATA/Malakand Stability Indices on the other hand. These latter monitoring systems have stronger frameworks and more highly developed indicators and approaches to measurement. Older programs in the sample generally relied more heavily on output indicators, which reflected not only the more limited state of knowledge at the time but also the emphasis within USAID on F process indicators. The older monitoring efforts also experienced more difficulties in acquiring timely baseline data and maintaining data quality. There was also less of an awareness in the older programs in the sample of the need to invest in the kind of rigorous M&E systems that would generate useful knowledge for CVE/I programming. Also, there was little expertise available at the time to help in shaping indicators for CVE/I programs so USAID staff and Implementing Partners (IPs) were largely working in isolation, trying to shape indicators on their own.

Investments in M&E have increased significantly and shifted toward more complex, abstract, and meaningful concepts like youth empowerment, community outlook for the future, attitudes toward violence, and stabilization. Indicator sets are more likely now to capture citizen experience, behavior and perception as well as on-the-ground reality. There is a recognition that citizen perception can be volatile in uncertain, high-risk environments and that much-surveyed populations are likely to deliver set responses to frequently asked questions. Much more effort is being invested in trying to capture concepts that are difficult to measure, such as the provision of justice (believed to be a vital factor in stability and resiliency in Pakistan and Afghanistan).

All the programs in our sample incorporate a mixed-methods approach to **data collection**, which allows M&E specialists to validate or cross-check the reliability of data from any given source and which also affords richer and more nuanced information for learning than reliance on any one method. Because of the different types of data required, the difficulty of obtaining data in insecure environments, and the limited availability and quality of government and other sources of data, the mixed methods approach is costly and burdensome. Household surveys (depending on sample size) probably represent the most costly form of data collection, but may provide the most value for money given the importance of experience and perception in these environments. Other forms of data collection such as expert rating panels, Focus Group Discussions (FGDs), in-depth qualitative interviews, and SMS surveys may be useful but pose challenges for aggregation and generalization. With the exception of the earlier CSP, all the COIN programs collect data at the district and sometimes sub-district and village level, which allows for careful monitoring of micro environments. The COIN/stabilization programs collect data more frequently than the CVE programs do in response to their more volatile environments.

The report concludes with a number of **recommendations for the VE/I Steering Committee**, including:

- Establish a wider M&E Community of Practice to advance sharing of approaches, instruments, tools and learning. Collect and share:
 - Good survey plans, sample frames and instruments for CVE/I programs

- Results Frameworks and Performance Monitoring Plans
 - Indices with the logic that underpins them and the definitions of their constituent parts
- Provide STTA and training on CVE/I M&E for relevant program staff and IPs
- Support development of measures of the overall VE environment
- Help solve some of the more difficult measurement challenges (such as how to define and measure social cohesion or the provision of justice) via USAID/Washington investment
- Consider training a carefully selected cadre of regional research organizations to implement CVE/I-relevant surveys
- Track some percentage of youth livelihood beneficiaries (for at least a few years post-project) so that some sense of the return on the investment can be acquired
- Track particularly interesting indicators/indices/measurement approaches over time to assess how useful they are and to determine their possible relevance to other programs/users
- Try to measure the connection between improved services and changes in perception of government legitimacy, and
- Track the utility and uses to which some of the new and more comprehensive systems that attempt to monitor stability conditions to determine cost versus benefits.

I. INTRODUCTION

Scope of Work

Management Systems International (MSI) was contracted by USAID’s Asia and Middle East Bureau to conduct an inventory of current and prior monitoring and reporting systems that track progress in countering violent extremism and insurgency (VE/I). This assignment grew out of the Agency’s recently released policy entitled, “The Development Response to Violent Extremism and Insurgency.” The policy builds on both research and insights from years of program implementation across the globe, including well-funded programs in Asia, Africa and the Middle East. As a follow on to this policy, USAID created a VE/I Steering Committee to guide the implementation of this policy. This scope of work grew out of a request from the Steering Committee to capture useful information/lessons from previous and current VE/I monitoring efforts as part of a larger endeavor to produce an operational guide or field handbook to assist USAID practitioners in the field. MSI’s task was made up of three components:

1. A desk study inventory of past and current monitoring and reporting systems or processes used by USAID to report on, learn about and adapt programs designed to address violent extremism and insurgency.
2. An assessment of the implementation of these same systems through a review of reports and data produced by them and through interviews with those who gathered the information and those who used or were supposed to use this data (including within USAID and IP organizations).
3. An inventory of indicators and drivers which were considered to be most relevant in validating program design and adapting programs.

The programs identified for review by the Scope of Work included:

| Program | Purpose | Donor/Implementing Partner |
|---|---|------------------------------|
| Community Stabilization Program (CSP) in Iraq | Reduce incentives for youth to engage in violence by employing/engaging at-risk youth | USAID/IRD |
| District Stabilization Framework (DSF)/Afghanistan, Pakistan, and South Sudan | Identify local sources of instability, inform programs to address them, and measure change in the stability environment focusing on program output, impact, and overall stability | USAID and DOD |
| West Bank/Gaza Monitoring and Reporting System* | Track project locations using GIS system | USAID Mission Program Office |
| Yemen Monitoring and Evaluation Project (YMEP) | Track stabilization progress and the impact of USAID programs in Yemen | USAID/IBTCI |
| IGRP and CELI Programs in Colombia | | USAID/OTI |
| Eastleigh Program in Kenya | Address youth grievances and reinforce identity; build confidence and capabilities of Somali and Kenyan youth in Eastleigh and neighboring peri-urban slums | USAID/OTI/Chemonics |

| Program | Purpose | Donor/Implementing Partner |
|---|--|---|
| Garissa Youth Program in Kenya (including partial application of DSF) | Address youth grievances; increase youth empowerment, skills, livelihoods, education and engagement in the community | USAID/Education Office/EDC |
| PDEV I in Niger and Chad | Deter marginalized populations from supporting violent extremism by 1) improving community development and local governance; 2) empowering at-risk youth; and 3) rendering superfluous ideologies promoting violence | USAID/West Africa/RPGO/AED |
| PDEV II in Niger, Chad, and Burkina Faso | Enhance community resiliency and social cohesion to constrain VE risk | USAID/West Africa/RPGO/IRD |
| MPICE (Measuring Progress in Conflict Environments) System | Catalogue of indicators for COIN programs emphasizing reductions in grievances and increased institutional capacity | United States Institute of Peace/US Army Corps of Engineers |
| HMEP in Afghanistan | Track stabilization progress and impact of PRT programs and interventions in Helmand | British Foreign Office/Coffey International Development |
| MISTI in Afghanistan | Track stabilization progress and impact of stabilization programs for the Stabilization Unit in USAID | USAID/Stabilization Unit/MSI |
| FATA/Malakand Stability Index in Pakistan | Track stability conditions in FATA and Malakand Division | USAID/OTI, FATA/KP and Program Offices/MSI |
| Regional Command (RC)-South Afghanistan Monitoring System | Streamline the assessment of counterinsurgency campaigns in the US Army's Regional Command (RC)-South | DOD/Caerus Associates |

Eleven monitoring systems were reviewed. MSI was unable to include all the programs listed above, for reasons explained shortly. The sample covered is small and diverse. The programs to which they were/are attached also differ greatly and include CVE and COIN programs. The earliest monitoring system reviewed dates to 2006 and the most recent ones are still under construction. Included in the sample are two categories of monitoring systems: 1) single-purpose frameworks that are attached to a single development program or project; and 2) multi-purpose systems that track progress in meeting program objectives and/or monitor the broader environment (the extent to which stability conditions exist or are changing). The Community Stabilization Program (CSP) in Iraq, the Eastleigh and Garissa youth programs in Kenya, and Peace through Development (PDEV) I and II in the Sahel fall into the first category of project-specific approaches. The Helmand Monitoring and Evaluation Programme (HMEP) in Afghanistan, the Monitoring Impacts of Stabilization Initiatives (MISTI) in Afghanistan, the Yemen Monitoring and Evaluation Program (YMEP), the Regional Command (RC)-South monitoring system in Afghanistan, the District Stabilization Framework (DSF) application in Afghanistan, and the FATA/Malakand Stability Index in Pakistan fall into the second category. The FATA/Malakand Index monitors the environment for stability, while the other category 2 meta-systems are multi-purpose in intent, looking at the impact of programs and at progress in meeting stability conditions.

The full scope of work can be found in Annex A.

Approach

MSI initiated a two-phase process to collect the necessary data:

Phase I: Review of secondary sources. MSI sent out an email of introduction to associated contacts for each program requesting any reports or documents that explain the program's VE/I strategy (Results Frameworks and any accompanying narrative), COIN or CVE specific documents used to shape the M&E system and/or programming, project logframe(s), relevant project descriptions, and/or project/program annual reports and evaluations. MSI reviewed the documents it received in order to gain a broader understanding of strategic approaches to VE/I efforts; to review indicators used to measure VE/I efforts; to become familiar with the design and intention of the system in each case/country; and to compare approaches and systems across the various users. Most of those contacted were USAID staff, with the exception of a few implementers where appropriate. We would like to acknowledge the help of Gavin Helf, Angela Martin, and Lisa Chandonnet of USAID in acquiring relevant program materials.

Phase II: Interviews. A number of telephone and in-person interviews were conducted with USAID staff and implementers to gain a more in-depth understanding of how the programs/systems were utilized, and to discuss any flaws, challenges, undocumented changes, and surprises. In addition, where possible, MSI attempted to investigate any associated costs in building or maintaining the system. The full interview guide that was adapted for each system being reviewed can be found in Annex B.

Outcome

A total of 11 system reviews were conducted from the above-noted list of systems and programs. While modest, this sample does give a sense of the efforts underway. A few of these monitoring systems are new and/or still under development and are thus far untried. The sample's size, diversity of purpose, and the newness of some of the systems included do pose a challenge for analysis and for making generalizations.

For a variety of reasons, including lack of sufficient documentation, a few of the systems included in the Scope were not reviewed, including the West Bank/Gaza Mission's M&E system, the DSF application in South Sudan, and the relevant Colombia program M&E systems. In addition, we excluded the DSF application in Pakistan from the review. OTI noted that they have applied the DSF in a more limited fashion (i.e., the culture matrix and the drivers of stability/instability matrix) in the FATA and Upper/Lower Dir regions of Pakistan and have used the data and analysis as the basis for work plan development and not for M&E. USAID/Pakistan has recently created a new Stability Index, as a supplement to its existing M&E processes for FATA and Malakand; that index is reviewed in this report. MPICE, a catalogue of stabilization metrics with some operationalized indicators (mainly of the popular perception variety), was not assessed in full due to limited evidence of its application in the field. The team found some examples that the MPICE compendium had been consulted in selecting indicators but it was never intended to be used *in toto*. The West Bank/Gaza system acts as more of a Geographic Information System (GIS) that tracks projects sites but does not serve as a means for specialized VE/I data collection and/or analysis. The team was unable to obtain adequate documentation to review the relevant Colombia or South Sudan M&E approaches.

In addition to analyzing these programs, feedback on indicators and data collection plans was provided to some USAID and Implementing Partner (IP) personnel working on current programs.

This review does not constitute a replacement for or comprehensive update of USAID's earlier guide, *Measuring Progress in Development Assistance Programs Countering Violent Extremism: an Introduction*, produced by one of the authors of this report in September 2010. Nor does it incorporate elements of that guide. Our purpose in this current assessment is instead to review real-world CVE/I M&E efforts, and understand how programs are meeting the M&E challenge and trying to learn how to do CVE/COIN programming better.

Organization of the Report

Sections II and III of this report present the team's findings and analysis. Section II of this report presents summaries, on a standard template, of the M&E systems reviewed. Section III analyzes and compares:

- The theories of change that could be discerned in these programs, where relevant
- The indicators used to measure common results
- Data collection approaches, data use, and learning (to the extent that this could be determined; this was an area where the team was able to uncover relatively little documentation)

Section IV presents recommendations for consideration by the Steering Committee.

Annex A presents the Scope of Work. Annex B provides the interview protocol. Annex C provides a list of documents reviewed and individuals interviewed.

II. PROGRAM SUMMARIES

CVE-Related Monitoring Systems

Program: Garissa Youth Program (G-Youth)

Type of M&E system: single project

Overarching purpose: To empower youth via leadership, education, and training opportunities

Implementer/Funder: EDC/USAID

M&E cost (partial): \$3k/quarter for community needs assessment (CNA) survey; \$714 for two radio audience surveys

Location(s): Kenya

Timeframe of program: 2009-2013

Summary of program: G-Youth empowers youth through leadership and civic education, service learning, skills training, radio programming and interactive radio instruction, and grants. Programming mostly addresses VE “push” factors and youth grievances, although radio programming on themes of tolerance, non-violence, and service learning do address pull factors. Five hours of Interactive Radio Instruction (IRI) on civic education for middle school students (younger age group) did not reach the main G-Youth target population but was an add-on.

Theory of change: The G-Youth program is intended to empower youth by improving opportunities for leadership, civic engagement, skill development and education (including English language training for madrasa students to allow them to better integrate into Kenyan society). Empowered youth are more resilient and less at risk of being attracted to/by VE groups.

Indicators: The project initially had two Strategic Objectives (SO): SO 1, improved youth opportunities and capacities; and SO 2, strengthened enabling environment for youth development. However, in 2010 the articulation of the project concept shifted to focus more broadly on youth empowerment via livelihood, learning, leadership, and community engagement opportunities, with corresponding revisions made to the PMP. There are no indicators at the SO or IR levels although indicators placed at the sub-IR level are sometimes at varying levels of ambition and not consistently related to the sub-IR statement; the indicators listed below for Sub-IR 1.1 demonstrate both problems as does Sub-IR 1.5. There is no measure of empowerment and thus little way to begin to test whether these interventions do in fact contribute to empowerment. F process indicators are not included in the indicators listed below, which represent the stronger indicators included in the PMP; many of these are outputs though some do suggest more significant accomplishments than others.

SO1: Improved youth opportunities and capacities

IRI: Increased engagement of youth in their communities

Sub IR 1.1 Improved capacity of youth to participate in community affairs

1.1 # of youth participating in youth summits

1.3 # of youth accessing the youth fund

1.4 # of public information campaigns conducted

1.5 # of youth summits conducted

1.6 Mentorship/curriculum guide; leadership training curriculum in place

1.11 # of community projects implemented by youth

1.12 # of hours of community service donated by youth

Sub IR 1.2 Increased awareness of youth of civic matters

4.1 # of IRI curriculum modules developed for grades 6, 7, and 8 (not the original target population)

4.2 # of upper primary students reached through IRI civic education program (not the target population for G-Youth but an add-on)

4.3 # of youth reached through radio civic education programming

- 4.6 # of 15-20 minute IRI programs in English for 6, 7, and 8 grade developed
- 4.7 # of IRI teacher training guides developed
- 4.8 # of MP3 players provided for IRI classes
- 4.9 # of 30 minute youth-led and –produced civic education radio programs
- 4.10 Increased level of civic awareness among participating upper primary school youth (pre/pst test based on 5 hours of instruction)

IR2: Increased leadership capacity of youth

Sub IR 2.1 Establishment of District-Wide Youth Organization to promote youth interests

1.8 # of youth clubs established

Sub IR 2.2 Promotion of youth leadership in all aspects of program implementation

1.2 # of youth who have completed leadership training

1.9 # of youth participating in project design, implementation, and M&E activities

1.15 % of participating youth feeling capable of leading other youth

2.2 # of youth and adults who attend mentor workshops

IR3: Improved capacity of youth to pursue employment or livelihoods

Sub IR 3.1 Improved skills & attitudes for work & livelihoods among youth

2.1 # of youth trained in work readiness program

2.4 # of youth gaining internships

2.6 # of youth employed through income-generating opportunities

2.7 % of youth feeling better prepared to enter job market

Sub IR 3.2 Improved linkages to financing & support for entrepreneurship

2.3 # of partnerships with local businesses

2.5 # of you accessing business loans through Youth Fund

IR4: Improved education opportunities for youth

Sub IR 4.1 Strengthened quality and access of formal secondary and post-secondary school education

3.1 # of youth receiving scholarships to continue education through the Youth Fund (10% of the Fund)

3.3 # of students enrolled in Cisco Networking Academy at the North Eastern Province Technical Training Institute (project maintains data on completion)

3.8 # of textbooks and other teaching and learning materials provided to secondary school teachers

3.10 # of various computer equipment provided to the North Eastern Province Technical Training Institute

Sub IR 4.2 Improved Madrassas' students' access to work and education through improved English skills

3.2 # of Madrassa students receiving English language instruction

3.9 # of textbooks and other teacher learning materials provided to Madrassa teachers

3.11 % of participating Madrassa students with improved English language competency

SO2: Strengthened enabling environment for youth development

IR 5: Improved capacity of youth-serving organizations (YSO) and institutions in youth development

Sub IR 5.1 Improved capacity of youth serving organizations to meet youth needs

1.7 Amount of financing mobilized through public-private partnership

1.10 # of youth action partnerships established

1.13 % of youth reporting feeling better supported and represented by YSO

1.14 % of youth reporting improved youth relationships with the community

Sub IR 5.2 Improved capacity of education institutions to meet youth needs

- 3.4 # of Madrassa teachers trained to deliver ESL curriculum
- 3.5 # of secondary school career counselors and head teachers trained
- 3.6 # of North Eastern Province Technical Training Institute teachers trained for Cisco Academy
- 3.7 # of administrators and officials trained
- 4.4 # of out-of-school civic mentors trained
- 4.5 # of in-school civic mentors trained

Data Collection: EDC has used a combination of methods: 1) before-and-after surveys with YSO grant recipients and participants in leadership trainings and civic education sessions; 2) a radio audience survey to establish numbers of listeners (only); 3) periodic meetings with the community to collect perception information; and 4) a community needs assessment (described in more detail below). It hopes to conduct follow up tracer studies of some of its beneficiaries in 2012, because it is difficult to form a picture of program impact (e.g., what happens after youth finish vocational training or obtain a loan) from the existing indicators.

Initially, EDC planned to use SMS surveys to follow program participants. However, this proved to be expensive and EDC was unable to obtain the phone numbers of those responding to surveys from the mobile phone service providers, so there was no way to track respondent locations or determine where the sender was based. It also was not possible to verify that the sender had actually been a participant. The program has used SMS blasts for IRI instructors with tips and ideas, and for youth about events and opportunities.

In 2011, EDC was asked by USAID to add a community needs assessment (CNA) as part of an adaptation of the District Stability Framework (DSF). The surveys are conducted quarterly. Data collectors have gone to houses during the day, resulting in an over-representation of the unemployed, students, and housewives. The sample may be skewed toward females. There have been dramatic differences in responses between quarters for some questions, raising issues about the reliability of the data and sampling. EDC believes that reliability has improved since issuing data collectors with electronic tablets for recording survey responses. The CNA contributes little useful information for G-Youth programming. Limited funds for monitoring program outcomes would have been better spent determining the effects of EDC supported radio programs on youth.

Data Utilization: They have used process or output data to make adjustments in their activities and approaches, such as revising a training curriculum. Regular consultation with the community has been important to establishing their credibility as well as allowing G-Youth to adapt approaches when they learn of concerns. This has helped them address and avoid misunderstandings in this conservative culture.

Successes, Innovations, and Challenges: It has been difficult to openly discuss the CVE purpose of the program with local staff, partners, and communities without risking suspicion, distrust, and the loss of effectiveness, making it difficult to have the Implementing Partner measure the VE environment directly.

The program has had trouble recruiting female youth because of the conservatism of the community, and even male youth who participate tend to be from more modern, forward-thinking families. Conservative families have been suspicious of this American-sponsored activity and are concerned about what their children will be taught. The madrasa curriculum had to be vetted extensively, which delayed the start of that particular component. However, the curriculum review process may have engendered some goodwill toward the US within the community as will the work with madrasas. EDC believes that it has successfully used paid and volunteer youth data collectors with tablets to inexpensively conduct the DSF CNA. Some of the youth have then been able to find jobs as data collectors.

The radio programming for youth has reached approximately 600,000 youth, but the project lacks funds to measure impact.

Program: Kenya Transition Initiative (KTI) Eastleigh Youth Empowerment Program

Type of M&E system: single project

Overarching purpose: Foster identity and confidence of Eastleigh youth to allow them to reject extremism

Implementer/Funder: Chemonics/KTI-USAID

M&E cost (partial): \$98k for the first application of Cognitive Edge; \$26k per annum for semiannual data collection by Infotrack, a local survey firm.

Location(s): Nairobi, Kenya

Timeframe of program: 2011-2013

Summary of program: The Eastleigh project works to empower and support youth to create a positive alternative to extremism. KTI will eventually hand the program over to the Mission.

Theory of change: Fostering a positive identity and culture for Eastleigh youth (a form of empowerment) will enable youth to reject extremism. The project has a Results Framework because KTI managers intend to turn the program over to the Mission as they phase out. This is unusual for an OTI program since OTI tends to follow a flexible approach that adapts objectives to rapidly changing environments.

Indicators: There are no indicators attached to the project purpose or IRs, though the **two bolded indicators** under IR 2 below could be used to measure the purpose if the first one were disaggregated for youth. Key indicators are shown below with the strongest *in italics*.

IR-1 Build capacity among youth and community for moderation and non-violence

IR-1.1 Enhanced advocacy capacity and sustained positive dialogue with GoK

- # dialogues/events with GoK
- # of youth reporting less frustrated with and more trustful of gov't agencies

IR-1.2 Improved communication channels for youth to discuss sensitive topics

- # of dialogue forums held on key issues
- # of public information campaigns conducted
- % increase of community perceiving improved communication channels for youth to discuss sensitive topics

IR-1.3 Improve capacity & structure of moderate voices including women and religious leaders

- # of community facilitators trained
- # of respected leaders addressing topic of violent extremism
- % increase of community members perceiving violence as illegitimate means for social change (only asked in the FGDs, with an unusual and not entirely apt question related to whether respondents would like to see a repeat of the 2007 election violence)
- % increase of moderate imams confident and well-trained to disseminate message of non-violence, moderation, & tolerance
- % increase of radical imams more predisposed to disseminate message of non-violence & tolerance

IR-2 Empower local youth

IR-2.1 Increased recreational and cultural activities to foster identity

- # of community members engaged in cultural activities
- % increase of community members having a stronger sense of identity

IR-2.2 Leadership development

- # of youth trained in leadership
- # of community members trained in leadership
- # of youth organization leaders trained
- % increase of youth envisioning and working toward a better future (a concept related to empowerment)

IR-2.3 Network of youth organizations

- # of youth organizations/CBOs supported to improve organizational capacity
- # of youth reporting they are better supported and represented by youth organization
- % increase of youth organization membership base

IR-2.4 Counseling and mentoring

- # of community members receiving counseling and mentoring services
- % increase of community members claiming that mentoring and counseling were crucial in helping them overcome personal issues

IR-2.5 Youth empowerment center

- % increase of youth seeing center as focal point for civic life

IR-3 Livelihood support for youth

IR-3.1 Improved vocational and professional skills

- # of public information campaigns conducted (same as under IR-1.2)
- # of youth trained in work readiness program and working, i.e. entrepreneurship, vocational training
- # of youth trained in professional skills building

IR-3.2 Improved attitude for meritocratic work

- # of small businesses created
- % increase of youth with improved attitude towards meritocratic work

IR-3.3 Improved linkages to finances and other support for entrepreneurship

- # youth accessing government and other financing
- # youth placed in jobs (defined as placed in internships which by nature are temporary)

No definitions of indicators are presented in the PMP. These are being developed as data collection is scheduled. This runs some risk that the actual data collected could diverge from what KTI staff hoped to gain via the indicators unless they are closely involved in reviewing data collection instruments.

Data Collection: Output data are being collected from activities (events, trainings, etc). A youth/adult survey with a sample of 1800 treated and control groups, FGDs, and key informant interviews have recently been implemented to collect data for other indicators. Data will be collected semi-annually. A local survey research firm is tasked with completing all three forms of data collection. The first survey results were to be available in August 2012. In addition, a special survey using the software Cognitive Edge (see below under Successes) is being piloted to determine its utility.

The project has a Steering Committee that includes local community leaders and government personnel. This committee and staff keep an eye on the larger VE context in target neighborhoods.

Data Utilization: The project is relatively new and much of the data collection is recent. Currently, joint team reviews are completed quarterly, according to OTI's model, and an external review is done annually. Grant making is still in early stages so there are not significant examples of using data (aside from the Cognitive Edge survey described below) to shape activities and grants.

Successes, Innovations, and Challenges: SMS micro surveys are implemented following events and are currently obtaining a response rate of about 10%. Currently questions are limited to the quality and utility of the event, but this could present an opportunity to track livelihood recipients' progress throughout time, or as an intervention to spark thinking and debate on substantive issues. One difficulty is that the population is transient and youth are sometimes unwilling to provide cell phone numbers due to suspicion about how this information will be used; this would particularly be the case for émigrés from Somalia.

The project is piloting a new software called Sensemaker by Cognitive Edge that analyzes qualitative data collected from interviews. In this first application, 2800 respondents were asked to tell a story, assume the character of one of the people in the story and respond to questions as that character. While the

first application, which illustrated youth problems and perceptions, was not useful for monitoring, it helped to provide ideas for grant opportunities. In the second application, results pending, the team planned to ask 400 youth and 200 adults for a story about whether anyone they know was ever approached to join a VE organization. This could be useful over the long term if the patterns in the responses/stories change. Sensemaker is very expensive so it needs to be clear that the resulting information justifies the investment. The project has found a way to reduce the costs by identifying a local analyst.

It is difficult to openly discuss the CVE purpose of the program with local staff, partners, and communities without risking suspicion, distrust, and thus a loss of effectiveness. As a result, it is difficult to have the IP measure the VE environment directly itself.

Program: Peace Through Development (PDEV) I

Type of M&E system: single project

Overarching purpose: Mitigate potential for violent extremism (VE) in the Sahel

Implementer/Funder: AED-FHI360/USAID

M&E cost (partial): \$400k for the endline survey; cost data for the baseline not readily available.

Location(s): Niger, Chad (Mauritania was to be included but the GoM did not approve the project)

Timeframe of program: 2008-2011

Summary of program: PDEV I worked on community development and participation in decision making; youth opportunities and inclusion; CSO capacity building; information access and messaging on moderation and tolerance; and work with religious leaders and with inter and intra-faith dialogue.

Theory of change: PDEV aimed to deter marginalized populations from supporting destructive and hostile ideologies that advocate violent extremism by 1) improving local governance (with more emphasis on community development than formal local governance); 2) empowering at-risk youth; and 3) rendering superfluous ideologies promoting violence.

Indicators: Key indicators are included with the stronger ones appearing *in italics*.

IR 1. Improved local governance in target communities (emphasis on urban and peri-urban areas)

1. % citizens in target area satisfied with local decision-making processes (*disaggregated by age and target region*)

IR 1.1 Increased access to quality services and resources—especially those in demand by at-risk groups

2. % of community members satisfied with access to services and resources (*disaggregated by age and target region*)

IR 1.2 Increased collaboration of public, private and civil society sectors

3. % of public, private, and civil society sector members stating collaboration with one another

4. # of initiatives jointly implemented by entities from two different sectors

IR 1.3 Increased and more effective citizen participation (emphasis on youth) in the governance process

5. % of citizens participating in local decision-making processes

IR 1.4 Increased capacity of local CSOs to respond to community needs

6. # of CSOs demonstrating increased capacity to respond to community needs

IR 2. At-risk youth are empowered participants in the community and economy

7. % of youth stating that their opinions are respected by community leaders

8. % of community leaders stating that youth participate constructively in community decision-making processes

9. % of youth expressing an improvement in their economic situation

IR 2.1 Increased quality opportunities in the social, cultural, and economic sectors for youth

10. # of youth participating in youth social, cultural, or economic opportunities w/ project support

11. % of youth reporting satisfaction with social, cultural, or economic opportunities

12. % of community leaders (gov't, traditional, business, civil society) reporting satisfaction with social, cultural, or economic opportunities for youth

13. # of youth that obtained employment with project support

14. # of target youth that started a business with project support

IR 2.2 Expansion of informal and formal social groups for youth

15. New or expanded positive youth social groups (informal and formal)

16. % of citizens stating youth groups make positive contributions to society

IR 2.3 Increased youth civic participation

17. % of youth participating in civic activities

18. % of youth stating positive changes in their communities as a result of their participation in civic activities

IR 2.4 At-risk youth increasingly possess knowledge and skills relevant to the needs of the economy

19. % of graduates satisfied that they have the skills and knowledge needed to obtain a job or start a profitable business

20. % of employers satisfied with skills of graduates from target training programs

IR 3. Ideologies promoting violence are rendered superfluous

21. Moderate ideologies are increasingly perceived as influential and meaningful

22. % of citizens stating that violence is an effective method to solve problems

IR 3.1 Improved information flow among citizens on peace and tolerance

23. % of citizens perceiving increased flow of information on peace and tolerance

24. % of citizens listening to radio programs about peace and tolerance

IR 3.2 Moderate voices strengthened

25. % of citizens declaring the messages of moderate voices as attractive

26. # of moderate leaders participating in PDEV activities delivering messages more effectively

IR 3.3 Traditional educational institutions strengthened

27. # of traditional schools that carry out visioning, strategic planning, and related activities

28. % of instructors in traditional schools that plan to use new teaching tools introduced by the project (activity introduced late)

29. # of traditional schools that implement school enhancements through project support

IR 3.4 Intra-faith dialogue enhanced

30. # of intra-faith dialogues facilitated

31. % of intra-faith dialogue participants stating willingness to engage in further dialogue with members of their faith

32. % of inter-faith dialogue participants stating willingness to engage in further dialogue with members of another religion (Chad only)

33. # of public information campaigns completed by USG programs

34. # of people from at-risk groups reached through PDEV-supported activities

35. # of individuals who received USG-assisted training, including management skills and fiscal management, to strengthen local gov't and/or decentralization

The overarching development objective stated in the narrative accompanying the Results Framework (RF) was “to mitigate the potential for terrorism and extremism in the Sahel.” This was not included in the RF itself, and the project did not try to measure it.

Data Collection: The program relied on survey data for the main outcome-oriented indicators. Due to differences in sampling, the baseline and endline surveys could not be compared. The midterm evaluation had a limited sample size and handled some questions differently from both the baseline and endline, so it also could not be compared to these surveys. The baseline was delayed and of poor quality due to technical issues with the local firm that was hired; for the endline survey, AED hired an international survey research firm to work with a regional West African organization to train data collectors and conduct the survey; this survey collected very good outcome data on the radio component.

Data Utilization: Reliable time-series data do not exist. It is difficult to know what impact the project had, except for the radio component.

Successes, Innovations, and Challenges: PDEV's area of focus shifted as the project progressed to better align with shifting VE risks in the region. Political issues in Niger interrupted the program and after regaining access, interventions were largely limited to media and outreach activities under goal 3. The activities under goal 3, “rendering superfluous ideologies promoting violence” were innovative and included inter-faith and intra-faith dialogues and radio programs promoting peace and tolerance. The Intermedia survey conducted in March 2011 in Chad found that 90% of listeners trusted the information on governance, and 70% trusted the information on youth programs. 98% and 81% of listeners to the youth and good governance programs, respectively, talked about the programs to others, with 88% of those sharing information about peace and tolerance. 60% and 68% of listeners to the youth and good governance programs, respectively, felt that listening to the program helped them or a family member to make a good decision. In Niger, 45% of youth program listeners

said it led them to join with other youth to take action on an issue, and 89% of listeners of the religion series said that it has helped them to become more tolerant of people different from them.

Building of trust and development of relationships was a challenge. It took local partners time to persuade communities to accept USG support and there was a considerable delay in getting to the point of actually implementing activities. It was very difficult to discuss CVE with program staff, and with partners and community members. Emphasizing the positive - peace and tolerance - was a much more successful way to frame program efforts.

Program: Peace Through Development (PDEV) II

Type of M&E system: single project

Overarching purpose: To increase community resilience against violent extremism.

Implementer/Funder: IRD/USAID

M&E cost (partial): \$1.26 million for the household survey for three iterations; in addition, an external contractor will collect data in control communities as well as qualitative research in target communities

Location(s): Niger, Chad, Burkina Faso

Timeframe of program: 2011-2016

Summary of program: PDEV II will continue the work begun by PDEV I in Niger and Chad, and begin new programming in Burkina Faso. PDEV II will continue to focus on empowering youth, community development, local government strengthening, and supporting moderate voices.

Theory of change: Through interventions in high-risk communities focusing on building social cohesion, participatory local development, holistic youth programming, and strategic communications promoting moderation and tolerance, PDEV II aims to increase community resiliency against violent extremism.

Indicators: Many of the program components that existed in the PDEV I framework have been included in the PDEV II framework, but have been realigned and segregated to simplify the presentation. The PDEV II framework also includes an overarching goal, and has changed IRs to SOs to create a sophisticated three-tiered framework. The indicators are under revision at this writing.

Goal: Increased community resilience against violent extremism

G1: Social cohesion index in target communities (under development)

G2: Resilience to violent extremism index in target communities (under development) (This is being revised to focus on individual-level resilience)

G3: Youth outlook index in target communities (under development)

SO 1: Youth are more empowered

SOI 1.1: % of adults stating that youth make positive contributions to society

SOI 1.2: % of youth who completed leadership training and are active participants in community events.

SOI 1.3: % of youth assisted by PDEV II who improve their livelihoods

SOI 1.4: % of learners who enroll in Grade 4 and write exams at the end of Grade 3 in PDEV II assisted schools.

IR 1.1: Expanded Youth Livelihoods

IND 1.1.1: # of individuals trained (Youth - Vocational)

IND 1.1.2: # of youth assisted to improve their livelihoods

IR 1.2: Increased Access to Education

IND 1.2.1: # of individuals trained (Youth - Adult literacy)

IND 1.2.2: # of formal and non-formal schools benefiting from PDEV II support (Value of support / # of learners)

IR 1.3: Strengthened Youth Leadership in Civil Society

IND 1.3.1: # of individuals trained (Youth - Leadership, Participatory theatre, Multi-media)

IR 1.4: Increased Youth Mobilization and Engagement in Local Development

IND 1.4.1: # of PDEV II supported community events with significant youth participation

SO 2: Increased Moderate Voices

SOI 2.1: % of individuals who report listening to PDEV II radio programs regularly in target populations

SOI 2.2: % of individuals who report sharing / discussing the thematic content of PDEV II radio programs

SOI 2.3: % of radio stations assisted by PDEV II that produce and broadcast their own CVE content

SOI 2.4: # of occasions on which leaders engage with PDEV II to deliver CVE aligned messages to target populations in public venues or through mass media

IR 2.1: Increased Capacity of Media Outlets

IND 2.1.1: # of individuals trained (Media Outlet staff)

IND 2.1.2: # of radio stations assisted by PDEV II

IR 2.2: Increased Access to Quality and Credible Information

IND 2.2.1: # of public information campaigns completed by PDEV II

IND 2.2.2: # of hours of CVE material produced, broadcast and rebroadcast by PDEV II

IND 2.2.3: # of hours of CVE material produced, broadcast and rebroadcast by PDEV II supported radio stations

IR 2.3: Increased Positive Dialogue among Religious Leaders and with the communities they serve

IND 2.3.1: # individuals trained (Religious leaders)

IND 2.3.2 # of inter and intra faith activities facilitated by PDEV II

SO 3: Increased Civil Society Capacity to address Community Issues

SOI 3.1: Organizational Capacity Assessment Tool (OCAT) scores of targeted CSOs

SOI 3.2: % of CSOs submitting grant applications to address community issues to PDEV II that successfully implement funded activities

SOI 3.3: % of individuals who report engaging with local power structures to effect change

SOI 3.4: % of individuals who report engaging with local power structures to effect change successfully

IR 3.1: Increased CSO Capacity

IND 3.1.1: # of individuals trained (CSO staff)

IR 3.2: Increased Citizen Participation and Advocacy

IND 3.2.1: # of individuals trained (CDC members)

IND 3.2.2: # of Community Development Committees formed and meeting regularly

IND: 3.2.3: # of people attending PDEV II supported community events

SO 4: Strengthened Local Government

SOI 4.1: % citizens in target areas satisfied with local decision making processes

SOI 4.2: Local Government Capacity Index (LGCI) score of targeted municipalities

SOI 4.3: % of individuals who report that local government takes into account the opinions of citizens in local decision making processes.

SOI 4.4: # of municipalities implementing identified best practice in participatory, accountable and transparent local decision making as a result of PDEV II intervention.

IR 4.1: Improved Local Government Capacity

IND 4.1.1: # of individuals trained (Local Gov. Officials)

IND 4.1.2: # of municipalities that learn about best practice in participatory, accountable and transparent local decision making as a result of PDEV II intervention.

IR 4.2: Increased Transparency and Accountability in Local Decision Making

IND 4.2.1: # of community events held to discuss local governance issues

IND 4.2.2: # of best practice in participatory, accountable and transparent local decision making identified or developed as a result of PDEV II

IR 4.3: More Participatory Local Development

IND 4.3.1: # of community development projects successfully completed with PDEV II assistance that respond to community development plans

Cross Cutting Indicators

IND 5.1.1: # of community events conducted under PDEV II

Initially, the goal was to decrease the risk for violent extremism. However, USAID/West Africa preferred to frame the program in terms of what the US stood for rather than what it stood against. The goal was revised to 'increasing resiliency' which is a more defensible position, and one that local people can buy into. However, the goal is considered outside the manageable interest of the program; this is an unusual structure for a project in that there is no one master

purpose or result for which the mission is accountable.

Data Collection: IRD is planning a survey that will be outsourced to Intermedia, which performed well on the PDEV I endline survey. The sample will be around 1,600 per country; they will pull samples proportionate to the population mix with an adequate sample of youth and women. There will be a baseline conducted in two batches, a midline, and an endline in year 5. An external group with a contract with the Mission will conduct the survey in control communities as well as conduct VE qualitative research in target communes as needed. Other forms of data collection are under review.

Data Utilization: IRD has not yet begun its first survey.

Successes, Innovations, and Challenges: This is an expensive monitoring effort with several different kinds of data collection, but it may do much to advance understanding of the role of development assistance in CVE. The goal level indices are innovative but work is still underway to flesh them out, deciding on weighting and aggregation, turning them into survey questions, and then testing their utility. Social cohesion could overlap with community resiliency or the two could move so much in sync that only one would be needed so most recently the team is considering emphasizing individual level resiliency for the resiliency index and leaving social cohesion at the community level. There are some hints in the literature that community resilience/social cohesion can provide more protection than individual level resilience because all individuals are affected by their environment and generally operate within the norms of that environment.

Framing the goal in a positive light enables USAID and IRD to talking openly about the program to staff, partners, government, and local communities, which in turn engenders trust and local ownership. However, not discussing the CVE purpose also means that it is difficult to have the Implementing Partner measure the VE environment directly itself.

COIN-Related Monitoring Systems

Program: Community Stabilization Program (CSP)

Type of M&E system: Single project

Overarching purpose: Reduce incentives for participation in violent conflict.

Implementer/Funder: IRD/USAID

M&E cost: Unknown

Location(s): Iraq

Timeframe of program: 2006-2009

Summary of program: This \$675 million project was the largest ever, civilian-led USG counterinsurgency project.

Theory of change: CSP supported USAID/Iraq's Strategic Objective 7 "Focused Stabilization: reduce the incentives for participation in violent conflict." Originally the project had three Intermediate Results (IRs): IR 7.1 Unemployment decreased with a focus on young men; IR 7.2 Conflict mitigated through increased civil society organization and community activities; and IR 7.3 Community infrastructure revitalized and essential services provided by local government. Mid-way through the project, USAID dropped the third IR because the US State Department and the US Army were also engaged in infrastructure support and the Iraqi government began undertaking more of this work itself. The shift to two IRs improves the logic as the program had a two-fold approach to reducing the incentives for participation in violent conflict. It focused on:

- Giving youth jobs, and
- Engaging youth to ground them and give them a stake in the community, and interact with youth from other backgrounds to foster tolerance.

The dropped IR relating to infrastructure and essential services really served the purpose of providing short-term jobs. The logic is clear that if unemployment and youth marginalization fueled the violent conflict, then jobs and youth tolerance and engagement programs would be effective stabilization programs. However, much of the youth tolerance and engagement programming involved sports and recreation activities, which may have had a less direct impact than activities with a conflict mitigation component. It is also unclear from documents why the program focused singularly on youth and left aside work with sectarian leaders and broader efforts to promote tolerance and mitigate conflict such as strategic communications.

Geographic Targeting. CSP operated in 15 strategic cities such as Baghdad, Mosul, Kirkuk, and Basrah—many of them in hotly contested oil-producing areas or in Al Anbar where Al-Qaeda and insurgent activity were focused. Some cities, like Kirkuk, made it an express policy to divide their work equally among all the districts irrespective of their insecurity level in order to avoid accusations of favoritism, which might exacerbate insecurity levels. Other CSP city programs like Basra targeted specific districts that were of strategic importance to the Provincial Reconstruction Teams (PRTs) and shifted the focus of their district level interventions over the course of the project in response to new security challenges. Each CSP city program's geographical targeting was a unique response to the highly dynamic security environments and extremely diverse ethnic and religious composition of specific governorates that was negotiated with a wide range of GOI and USG partners.

Gender Targeting. From May 2006 to June 2008, the principal focus of the CSP was on male youth. Even with this focus, a growing number of city programs had quietly begun adding programs to better address the needs of women—especially young widows with dependent children. The justification for this change was twofold. First, many of these women were destitute with children—especially young boys—who would be prime recruits for insurgent activity. Second, CSP city programs found that helping widows was a quick way to reduce local suspicions about the project being a military operation and to gain wider community appreciation. Once the number of female suicide bombers (many of them widows) became a strategic concern in late 2008, this gradual shift was supported by USAID.

Indicators:

The M&E plan had fifteen indicators, as follows:

SO 7 Focused Stabilization: reduce the incentives for participation in violent conflict

Perception of citizens on the effectiveness of local government to provide services

Number of insurgent incidents decreased

IR 7.1 Unemployment decreased with a focus on young men

7.1 Unemployment rate decreased

Sub IR 7.1.1 Jobs created

7.1.1.1 # of person months of employment generated for short term employment

7.1.1.2 # of long term jobs created

7.1.1.3 # of person days of employment

Sub IR 7.1.2 Employable skills improved

7.1.2.1 # of participants completing vocational skills training

7.1.2.2 # of apprentices placed

Sub IR 7.1.3 Businesses created and expanded

7.1.3.1 Total value of financial support to businesses

7.1.3.2 # of businesses receiving assistance

7.1.3.3 # of people completing business skills training

IR 7.2 Conflict mitigated through increased civil society organization and community activities

7.2.1 # of youth participating in non-formal education programs

7.2.2 # of activities completed thru integrated decision-making between LG and CSOs

IR 7.3 Community infrastructure revitalized and essential services provided by local government

7.3.1 # of Public Works Projects completed

7.3.2 # of Public Works Projects on-going

All the IR and sub-IR level indicators are output indicators and do not measure impact. The two SO-level indicators attempt to measure impact but both are problematic. The first, perception of citizens on the effectiveness of local government to provide services, does not track very well with the thrust of the program. Program documents do not articulate a development hypothesis noting that citizens' perception of increased local government effectiveness will reduce incentives for participation in violent conflict. To gather perception information, a CSP subcontractor conducted a survey asking participants if their community was safer, government services were better, and relations between religious and ethnic groups were less tense due to the program. Data should have been compared with data of nearby areas that did not benefit from the program to increase confidence in observations and analysis regarding correlations. There was no measurement of support for violent groups/ideologies in the target communities or trend line measurement.

The second SO-level indicator, number of insurgent incidents decreased, is problematic as the number of confounding variables makes any inferences of causation between the CSP and the number of violent incidents potentially misleading (e.g., the effect of similar programs like CERP or criminal activity not disaggregated from violent incident statistics). Better impact indicators were needed to assess causal linkages between CSP projects and program objectives. The project should have attempted to measure the impact of short-term employment on local stability and whether there was a reduction in propensity to violence by the participants. This might have been done through exit polls or focus group discussions with program participants or through general surveys of comparable districts/cities that participated or did not participate in the CSP.

Data Collection: The program used data collected from CSP bi-weekly reports (often compiled from grantees), the Brookings Institution index for insurgent activities, national unemployment statistics, and household survey data collected by a sub-contractor (the Lincoln Group). IRD hired local Iraqi M&E staff, which allowed USAID to monitor activities because ex-patriot staff were restricted from traveling to most program sites. Given the unstable environment and the lack of a national ID program, IRD found it easier to monitor the business development program than the vocational education program since the former was tied to a location.

Data Utilization: CSP's M&E system showed that the cost of creating a job under the business development program increased with the size of a grant, suggesting that smaller grants were more cost effective. Smaller grants entailed more work for the Implementing Partner, however. The data led USAID to establish a \$50K ceiling for routine grants and require USAID approval for larger amounts.

Successes, Innovations, and Challenges: The monitoring systems, including the evaluations carried out by a third-party monitor, did not explore the link between participation in violent conflict on the one hand, and program activities - jobs, business grants, community infrastructure, and youth activities - on the other, so measuring program impact is problematic.

Initially, IRD's local M&E officers reported to the program manager for their respective province, but in 2008 they began reporting to an M&E Director with a Ph.D. that IRD hired, which notably improved CSP's M&E efforts.

Program: District Stability Framework (DSF)/Afghanistan

Type of M&E system: multi-purpose

Overarching purpose: DSF identifies local sources of instability, informs programs to address them, and measures change in the stability environment focusing on program output, impact, and overall stability.

Implementer/Funder: USAID and DOD with a recent effort to involve Afghan officials

M&E cost: unavailable

Location: Afghanistan, predominantly in RC-East, RC-Southwest and to a limited extent in RC-South, RC-West and RC-North.

Timeframe of program: 2010 - present

Summary of program: DSF provides a standardized framework that utilizes a range of situational awareness, planning and programming tools to inform stability programming. In early 2010, OTI's Afghanistan Stabilization Initiative-East (ASI-East) became the first program to use the DSF as a programming tool. Later that year, Regional Command-East (RC-East) began using DSF. Both ASI-East and RC-East made contributions to the monitoring component of DSF, which is discussed below.

Theory of change: The DSF analyzes the causes behind sources of instability and directs programming to respond to those sources. It is designed for use at the district level or below. Stability working groups develop a series of matrices, including a tactical stability matrix (TSM) for their district to design programmatic interventions and associated indicators. Components of these matrices include: analysis of local political and cultural environment, sources of instability, perceived and actual causes of sources of instability, programming objective, and indicators at three levels – output, impact, and overall stability.

Indicators: The DSF prescribes monitoring at three levels – output, impact, and overall stability.

For the first level in ASI-East, OTI provided a list of standard output indicators that stability working groups could draw from. For monitoring at the second level, ASI-East reviewed DSF workbooks created at the field level and identified 11 impact indicators that could track progress in addressing the five most common sources of instability identified across all districts in which the program worked. For each indicator, assessors rate change on a five-point scale, including: strong positive change, some positive change, no change or inconclusive, mixed results, or negative change. The following table describes the relevance and data sources of the 11 impact indicators.

| ASI-East Impact Indicators | Relevance to ASI-East Objectives | Data Sources | | | |
|--|---|--------------|-----------|-----------|-----------------------------|
| | | Elders | Residents | Officials | Stability Index (see below) |
| District Government | | | | | |
| District/informal government responsiveness | ASI-East activities were designed to demonstrate the government's responsiveness to community's needs | x | x | | |
| Visits to the District Center | Visits to the district center are one measureable way to assess community engagement with district government | | | | x |
| Satisfaction with officials' involvement in development | ASI-East "hard" activities were designed to showcase positive government engagement in development work | x | x | | x |
| Perception of district/informal government officials' capacity | Capacity building for officials was a component of almost all ASI-East district programs | x | x | x | |
| Communication with residents | ASI-East supported strategic communication between district governments and residents, through media and traditional channels | | x | x | |

| | | | | | |
|---|--|---|---|---|---|
| Trust in district government/ informal government | ASI-East activities presented an opportunity for residents to engage with officials and build trust | x | x | | |
| Customary Leadership | | | | | |
| Relationship between elders and district/informal officials | ASI-East activities presented an opportunity for elders and officials to work together | x | | x | |
| Tribal Cohesion | | | | | |
| Perception that tribes are united | ASI-East activities were designed to bring people from different tribes to work together on common problems | x | x | x | |
| Dispute Resolution | | | | | |
| Satisfaction with dispute resolution of district/ customary leaders | ASI-East activities sometimes promoted dispute resolution by district leaders, or directly resolved disputes with infrastructure | x | x | | x |
| Youth | | | | | |
| Perception that district/ informal government works for youth | District leaders were encouraged to lead ASI-East youth activities to demonstrate their commitment | x | x | | |
| Perception that youth take an active role in the community | ASI-East activities provided an opportunity for youth to engage with elders to solve community problems | x | x | x | |

For the third monitoring level, ASI-East developed the methodology for measuring the DSF's seven stability indicators, as shown in the table below. The individual variables, which are all articulated as percentages, are aggregated at the district level for each stability indicator as an overall percentage and the seven stability indicators are then aggregated to generate an overarching stability index for the district. ASI-East is then able to compare district-level progress over time. All seven categories are treated equally in the aggregation.

ASI-East Stability Indicators

1. District Government Recognition

- Percentage of respondents who believe that the district government represents one's interests either always or most of the time
- Use of district government for justice services (percent of those using services reporting that they came for justice related services)
- Percentage of the population visiting the district center

2. Civilian Security

Percentages of people reporting:

- Attacks against civilians in past six months
- Fear of such attacks for self and family
- Attacks perpetrated by the Taliban, Afghan Border Police (ABP), Afghan National Police (ANP), Afghan National Army (ANA), criminals, other civilians and security companies

3. Market Activity

- Percentage of shopkeepers reporting improving business
- Percentage of shops that are newly opened in the district

4. Afghan National Security Forces (ANSF) Presence

Percentages of people reporting:

- ANA and ANP patrols at the village level
- Checkpoints outside their district center
- ANA and ANP provided security in their district
- ANA and ANP effectively protected the community

- ANSF was a source of insecurity (accounted for negatively)

5. Freedom of Movement

Percentages of people reporting:

- Never feeling safe leaving the house, village and district
- Encountering anti-government elements on the road
- “Always” encountering trouble on the road
- “Never” leaving the district
- “Always” feeling safe leaving their house, village and district (accounted for negatively)
- Percentages of drivers reporting encountering trouble on the roads in the daytime

6. Perceptions of Security

Percentages of people reporting:

- Family is safe or very safe in village
- District has become safer in last three months
- ANA always effectively protects community
- ANP always effectively protects community
- Afghan Local Police (ALP) always effectively protects community
- No sources of insecurity in district

The index also accounts negatively for percentages of people reporting that ANA, ANP, ALP, ABP are sources of insecurity and nobody provides security in the district

7. Perceptions of Governance

Percentages of people reporting:

- Positive trust in district government, provincial government and national government
- That they use the district government for justice services

After RC-East began using DSF as a programming tool, USAID incorporated learning from ASI-East and developed a set of 30 impact and stability indicators. The regional platform expected Field Program Officers at DSTs (district support teams) and PRTs (provincial reconstruction teams) to track output data for their programs but not to report it up to the regional level. This reduced the reporting burden for district stability working groups from 316 metrics down to 30 indicators. The 30 indicators included 16 quantitative metrics, nine perception-based metrics and five subjective judgments of USG officials. The 18 measurable metrics represented an advance in using more objective data. The selection of the 30 indicators gave considerable thought to the safety of the data collectors charged with obtaining data for those indicators.

| RC-East Impact and Stability Indicators | Data Source |
|---|--|
| Afghan on Afghan violence | |
| Population feels safe to travel and that Government of the Islamic Republic of Afghanistan (GIROA) protects them from insurgents and criminals. (Score: 1-10) | Recommendation by DST/PRT Point of Contact (POC); validated by Platform section chief |
| Number of violent incidents reported to local ANSF | ANSF reporting |
| Number of Afghan patients treated at health clinic due to Afghan on Afghan violence | Ministry of Public Health reporting (provincial or district level), if possible. Request for Information to local clinic if necessary. |
| Percentage of village clusters (mantaqas) which ANSF patrols at least every two weeks. | Coalition Forces and/or ANSF patrol reports |
| Number of attacks on GIROA officials and ANSF personnel and/or infrastructure | Significant Acts reporting |
| "Who brings the most security to your mantaqa?" | BINNA household survey percentage (aggregate ANSF/GIROA) |
| "Who brings the most insecurity to your mantaqa?" | BINNA household survey percentage (aggregate anti-GIROA) |

| | |
|---|---|
| Community advocacy | |
| District government enables basic functioning of society. (Score: 1-10) | Recommendation by DST/PRT POC; validated by Platform section chief |
| Number of citizens traveling to district center for services or grievance resolution | District center administrative log book |
| Number of citizens participating in district council meetings or its local equivalent | Coalition Forces, GIRoA, or FSN counts participants at district council or GIRoA linked shura/jirga meetings. |
| "If you had a problem, who would you go to for help?" | BINNA household survey |
| "If the Taliban returned to power, would it be good/bad for the people and the country?" | BINNA household survey |
| "What are the top three reasons to support the Taliban?" | BINNA household survey |
| Dispute resolution | |
| Formal or legitimate informal institutions resolve disputes effectively for the local community. (Score: 1-10) | Recommendation by DST/PRT POC; validated by Platform section chief |
| Number of disputes resolved monthly by district governor | District Center administration reports |
| Number of disputes resolved monthly by formal judge(s) | District Center administration reports |
| Number of disputes resolved monthly by legitimate informal structures (local shuras/jirgas with haqooq validation) | Haqooq records |
| Economic activity | |
| Local population demonstrates increased economic activity and investment (Score: 1-10) | Recommendation by DST/PRT POC; validated by Platform section chief |
| Number of commercial vehicles using the district's main market road on a given day each week | On local market day, count commercial vehicle traffic for 12 hours (observable by ANSF, GIRoA, or worst case, Coalition Forces/FSN) |
| Number of hectares irrigated on a quarterly basis | Analyze satellite imagery and compare to same time period last year |
| Number of local Afghans involved in programs and training provided by Directorate of Agriculture, Irrigation and Livestock and/or agriculture extension agent | Count number of Afghans attending local programs by Directorate of Agriculture, Irrigation and Livestock or agriculture extension agent |
| Number of non-agricultural businesses added along main market road (stores, cafes, service providers) | Count number of new businesses along main market road |
| Number of hectares planted with perennial crops (mostly nut and fruit producing trees) | Analyze satellite imagery and compare to same time period last year |
| Number of hectares planted with annual crops (vegetables, grapes, etc.) | Analyze satellite imagery and compare to same time period last year |
| Development activity | |
| Local population perceives GIRoA as providing services (Score: 1-10) | Recommendation by DST/PRT POC; validated by Platform section chief |
| Number of patients treated at Ministry of Public Health clinic monthly | District Ministry of Public Health representative report; GIRoA clinic reporting |
| "What are the top three reasons to support the government?" | BINNA household survey |
| Percentage of people satisfied with development in their district | BINNA household survey |
| Percentage of people that say they are better off than they were six months ago. | BINNA household survey |
| "What is the biggest problem facing the district?" | BINNA household survey |

Data Collection: The ASI-East implementer DAI conducted an impact assessment in four districts in 2011, and three districts in 2012 as well as an impact baseline study in four districts in 2012. The M&E team chose a qualitative approach for expediency (the program is short so data were needed quickly) and the value of in-depth data for program teams. Data sources included interviews with elders and district officials, FGDs with residents, and quantitative and qualitative data from the overall stability assessment (described further below). To interpret the data and provide necessary context, the team also conducted interviews with DAI and OTI staff and with military and civilian officials at the relevant Forward Operating Bases, and examined ASI-East documents, including the DSF workbooks, district strategies, program documents, grantee reports, activity final monitoring reports, and any district studies. The second impact assessment included input from 81 individual interviews and 14 FGDs with more than 100 residents.

ASI-East also engaged a third-party research firm (RSI) to conduct two rounds of an overall stability assessment—one in June 2011 that covered 10 districts and one in April 2012 that covered 12 districts. For the second assessment, the firm conducted a household survey targeting 3,249 adult men and women in 162 communities; 32 FGDs (24 with men and eight with women); one month of data collection inside the district government office for all 12 districts; and convenience samples including a drivers' survey, a government officials' survey, a shopkeepers' survey, district market observation, and basket of goods price-tracking.

Going forward with Community Cohesion Initiatives (CCI), the successor program to ASI-East, OTI intends to coordinate with MISTI on data collection twice a year.

RC-East conducted a baseline of the 30 indicators in summer 2012 and planned to collect data quarterly thereafter. District stability working groups comprised of USG civilian and military officials, Afghan officials, and local elders, generate the data. RC-East uses Afghan partners to collect data for the quantitative measures, seeing this as a way to build Afghan capacity in monitoring dynamics in their own area, i.e. commercial traffic along their main market road and Afghan-on-Afghan violence. Several metrics draw on BINNA, a household survey conducted quarterly at the district level by the International Security Assistance Force in Afghanistan. The high frequency of “don't know” responses and overall survey fatigue, however, pose significant reliability changes to the data. Also, social desirability bias may result as Afghans know what answers they should give. In the Tactical Conflict Assessment and Planning Framework (TCAPF)¹ survey, for example, respondents surmised if they said the national government should be responsible for fixing a problem, the district might receive funds from the national government.

Data Utilization: Data from ASI-East's impact and overall stability assessments were used to inform programming, particularly through day-long roundtable discussions with program teams in 2011. The assessments have also been used to provide situational awareness and data to other USG programs, military entities, donors, and monitoring and evaluation firms. They were also used to inform programming in the follow-on program, Community Cohesion Initiatives (CCI).

As RC-East has just begun collecting data on its revised set of metrics, it is too soon to evaluate utilization.

Successes, Innovations, and Challenges: Recognizing the merits of ASI-East's overall stability assessments, MISTI will measure overall stability for the USAID Stabilization Unit districts.

RC-East added observable, objective indicators to complement subjective data points and worked to involve Afghans in data collection. In addition, it modified its approach to measuring stability. It observed significant correlation between the impact indicators that monitored daily Afghan life and realized that by looking at those impact indicators holistically they could obtain an accurate portrayal of the extent to which people were able to function and live their lives in a more

¹ TCAPF was the predecessor for DSF and had many of the same elements in its final iteration before the name change.

normal fashion; this gave programmers a concise and observable way to measure stability trends. High turnover of personnel in RC-East and changing leadership buy-in to DSF monitoring efforts hampered implementation of some innovations. For example, DSF personnel at RC-East developed a stability index based on the ASI-East's overall stability assessment that focused on five condition statements, but this was never used in practice.

Scoring on a scale of one to ten for the five subjective judgment indicators used in RC-East asks USG officials to make distinctions between conditions on the ground that are too fine. The difference between a six and a seven may be difficult to determine reliably across time and across people.

For the stability indicators used in ASI-East, the percentages reported for individual variables are simply aggregated for each stability indicator and the seven stability indicators are then aggregated to generate an overarching stability index for the district without any standardization of the responses or weighting of the individual variables or seven stability indicators.

Program: FATA/Malakand Stability Index

Type of M&E system: measure stability conditions

Overarching purpose: To provide USAID and other USG stakeholders with a broad, macro-level perspective of stability in the region and the extent to which stability changes over time. At the same time, USAID wants the exercise to test the underpinning theory of change but to explore and utilize different methodologies for data collection. In understanding how local perceptions influence overall stability; identify and define actions needed to address root causes of instability; and assess the progress toward stability in specified regions.

Implementer/Funder: MSI/USAID

M&E cost: Roughly \$900,000 including Implementing Partner (IP) staff time to implement and design

Location(s): FATA and Malakand Division, Pakistan

Timeframe of program: N/A

Summary of program: The FATA/Malakand stability index offers a streamlined way for USAID to track progress in meeting conditions for stability.

Theory of change: The FATA/Malakand stability index is not intended to measure program performance or impact, but is meant to measure the stability conditions and to guide programming. For example, if some areas or aspects of stability are deteriorating, additional resources can be moved to program in those areas or related to these aspects (provided the latter are amenable to development assistance programs). USAID pursues three end state objectives in the region:

- The tribal populations of FATA and PATA are not susceptible to the messaging of violent extremist groups;
- The legitimacy and writ of the GOP are acknowledged, the state has a monopoly on armed force, and there is a recognized acceptance of the need for an end to extremist violence; and
- FATA and PATA become inhospitable to the perpetrators of organized, premeditated, religious/political violence.

Indicators: The index identifies four elements as key to stability: security, governance, economic security, and social structures. Under these four metrics fall a number of indicators; the indicators themselves are composites.

Goal: *FATA and Malakand Division reach a level of stability within the GOP framework and societal norms.*

Security

Indicator 1: Freedom of Movement

Indicator 2: Incidents of Violence

Indicator 3: Territory Control/Presence

Governance

Indicator 4: Basic Services

Indicator 5: Justice

Indicator 6: Civic and Political Participation

Economic Security

Indicator 7: Livelihoods

Indicator 8: Market/Bazaar Activity

Social Structures

Indicator 9: IDPs

Indicator 10: Social Gatherings

Indicator 11: Status of Women

It should be noted that the instrument is not an index in terms of measurement, producing a single number. The output will be a comparative analysis of the data semi-annually.

Data Collection: Data will be collected using mixed-methods, combining annual household surveys, FGDs held outside of FATA (for safety and in the hope of reducing social desirability bias), direct observation (as with bazaar activity), media monitoring (e.g., for violence and social events), and reporting by IPs (e.g., on freedom of movement). An expert panel will be used to help capture justice on an annual basis. USAID's M&E project (MEP) will conduct annual surveys through a local research firm in both FATA and Malakand. For FGDs, social mobilizers hired and trained by the local survey firm will recruit participants; more than half the mobilizers are female and they will be accompanied by chaperones. Participants will be brought to Peshawar for the FGDs to ensure their safety and that of the participants and facilitators. Participant travel and per diem costs will be paid and they will receive remuneration for their time. Women will be accompanied by chaperones.

Data Utilization: Index details and methods of operationalization are still being worked out. Data have not been collected as of Sept. 2012.

Successes, Innovations, and Challenges: This is an interesting (but still very new and thus untested) attempt to boil down complex phenomena to a manageable number of indicators. Because of the different types of data required, the difficulty of obtaining data (and genuine responses in surveys) in insecure environments, and the limitations on (and reliability of) government and other sources of data, the approach is costly but the narrower set of indicators make it less costly than some other approaches. Cost will diminish over time, after instruments are tested and refined and survey firm trained.

The approach to FGDs worked well in its first iteration, particularly the attempt to secure female participation. Many of the female participants had not been outside FATA and seemed pleased by the opportunity to be in Peshawar. (Note that HMEP gave up trying to conduct FGDs with women and replaced them with individual interviews because of difficulties getting women to participate in group interviews – see the subsequent matrix on HMEP.) Skype was used so that staff could unobtrusively monitor FGDs and provide timely and discreet suggestions to facilitators on topics to pursue further.

There are some gaps in the index. It does not look formally at overall confidence levels in the GOP and how that is linked to legitimacy – that will be partly an analytic exercise drawing on other data collected. There is also nothing included about popular perception of the Political Agents (the dominant GOP political authority) in FATA due to sensitivities. There also is little on tribal governance and writ, except via jirgas for dispute resolution and crime.

Program: Helmand Monitoring and Evaluation Programme (HMEP)

Type of M&E system: multi-purpose

Overarching purpose: Develop a comprehensive plan to monitor and evaluate progress against the Helmand Plan.

Implementer/Funder: Coffey International Development (ID)/DFID and the Helmand Provincial Reconstruction Team

M&E cost: \$2 million per year

Location: Afghanistan

Timeframe of program: 2009-2014

Summary of program: Coffey ID designed an M&E system to evaluate progress for all ongoing development projects in Helmand.

Theory of change: The Helmand theory of change focuses on state-building for a sustainable transition and focuses on the strengthening the “capability, accountability, and responsiveness of the state, building a political settlement, and managing spoilers.” Concentrating on all of these aspects addresses the underlying causes of conflict and thus creates space for a sustainable transition.

Indicators: There are no exact criteria that define a district-level transition given the variation in political realities and conditions in each district. Instead, HMEP helps assess transition readiness by comparing developments in each district using six key perception indicators. Each indicator is compared against the levels achieved in Lashkar Gah when the decision was made to transition. Based on representative surveys of heads of household, the indicators cover:

- **Security:** Percentage of heads of household rating the security situation in their area as somewhat good or good.
- **Afghan National Police (ANP):** Percentage of heads of household saying that the ANP in their region act in the interest of the local people
- **District Government:** Percentage of heads of household saying that the district government listened to and acted on behalf of the people rather than in the interest of its officials. (Captures responsiveness and accountability.)
- **GIRoA:** Percentage of heads of household agreeing that GIRoA, generally speaking, is doing a good job.
- **Justice:** Percentage of heads of household saying they are satisfied or very satisfied with statutory justice services.
- **Taliban:** Percentage of heads of household who do not think it would be a good thing for the people and the country if the Taliban were to return to power and govern Afghanistan.

HMEP has also identified a range of indicators that fall under the three lines of operation (security, governance, and development).

1. Security

- Perceptions of overall security and security on roads
- Perceptions of GIRoA’s influence in delivering security
- Satisfaction with key actors’ (government vs Taliban) efforts in delivering security
- Perceptions of police capacity
- Confidence and trust in the police
- Trust in the army

2. Governance

- Legitimacy of state institutions (GIRoA, district government, and district community councils)
- Size of provincial government’s staff
- Amount of development funds drawn down by provincial government from the central government
- Proportion of respondents believing that the district government does a good job in developing their district
- Proportion of respondents believing that local government bodies are effective in dealing with problems in the community
- Proportion of respondents reporting that the district government had improved access to information and news over the last three months
- Proportion of respondents who would welcome a stronger national government presence in the area

- Proportion of respondents willing to reintegrate anti-government elements in order to end the conflict
- Proportion of respondents who consider job creation essential to help reconciliation
- Proportion of respondents who expect successful reconciliation
- Proportion of respondents who view corruption as a serious problem
- Proportion of respondents who believe corruption is worsening
- Number of statutory justice mechanisms
- Proportion of respondents satisfied with state justice
- Proportion of respondents reporting that the government had improved the effectiveness of justice services
- Proportion of respondents confident in future improvements
- Size of the poppy economy in Helmand
- Support for the Taliban

3. Development

- Proportion of respondents satisfied with government education services and the availability of government schools
- Proportion of children attending government school
- Proportion of girls and boys attending government and non-government schools
- Proportion of respondents satisfied with the provision of health services
- Proportion of respondents noting an improvement in healthcare over the last three months
- Proportion of respondents optimistic about the state of their economy
- Proportion of respondents agreeing that the district government was going in the right direction with regard to the local economy
- Average household incomes
- Proportion of households living under the poverty line
- Percentage of farming households saying that they do not grow poppy and that the sale of licit crops is their main source of income
- Percentage of households in Helmand province that have electricity within their home
- Proportion of Helmandis experiencing shortages of clean drinking water
- Proportion of respondents satisfied with roads

A comprehensive M&E framework for the Helmand Program was not developed until 2009, nearly three years after the UK government entered the province in mid-2006.

Data Collection: Data are collected from several sources. Afghan teams collect quantitative and qualitative data, complemented by primary and secondary data from third-party sources. HMEP also uses a quarterly head of household survey, with roughly 4,000 respondents in 11 districts; qualitative interviews of 80 men and women in 8 districts to provide additional insight and validate quantitative findings; and triangulation of data using other PRT and RC-SW data, other surveys, and secondary sources. Many of the interviews have been conducted with men; FGDs have not worked with women so now interviewers talk to women individually.

Security issues have had an impact upon data collection. Interviewers may only go to cleared areas and grey areas, but not to no-go areas held by the Taliban. Interviewers were also supposed to carry GPS systems to verify that they went to specified places, but due to perceived risks of carrying a GPS some interviewers refused to do so.

Data Utilization: HMEP produces quarterly and annual reports that provide evidence for further investment in programs. They use heat maps to show the transition readiness of each district. Formerly, 2-3 day discussions with leadership were held each quarter to discuss the data; those discussions now take place every six months. It produces different kinds of reports for different types of audiences with higher levels of aggregation for higher command levels. Because it works at the provincial level (and not regional or national), it does not find aggregation at highly synthetic levels much of a risk. HMEP also generates “dashboards” – simple graphic summaries updated regularly that give senior officials a sense of progress over time against key indicators. The HMEP website is unclassified, so those who receive prior authorization from the UK government can access the data. The primary users are coalition government officials, their contractors, and UN agencies. Most end users just read the reports they receive via email. Some analysts in the survey and defense community who work on Helmand have worked with the raw data. Data sharing deals have been set up with Rand and USIP but they have not made use of the data so far.

Successes, Innovations, and Challenges: HMEP has been able to construct outcome-level proxies for assessing campaign-level effects, a technique that could be replicated by other projects. AusAID has just asked Coffey to replicate HMEP in the adjacent Uruzgan province.

HMEP attempts to monitor all of the programs in Helmand, but monitoring has brought up political sensitivities with some stakeholders in the international community when progress on indicators is lackluster. Furthermore, some individual projects have their own light monitoring systems in place. The ability to monitor progress is somewhat constrained by the lack of baseline data prior to 2009-10 when HMEP became operational.

HMEP tries to use qualitative data to explain why something is or is not changing and to help inform adaptations in programming but has found it difficult to integrate such data in a systematic fashion into its assessments and reports and to make full use of data that are collected in part because the data were sometimes partial, spotty and opportunistic so it was not easy to gauge utility. Coffey intends to modify its qualitative methodology in Uruzgan province by interviewing key informants in the fields of medicine, education, religion, and infrastructure as well as in government, ANA, ANP, and Coalition forces.

Program: Monitoring Impacts of Stabilization Initiatives (MISTI) in Afghanistan

Type of M&E system: multi-purpose

Overarching purpose: Measure and map stabilization trends and impacts for the USAID/Afghanistan Stabilization Unit (STAB-U).

Implementer/Funder: MSI/USAID

M&E cost: \$15 million over three years, \$5.5 million in additional funding seems likely; implemented nationwide and includes 11 program evaluations, one experimental, and two quasi-experimental impact assessments and a 40,000 sample household survey.

Location: Afghanistan

Timeframe of program: March 2012-2015

Summary of program: MISTI supports the M&E capacities of Implementing Partners (IPs) and conducts mid-term and final performance and impact evaluations of SIKA, CCI, CDP, and other stabilization programs in Afghanistan.

Theory of change: USAID/Afghanistan STAB-U programs seek to help reduce key sources of instability (SOI) by engaging and supporting at-risk populations, extending the reach of the GIRoA to unstable areas, providing income generation opportunities, building trust between citizens and the government, and encouraging local populations to take an active role in their development.

Indicators: MISTI has compiled a set of indicators from existing programs in Afghanistan that are complemented by indicators developed specifically for MISTI. Indicators are classified into three types:

- Common Indicators (C): activity-level indicators common to all STAB-U programs. These include four standard F indicators, and one gender-specific F indicator.
- Stability Index Indicators (SI): Gathered by MISTI, these indicators will be used in the Maturity Model, to assess each district's overall stability level, stability trends, and the impact of programs. The maturity model aggregates data for these indicators into five phased categories: normal (most desirable); stable; fluid; unstable and insecure (least desirable) to insecure
- Program Specific Indicators (PS): activity-level indicators specific to ACAP II.

Note that the MISTI Results Framework is under revision as of November 2012.

AO: Sufficient stability achieved to enable transition to Afghan-led sustainable development.

7.a % of Afghans reporting their environment has become more stable (SI)

7.b % of Afghans reporting that the country is moving in the right direction (SI)

7.c % of Afghans sympathizing with motives for anti-government violence (SI)

7.d % of Afghans who participate in resolving community problems (SI)

7.e % of targeted communities reporting that their lives have changed for the better (SI)

IR 1: Community instability arising from war affected families reduced

7.1.a # of target districts with completed stabilization activities executed against identified SOI (C)

7.1.2.a # of war-affected families assisted (PS)

7.1.2.b % of families who report that tailored assistance provided has helped to recover and rebuild their lives (PS)

7.1.2.c % of families who report that tailored assistance has been delivered in a fair and transparent manner (PS)

7.1.2.d # of person days of employment created through stabilization projects (C)

GNDR-2 proportion of female participants in USG-assisted programs designed to increase access to productive economic resources and basic services (assets, credit, income, or employment) (C)

IR 2: Initial access to government established

7.2.a % of Afghans reporting that GIRoA listens to their grievances (SI)

7.2.b # of projects completed with GIRoA involvement (C)

7.2.1.a # of community-based organizations trained to develop and implement community development projects (C)
7.2.1.b # of stabilization activities completed with community contribution (C)
7.2.2.c # of visits and community engagements by provincial and district GIRoA officials (SI)
1.6.1-13 % of people who have access to information about government services and programs via the media, social media, or other lines of communication (C)

IR 3: Intervention areas' resiliency increased

7.2.c # of projects completed with community involvement (C)
7.2.2.c # of visits and community engagements by provincial and district GIRoA officials (SI)
1.6.1-13 % of people who have access to information about government services and programs via the media, social media, or other lines of communication (C)
1.6.2-14 # of people participating in Stability Analysis Tool stabilization work sessions (C)

IR 4: District entities able to mitigate SOIs more effectively

7.1.a # of target districts with completed stabilization activities executed against identified SOI (C)
7.1.1.a # of target districts in which the DSF has been used to develop planned activities (C)
7.1.1.b # of stakeholders trained in stabilization methodology (C)
7.2.b # of projects completed with GIRoA involvement (C)
7.2.c # of projects completed with community involvement (C)
7.2.1.a # of community-based organizations trained to develop and implement community development projects (C)
7.2.1.b # of stabilization activities completed with community contribution (C)
7.2.1.c # of stabilization activities implemented through grants to CBOs (C)
7.2.2.a % of targeted communities reporting availability of GIRoA-delivered basic services (SI)
7.2.2.b % of tashkeel fill in targeted districts (SI)
7.2.2.c # of visits and community engagements by provincial and district GIRoA officials (SI)
7.3a # of target communities with follow-on development activities (C)
7.3b % of beneficiaries reporting GIRoA development efforts (SI)
GNDR-2 proportion of female participants in USG-assisted programs designed to increase access to productive economic resources and basic services (assets, credit, income, or employment) (C)
2.2.3-8 # of sub-national government entities receiving USG assistance to improve their performance (C)
1.6.2-12 # of stabilization work sessions conducted for district entities (C)
1.6.2-14 # of people participating in Stability Analysis Tool stabilization work sessions (C)

Regarding indicators for the AO, 7a is more specific to stability than 7b. 7c is a sophisticated indicator, because it requires the respondent to unpack the drivers of anti-government violence. 7d has a Western orientation. It could be that only the elders participate, which people see as sufficient to represent their interests. 7e is a strong indicator because it is based on personal experience.

Data Collection: MISTI will use data collected from IPs and other sources in addition to its own surveys. MISTI surveys will collect household data involving quasi-experimental and possibly RCT sub-samples to measure intervention impacts and track stabilization trends over time. The survey aims to include around 40,000 respondents in 83 districts across Afghanistan. Within every USAID district, MISTI plans to identify matched pairs of 12 villages with one set for each treatment and control (24 villages total). These pairs will be selected from the larger pool of villages sampled as part of the baseline survey. With this large a sample, MISTI should be able to test different theories of change.

The more reliable questions in the survey ask people about their own experience. For security indicators, they ask how secure respondents feel in their home or when traveling. For quality of life indicators, they ask people to compare their quality of life at present to 1 to 2 years ago, and to the future.

Data Utilization: MISTI will organize a series of summits for stabilization program M&E practitioners and interested parties designed to share best practices and lessons learned. This group will meet on at least a quarterly basis. MISTI is also developing a knowledge management portal that will include evaluations and reports, raw datasets, maps, and GIS information.

Successes, Innovations, and Challenges: The use of common indicators across all stabilization programs along with a stability index and program-specific indicators is a good innovation.

MISTI had also planned to conduct RCTs, but those plans are currently on hold as some of the IPs may not be able to select villages based on MISTI's RCT criteria to guarantee a sufficient degree of control. However they can do quasi-experiments because of the large sampling size in the survey. They will include enough villages to do an impact evaluation.

USAID/Afghanistan leadership has stressed the importance of MISTI's regular M&E Summits in bringing together USAID's M&E community and sharing best practices.

Program: Regional Command-South (RC-South)

Type of M&E system: multi-purpose

Overarching purpose: Streamline the assessment of counterinsurgency campaigns in the US Army's Regional Command (RC) South.

Implementer/Funder: Caerus/10th Mountain Division, US Army

M&E cost: Unknown.

Location: Afghanistan

Timeframe of program: 2010-2011

Summary of program: Caerus refined a monitoring system that collects and uses of a variety of data to inform counterinsurgency activities.

Theory of change: Caerus and the 10th Mountain Division identified a model of stability applicable to RC-South. The model depicted a positive cycle of increased **security** contributing to **governance institutionalization** and, in turn, to greater **public confidence** in non-insurgent institutions, which then prompted greater **community resiliency** and increased **security** (see model below). The model could also depict a negative cycle of decreased security undermining the other conditions.



Figure 2.2. District Stabilization Model. In an improving stability environment, improvements in real-world stability conditions (*Loop 1*) drive improved popular confidence, which in turn contributes to community resiliency (*Loop 2*), prompting further real-world improvements. The same effect operates in reverse in a declining stability environment.

-Developed by Caerus for RC-South through field observations, 2010-2011.

Indicators:

Caerus reduced the number of metrics (in bolded text below) used in the RC-South monitoring system from 84 to 11, described by 18 indicators (in non-bolded text), as follows:

Resiliency of government institutions

1. Government Official Tenure & Quality
2. Depth of Government
3. Budgetary Process

Survivability of Government Officials & Community Leaders

4. Survivability Assessment

Public Confidence in Government Legitimacy and Effectiveness

5. Popular Perception of Government
6. Popular Usage of District Center

Predictability and Acceptability of Corruption

7. Popular Perception of Corruption

Strength of Shura and Councils

8. Shura and Council Assessment

Civilian Freedom of Movement

9. Freedom of Movement Assessment

Function and Effectiveness of Rule of Law

10. Reliability and Accessibility of Rule of Law
11. Popular Perception of Rule of Law

Degradation of the Insurgency

12. Effectiveness of the Insurgency
13. Capacity of the Insurgency
14. Will of the Insurgency

Function and Effectiveness of Government Security Institutions

15. Popular Perception of Government Security Forces and Institutions
16. Afghan National Security Forces' Ability to Defeat the Insurgency

Functioning Community Security

17. Popular Perception of Local Security

Resilience of Subsistence Economy

18. Economy Assessment

One of the innovations of this approach was developing a system whereby progress against all indicators can be tracked and compared. Using a system of Rating Definition Levels, each indicator was assigned a corresponding 1-5 Likert scale with pre-determined definitions for each level. Using this system, all relevant information can be categorized and compared regardless of source, from perception surveys, to behavioral observations, to key information interviews. Equal weight is applied to all 18 indicators, which can also be clustered under the four variables within the Theory of Change—security, governance institutionalization, public confidence, and community resiliency.

This information is displayed on a radar diagram (shown below) that offers a functional, visual description of where the campaign is falling short, where it has reached the desired end state, and how commanders can shift resources accordingly. Radar diagrams also show how the interplay between each of the 11 metrics can affect campaign objectives.

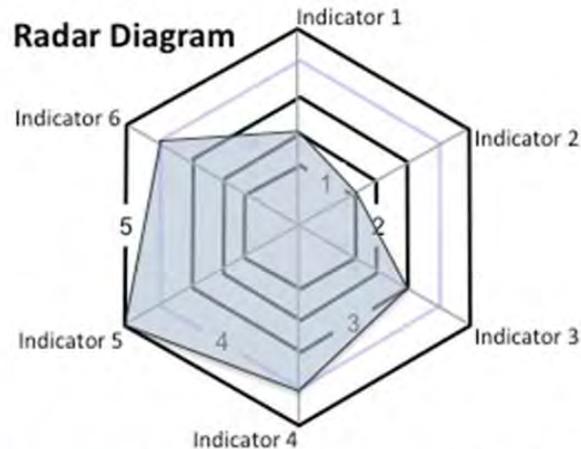


Figure 2.4 Dive deep; present simply. In RC-South we used radar diagrams, and other off-the-shelf formats, to provide key decision makers a holistic understanding of the rigorous triangulated measures of stability.

Data Collection: The Army reports at the local level and feeds the information up to the regional command/division headquarters. Caerus also supports a local research network that feeds information back to the assessment team. Each network is composed of three-four Afghan research managers who each manage 5-30 people in their district, who can then perform qualitative “deep dives,” meaning interviews and FGDs, as needed. The use of local researchers has been critical as they can go places that the US military cannot go. To date, there has been no compromise in security from this approach, and security protocols are in place for each of the local researchers. Three analysts from Caerus were embedded with the RC-South Assessments Cell at division headquarters to help with data collection and analysis, and had access to all of the information that came through the Assessments Cell.

Data Utilization: The analysis informs the RC-South Commander’s and the Department of State Senior Civilian Representative’s decision cycles. However, in practice it is unclear if and how the data are shared with civilians in the field.

Successes, Innovations, and Challenges: The streamlining of the metrics has enabled assessments to inform command decisions at the operational level more quickly, and thus more frequently. When first implemented, the refined system cut the assessment process from six weeks down to two weeks, and finally to a matter of days. Eventually a comprehensive campaign assessment could be produced virtually on demand.

The use of the Rating Definition levels and the Likert Scale has enabled comparisons across indicators.

Monitoring efforts need to be aware of survey fatigue in populations subject to such frequent surveys that “canned” responses are likely. Social desirability bias is also a significant risk in unstable settings where respondent security concerns loom large.

The actual cost of such a monitoring program is unclear since much of it falls within the army's personnel budget. The level to which civilians have access to data and make use of these indicators is also unclear.

Program: Yemen Monitoring and Evaluation Program (YMEP)

Type of M&E system: multi-purpose

Overarching purpose: Tracking stabilization progress and the impact of USAID programs in Yemen

Implementer/Funder: IBTCI/USAID

M&E cost: \$7.7 million over three years; with two option years for \$11.8 million

Location: Yemen

Timeframe of program: 2011-2013

Summary of program: This monitoring and evaluation program sought to provide innovative assessments of stabilization for USAID/Yemen's 2010-2012 stabilization-focused strategy. However, the Arab spring and ensuing conflict led to evacuations of USG and Implementing Partner (IP) staff in early 2011, which stalled implementing projects and YMEP. When operations resumed with a pared-down USG staff, new leadership in the Embassy and Mission dropped the stabilization innovations and asked YMEP to focus on monitoring discrete activities of the IPs.

Theory of change: The 2010-2012 U.S. strategy toward Yemen was two-pronged: 1) strengthen the Government of Yemen's ability to promote security and minimize the threat from violent extremists within its borders, and 2) mitigate Yemen's economic crisis and deficiencies in government capacity, provision of services, transparency, and adherence to the rule of law. The strategy aimed to improve stability and security by improving governance and helping to meet pressing socio-economic challenges. Two assistance objectives were expected to contribute to achievement of the USAID strategy, as laid out in the 2011 PMP:

Assistance Objective 1: Livelihoods in vulnerable communities improved

Intermediate Result 1.1: Employment opportunities increased

Intermediate Result 1.2: Access to and delivery of quality services improved

Assistance Objective 2: Governance capacities to mitigate drivers of instability improved

Intermediate Result 2.1: Public policies and institutions facilitate more equitable socio-economic development

Intermediate Result 2.2: Local governance and basic service provisions addressing community-level needs improved

Intermediate Result 2.3: Community-based institutions and mechanisms to ensure active participation in governance and locally-driven solutions strengthened

The underlying hypothesis was that improving services and opportunities in underserved communities would lead to improved political and social stability. Also explicit in the strategy was a geographic focus on villages, communities, tribal areas, districts, governorates, and/or regions that were vulnerable and posed probable risk to Yemen's overall stability. The ten governorates given greatest priority based on the Interagency Conflict Assessment were: Amran, Al Jawf, Hajjah, Marib, Sana'a, Shabwah, Abyan, Al Dhale'e, Lahj, and Aden. IPs were expected to develop a stability framework (a modified district stability framework) for each district or theme (such as microfinance or participatory governance). The stability framework is a compilation of: instability-related drivers (e.g., grievances) and/or triggers present for a local area or broader theme; the activities programmed to address these grievances/triggers; and the indicators used to measure success.

This strategy represented a shift from the customary, sector-based development interventions to a focus on key grievances and drivers of instability in targeted, disadvantaged communities. This shift was in response to escalating development challenges and the deteriorating security context, especially the increasing evidence of Al-Qaida operatives utilizing Yemen's territory as a base for operations as well as the Sixth Sa'ada War and intensified violence in the south of the country.

As noted in the program summary above, the turnover of USG leadership in Yemen and the down-sizing of Mission staff led to a shift away from the more innovative stabilization focus and to a more traditional development focus. This shift will likely be captured in the strategy under development for the new fiscal year. IPs did not complete stability frameworks.

Indicators:

The 2011 PMP laid out high-level indicators to track whether the overall stability situation was static, improving or declining, but these weren't used in practice:

- No. of districts/themes with significantly improved stability outlook
- No. of districts/themes with marginally improved stability outlook
- No. of districts/themes with unchanged stability outlook
- No. of districts/themes with marginally reduced stability outlook
- No. of districts/themes with significantly reduced stability outlook
- Estimates of time interval until stability issues would no longer significantly impede standard development efforts: >36 months; 12 – 36 months; 6 – 12 months; 0 – 6 months; 0 months

In addition to locally focused stabilization indicators, the PMP included democracy, economic, and human capital indicators at the national level that tracked changes in the overall country context. These national level indicators were drawn from the Monitoring Country Progress (MCP) analysis used by USAID's Europe and Eurasia Bureau and ensured YMEP would not lose sight of macro-level changes. They included ratings on a 5-point scale for: civil liberties, political liberties, corruption, rule of law, media, government effectiveness, regulatory quality, business environment, budget balance, trade, GDP per capita growth, environment, uneven development, energy security, domestic credit, exports, FDI, inflation, GDP per capita, school enrollment, health and education expenditures, life expectancy, under 5 mortality rate, and literacy rate.

YMEP also included indicators for the six projects that implemented the strategy. The two largest projects, representing roughly 95% of the budget, were the Community Livelihoods Program (CLP) and the Responsive Governance Program (RGP).

Data Collection: The PMP proposed identifying IP panels and external experts to make assessments of progress towards stability for each district or theme on a quarterly basis. IP panels would consist of the IPs currently operating in each district or theme, while experts would include academics, USG partners, USAID officials, or representatives from other donors or delivery organizations with specialized knowledge of the districts or thematic areas. They would enter their assessments in a decision assessment tool and load them onto the YMEP web-based information system. The decision assessment tool would prompt panelists to examine: the stability outlook (static, improving or declining); estimates of time interval for stability improvement; relevance of existing instability drivers or triggers; effectiveness of activities in addressing stabilization; suggestions for more effective activities; usefulness of indicators for monitoring stability; and stabilization effects that were not measured and suggestions for how to measure them. YMEP also anticipated partnering with a third party opinion polling firm in Yemen to administer baseline surveys across targeted communities regarding their perceptions of stability, but this did not happen.

Data Utilization: USAID can submit queries to the data system, which generates customized reports. In practice, it's unclear how much USAID has used this function.

Successes, Innovations, and Challenges: The monitoring program laid out innovations in stabilization indicators and data collection, namely relying on IPs and experts to evaluate stability in a district/theme, but the shift in circumstances precluded their use.

III. COMPARISON OF PROGRAMS

Theories of Change

This inventory examines two kinds of monitoring systems: those belonging to implementation programs and those encompassing multiple implementation programs and/or focused on the broader stabilization environment. The implementation programs include all the CVE programs in our sample as well as one stabilization program, the Iraq Community Stabilization Program. The broader programs include all the other stabilization programs.

What is a theory of change?

A theory of change articulates a set of beliefs about how and why change happens. It provides a semi-structured change map that links strategic actions to changes that are desired. It enables often implicit, untested assumptions which guide strategies to be drawn out, scrutinized, validated, and tested. It provides a common understanding to all stakeholders and sharpens planning. It is critical to the development of useful and cost-effective monitoring systems.

The implementation programs we looked at have results frameworks that articulate a theory of change. The Community Stabilization Program (CSP) in Iraq, for example, was a counterinsurgency program aimed at reducing incentives for participation in violent conflict. The theory of change embedded in CSP was that incentives could be reduced by employing or engaging youth, ages 12 to 35. CSP had three Intermediate Results (IRs):

- IR 7.1 Unemployment decreased with a focus on young men;
- IR 7.2 Conflict mitigated through increased civil society organization and community activities; and
- IR 7.3 Community infrastructure revitalized and essential services provided by local government

To better align activities with the program focus on employing or engaging at-risk youth, an early modification to the project placed more emphasis on youth in the conflict mitigation component (IR 7.2) and more emphasis on youth employment generation in the community infrastructure component (IR 7.3).

Similarly, the CVE programs in this inventory articulate a clear theory of change. The Eastleigh and Garissa Youth programs in Kenya aim to employ, empower, and integrate youth into their communities and the broader society so that they are not attracted to anti-system groups. PDEV I and II aim to mitigate the potential attractiveness of extremist elements by developing stronger, more resilient communities and more capable youth. They include a focus on youth employment and engagement, but also address the broader community with initiatives to discredit ideologies that promote violence, increase civil society capacity to address community issues, and strengthen local government and community decision making. The Results Frameworks in use on the CVE programs under review, however, do not necessarily capture the desired end result in terms of increased resistance to VE.

By contrast, the broader monitoring projects appear to have weaker or less explicit causal theories of change, with the exception of RC-South. The Helmand M&E approach, for example, combines existing strategies and plans into a larger framework, but is starting from the plans and working up rather than starting with the problem analysis, identifying a theory of change and then devising programs that would respond to the theory. It pulls together plans from ISAF, the donors (principally the British, US, Danes, Estonians, and UN), and the Afghan government at the national and provincial levels.

Similarly, MISTI pulls together four existing USAID stabilization programs—Afghan Civilian Assistance Program II (ACAP II), Community Development Program (CDP), Community Cohesion Initiatives (CCI), and Stabilization in Key Areas (SIKA)—which contain divergent theories of change. It articulates an overarching logic that simply incorporates the purpose of each program: “The USG’s stabilization programs seek to help reduce key SOIs by engaging and supporting at-risk populations [ACAP II], extending the reach of GIRoA to unstable areas [CDP], providing income generation opportunities [CDP], building trust between citizens and their government [CCI], and encouraging local populations to take an active role in their

development [SIKA].”² To some extent, MISTP’s very large household survey, described later in this paper, should enable it to test different theories of change.

YMEP pulls together a number of existing programs as well: Basic Education Support and Training Project (BEST) and Community Livelihoods Project (CLP) in support of the livelihoods component; and Promoting Civic Youth Engagement and Non-Traditional Actors, Strengthening the Role of Youth in Cross-Tribal Conflict Mitigation Processes and Community Development, Yemen Community-Based Conflict Mitigation Project, and Responsive Governance Program (RGP) in support of the governance component. The PMP posits a clear theory of change that improving services and opportunities to underserved communities leads to improved political and social stability. Of the programs we classed under the COIN rubric, YMEP shows the clearest dual focus on countering violent extremism and countering instability; others are not unconcerned with VE but stabilization receives more emphasis.

The FATA/Malakand Stability Index and the RC-South approach also fall into this category of broader contextual monitoring as opposed to capturing the outcomes of implementation projects. RC-South does

Understanding the Intervention Logic – HMEP

The HMEP team formulated a results chain which made the linkages within and between programs explicit. In health programming, for example, the team articulated the following problem statement:

“The population of Helmand lacks access to quality basic services. Lack of access to health services weakens the social contract by signaling to the populace that the state is unable or unwilling to respond to its needs. States which have a weakened social contract with their people are vulnerable to challenge from belligerent communities and resolution of differences through conflict.”

The statement led to the following evaluation questions: Are those with access to healthcare more likely to perceive the government as legitimate or less likely to express preference for the Taliban? What constitutes an acceptable level of access?

Source: Adapted from “Integrated Planning and M&E in Stabilisation Contexts,” Stabilisation Unit UK, February 2012, prepared by Coffey International Development.

articulate a strong causal logic. It posits several factors as key to stability: security, governance institutionalization, community resiliency, and public confidence in non-insurgent institutions, which it depicts as operating in two dynamic loops. In this approach, economic resiliency is part of community resiliency.

The FATA/Malakand Stability Index, unlike the others, does not provide a clearly specified logic model because it is not meant to capture the combined impacts of program interventions. It considers four elements as being key to stability: security, governance, economic security, and social structures. For OTI, important aspects of its work in FATA have long included trying to strengthen community-government interaction and community participation in governance, as well as improving services and objective community welfare in order to increase support for the legitimate government. These aspects do play a role in the index; the target population’s view of the government is however derivative of other elements in the index and not an element.

Unlike the other programs that we looked at, the DSF in Afghanistan does not have an over-arching theory of change due to its focus on district-level dynamics. The DSF focuses on the root causes that are contributing to instability in a district in order to identify the desired conditions and activities that would diminish instability. The theory of change is embedded in this exercise and the resulting tactical stability matrix (TSM). There is a different TSM for each district, but no overarching one for the whole country, so there is no easy way to roll up TSMs to a higher level.

In a COIN environment, one might expect the support to at-risk populations to particularly highlight youth, but most of the programs considered (or at least their monitoring frameworks) do not appear to do that. One exception is the DSF/ASI-East framework in Afghanistan. Generally speaking, however, youth seem to be a heavier emphasis in purely CVE programming.

² USAID Stabilization Unit Afghanistan Performance Management Plan, FY 2012 - FY 2015

The accompanying table (see pg 44) depicts the main components of each program. The last five columns call out whether the program includes a focus on livelihoods, governance, security, tolerance, or youth engagement. As can be seen in the column labeled “livelihoods,” all the programs that we looked at focus on livelihoods as a way to counter insurgency and violent extremism. Hypotheses include:

- With adequate livelihoods, people can meet their needs, are less aggrieved at being excluded, and might be less attracted to illicit means of earning income or anti-system actors.
- People who have work to occupy their time and give them a sense of purpose might feel more a part of the larger society and be less attracted to anti-system agendas and less susceptible to recruitment by anti-system groups.

Activities include cash for work on short-term infrastructure or cleaning projects, training, apprenticeships, micro-grants, and business development. The CVE programs and CSP, Yemen, and the FATA Stability Index among the stabilization programs focus on livelihoods. CSP and Yemen focus on livelihoods for youth, whereas the other stabilization programs do not single out that population.

The column labeled “governance” shows that most of the programs focus on governance as well. The stabilization programs place greater emphasis on setting up a functioning state and on government legitimacy than the CVE programs. Stabilization programs operate at both the national and local levels but tend to prioritize work at the local level. Where they include this component, CVE programs focus exclusively on local governments. The injection of government is assumed to contribute to stabilization. Activities under this component aim to promote accountable, responsive government, which reduces citizen grievances and increases government legitimacy and, in turn, makes anti-system groups less appealing. Depending on how severe the problem is and how generous the resources are, the approaches vary, but draw from the following elements: strengthening local government and improving interaction with citizens; improving delivery of services and infrastructure; reducing corruption; and improving the provision of justice (although the CVE programs in the sample do not include justice programming and do not appear to have an overt emphasis on corruption). Service provision seems to be a key lynchpin in the COIN programs, with the idea that improvements in services will enhance government legitimacy (assuming government provision of services or the popular perception thereof) and undermine the appeal of anti-system actors.

The stabilization programs tend to include an emphasis on security, which is unsurprising. Generally, the CVE programs do not do so in their programming, though their selection of indicators may be affected by concerns about a backlash from asking sensitive questions. For COIN programs, the aim is to improve security conditions so that people feel less pressured to support or give way to extremist actors and so that some semblance of normal life can return. RC-South also focuses on community-provided security as a form of security “redundancy” while most others focus on government-provided security and the reality on the ground in terms of how security affects people’s lives. Development agencies play a smaller role in this stabilization effort than defense and police professionals, and tend to focus on legal frameworks, civilian management of security institutions, and cooperation among security-related and civil institutions.

By contrast, all the CVE programs focus on tolerance and moderation, whereas only CSP and YMEP appear to do so among the stabilization programs. The tolerance component aims to undermine the attractions of violent groups and ideologies through messages on tolerance and non-violence, opportunities to interact with people from other groups, and conflict mitigation training. The CVE programs focus principally on tolerance messaging and to some extent on inter-and intra-faith dialogue, whereas CSP and YMEP focus principally on bringing diverse groups together and conflict mitigation training. Radio programs dealing with tolerance and moderation and that give youth and community members a change to engage in dialogue seem in the cases assessed to have had strong impact.

Similarly, the CVE programs focus on youth engagement, whereas only CSP and YMEP do so among the stabilization programs. The youth engagement component aims to empower youth, integrate them fully into the community, and bolster self-respect and identity so that they are not attracted to anti-system groups. Activities include cultural and social activities, life skills training, counseling, mentoring, and leadership training.

Table 2. Focus of Theories of Change or Conditions Captured by Monitoring Systems

| M&E System or Program | Objectives | Key Components or Conditions | Livelihoods | Governance | Security | Tolerance | Youth Engagement |
|---|--|---|-------------|------------|----------|-----------|------------------|
| Stabilization/Counterinsurgency Programs | | | | | | | |
| Community Stabilization Program, Iraq | <ul style="list-style-type: none"> Reduce incentives for participation in violent conflict | <ul style="list-style-type: none"> Give youth jobs/cash for work Engage youth to ground them, give them a stake in the community, and foster tolerance | X | | | X | X |
| DSF, Afghanistan | <ul style="list-style-type: none"> Stabilize the district | <ul style="list-style-type: none"> Varies by district but based on specific drivers of instability in a given district | X* | X* | X* | X* | X* |
| FATA Stability Index, Pakistan | <ul style="list-style-type: none"> Stability | <ul style="list-style-type: none"> Strengthened security Improved GOP's responsiveness and accountability; improved services Improved livelihoods and economic activity Strengthened civic and political participation Strengthened provision of justice Gender | X | X | X | | |
| Helmand, Afghanistan | <ul style="list-style-type: none"> Reduce support for the Taliban | <ul style="list-style-type: none"> Build capable security institutions Support responsive, accountable governance Improve essential services and economic opportunities | X | X | X | | |
| MISTI, Afghanistan | <ul style="list-style-type: none"> Stabilize | <ul style="list-style-type: none"> Support war-affected families Make infrastructure repairs and provide jobs Build trust between citizens and their government Improve district governance and service delivery that respond to citizens' needs and grievances | X | X | | | |
| RC-South, Afghanistan | <ul style="list-style-type: none"> Stabilize | <ul style="list-style-type: none"> Increase security Improve governance institutionalization Strengthen community resiliency Increase public confidence in non-insurgent institutions | X | X | X | | |
| YMEP, Yemen | <ul style="list-style-type: none"> Strengthen the GOY's ability to promote security and minimize the VE threat Mitigate Yemen's economic crisis and deficiencies in government capacity, provision of services, transparency, and adherence to the rule of law | <ul style="list-style-type: none"> Improve livelihoods in vulnerable communities Improve governance capacity to mitigate drivers of instability | X | X | | X | X |
| Counter Violent Extremism Programs | | | | | | | |
| Eastleigh Youth Program, Kenya | <ul style="list-style-type: none"> Foster the identity and confidence of youth to reject VE | <ul style="list-style-type: none"> Build community and youth capacity for moderation Empower local youth through cultural and social activities, counseling, mentoring, leadership training Support youth livelihoods | X | | | X | X |

| M&E System or Program | Objectives | Key Components or Conditions | Livelihoods | Governance | Security | Tolerance | Youth Engagement |
|--|---|--|-------------|------------|----------|-----------|------------------|
| Garissa Youth Program, Kenya | <ul style="list-style-type: none"> Primarily, improve livelihood, learning, and community engagement opportunities for youth in Garissa Secondarily, address “pull” factors of VE messaging | <ul style="list-style-type: none"> Increase engagement of youth in their communities Increase leadership capacity of youth Improve capacity of youth to pursue employment/livelihoods Improve education opportunities Develop capacity of youth serving organizations Develop civic education and radio programming and service opportunities linked to tolerance, conflict resolution, and community engagement | X | | | X | X |
| PDEV I, Niger and Chad | <ul style="list-style-type: none"> Mitigate the potential for terrorism and extremism in the Sahel | <ul style="list-style-type: none"> Improve community governance and reduce ungoverned and poorly governed space Empower at-risk youth to become active participants in their communities and the economy Discredit ideologies that promote violence | X | X | | X | X |
| PDEV II, Niger, Chad, and Burkina Faso | <ul style="list-style-type: none"> Decrease the risk of extremism in higher risk areas Mitigate grievances in marginalized communities | <ul style="list-style-type: none"> Empower at-risk youth to become active participants in their communities and the economy Discredit ideologies that promote violence Increase civil society capacity to address community issues Strengthen local government | X | X | | X | X |

X* Likely to include this component, but specific interventions vary by district.

In our sample, the Iraq CSP and YMEP experienced notable shifts during the life of the program. CSP dropped one of its IRs, which improved the logic of the approach, and shifted its targeting to include women. In response to a shift and reduction in Mission personnel, YMEP dropped its innovative focus on monitoring the impact of programming on stabilization and focused just on the impact on livelihoods and governance. This represented a shift from a COIN focus to a development focus (see the CSP and YMEP summaries for more detail).

PDEV II and also MISTI are undergoing significant shifts in framing objectives; changes in approach and indicators are likely to follow. PDEV II is shifting from a negative focus on CVE to a positive focus on community resiliency and social cohesion, with the logic that improving both will better enable communities to meet their own needs and also make them more resistant to ideologies and actors that do not in any case fit well with local traditions. This is partly (but not entirely) a matter of optics and strategic communications, and of being able to declare what the US stands for rather than what it stands against.³ This framing is more acceptable for host country populations. In order to inspire local ownership, project staff are describing PDEV II (as PDEV I staff did) as a program focused on peace, tolerance, inclusion, fairness, and non-violent ways of resolving conflict.

In a similar way, MISTI is considering a shift in its framing, but in response to coming changes in the security environment. With the withdrawal of ISAF troops according to a given schedule, security could deteriorate and local actors could reconsider the cost-benefit calculus of resistance to the Taliban and cooperation with the Afghan government and international donors. At the same time, development resources and government presence and services may diminish, requiring communities to do more to meet their own needs. This could

³ As noted by Robert Kevlihan, USAID/West Africa’s regional conflict advisor in August 2012 in a telephone interview with Lynn Carter.

make backsliding on stabilization unavoidable. USAID's Stabilization Unit and the MISTI program are now considering re-orienting program measurement to the concept of community resiliency, in which cohesive communities are able to work together to solve problems and resist malign actors. Programs already work to strengthen aspects of resilience through such efforts as strengthening local dispute resolution and problem solving. RC-South's M&E implementer, Caerus, has raised this concern, even though its model and measures go some way to highlight the importance of community resiliency.

Indicator Selection and Organization

Introduction

Overall, indicators developed for the 11 systems in the sample reviewed show steady improvement over time, from 2006 when the first programs in the sample started. Indicators (and just as critically, their operationalization) have in the last few years grown more sophisticated, compared to indicators developed for earlier programs. We can contrast the weak monitoring system developed for the costly \$675 million Community Stabilization Program (CSP) in Iraq, which ran from 2006-9, with the far more sophisticated efforts apparent in the emerging (but in some cases still under revision) PDEV II indices and the RC-South/Afghanistan, MISTI/Afghanistan, and FATA/Malakand Stability indices on the other hand. These latter monitoring systems have better elaborated frameworks and more highly developed indicators and approaches to measurement. In the CSP M&E plan, several problems are evident but two are worth noting here:

- The two Strategic Objective indicators did not accord well with the major thrusts of the program (one was later dropped).
- No attempt was made to measure whether target youth (the rationale for the program) were less inclined to become involved in insurgent activities or were less attracted to VE groups and ideologies. An assumption was made that the interventions were of the right kind to mitigate youth risk.

The older M&E plans in the sample are linked to projects and tended to be heavier on output indicators, which reflected not only with the state of CVE/COIN knowledge at the time but also the emphasis within USAID on F process indicators. Program staff for these older programs also had problems acquiring timely baseline data and maintaining data quality; there was little sense of urgency given the other challenges in programming. There was little recognition in the older programs in the sample of the need to invest in the kind of rigorous M&E systems that would generate useful knowledge for CVE/I programming. Also, there was little expertise available at the time to help in shaping indicators for CVE/I programs so USAID staff and IPs were largely working in isolation, trying to shape indicators on their own. They were also sometimes caught between programs articulated primarily as development programs (at least in the CVE cases) and the CVE rationale for the program, leaving a question about which of the two they were trying to measure.

There is more collaboration on CVE/I measurement now (witness the now-quarterly M&E summits sponsored by USAID/Afghanistan's Stabilization Unit and the MISTI team), and there is more research upon which to draw, but the field still has much to learn. Investments in M&E have increased and the move toward more complex, abstract, and meaningful concepts, such as youth empowerment; community outlook for the future; attitudes toward moderation, tolerance and non-violence; the provision of justice and dispute resolution; and the like continues. Indicator sets are more likely now to capture citizen behavior and perception as well as on-the-ground reality, and much more effort is being invested in trying to capture inherently difficult-to-measure concepts such as the provision of justice (a vital factor for stability and resiliency in Pakistan and Afghanistan) and the acceptability of corruption (how much is socially acceptable, how much de-legitimizes government officials?). So, for example, the concept of security, vital in COIN programming, might be measured now by assessing some combination of:

- Degree of territorial control by government forces
- Number and type of violent incidents
- Presence/activity of police/army (patrols, checkpoints) and citizen perceptions of the behavior of security forces (e.g., percent seeing them as a source of **in**security)
- Number of patients treated at health clinics suffering injuries from Afghan on Afghan violence
- Citizens and USAID IP views of their own freedom of movement
- School attendance as a sign of what parents think of security
- Trader perception of freedom of movement for commercial activity
- Citizen perception of overall security

This kind of triangulation for a single, albeit complex, concept provides a much richer and more accurate picture of reality and tells planners much more about progress, allowing programs to be adapted where needed. If security has actually improved, for example, but citizens do not feel more secure, they may not take advantage of their greater freedom of movement. Children will not be sent to school, economic activities may not resume. The misperception, if unaddressed, could then adversely affect elements of the stabilization strategy. Increasingly, M&E systems show that programmers understand they are dealing with complex, composite phenomena with many factors that must be taken into account.

Many of the concepts related to achieving stability, defeating insurgency, building community resistance, or mitigating violent extremism are inherently difficult to measure. It is a challenge to unpack and develop precise measures for a thick concept like *social cohesion*, which some believe is linked to resistance to violent ideologies/actors. First, it must be broken into its constituent parts, then indicators need to be developed for those parts, ways of measuring the indicators determined, gradations of change established, and methods of aggregation determined. Other thick concepts included in the sample include the degree of community (or individual) resiliency, level of youth empowerment, the amount and type of community participation, and the degree to which genuine justice is provided. Breaking apart these concepts into their constituent parts is key; those parts can then be converted into indicators that can be clearly defined and turned into questions on a survey or FGD guide, or into some other form of data collection. While progress in working with such rich concepts is clear, this review suggests that we are still at an experimental stage, but that many interesting attempts are underway. In Pakistan's FATA and Malakand regions, for example, USAID is experimenting with a combination of an expert panel and questions on a household survey asking where respondents go to resolve different kinds of disputes and rating various organizations involved in the provision of justice on a number of aspects including cost, speed, fairness, enforcement of decision, and the like. While labor intensive, the picture formed across these different kinds of data should yield far better information on the state of rule of law and justice in these regions than would an expert judgment alone or a few questions asking about perception of the rule of law on a household survey. In another example, in the PDEV II project in the Sahel, USAID is trying now to develop indices for social cohesion, resiliency, and youth outlook on the future as its major goal-level indicators.

With growing sophistication and complexity in the concepts and accompanying indicators come rising costs for monitoring and analysis. It is too early to gauge cost versus benefits and the uses to which data can and will be applied for many of these indicators but this will need to be assessed moving forward. If we look at the provision of justice annually, particularly for an environmental indicator that is not much addressed via programming, is there likely to be significant change year-on-year to justify the measurement success? Is such fine grained information useful or usable by some party with access to it? Some of this investment in measurement can be recouped to the extent that a successful attempt to measure justice in one place can be shared and adapted for other environments. Cost is perhaps less an issue in insurgency environments (where funding (at the moment) is substantial and planners must track local level changes with care) than it is in some of the CVE environments where resources are limited and must be stretched far. At this point we do not know how much measurement is enough, though there have been attempts to winnow the number of indicators in two cases in the sample, the RC-South and FATA/Malakand stability M&E approaches.

There is a question of the technical and management burden that sophisticated results and indicators place on IPs. It is not simply enough to develop indicators and Indicator Reference Sheets. Attention needs to be paid to how indicators are operationalized in the form of questions on a survey questionnaire, how answers on a questionnaire are combined, how different forms of data are scaled or aggregated (with what weighting, if any), and the like. Increasingly, M&E staff on field projects are CCNs and the skill set may vary widely between Afghanistan and Niger on the one hand and Indonesia and Lebanon on the other. Even home office IP M&E staff may lack the skills required for more sophisticated forms of monitoring and analysis. With USAID Forward, as well, as more awards and contracts are put in place with host country institutions, there may be not only limits to local technical capacity to put in place sophisticated M&E systems, but also sensitivities about what local organizations are comfortable measuring and analyzing. Issues of bias are also possible.

Despite signs that CVE/I measurement is being taken seriously, some weaknesses were found in some cases and are worth noting in this report. Indicators can appear solid and relevant to the result being measured on the surface, but both indicator definitions and how indicators are articulated in data collection instruments can substantially alter their meaning and utility. Examples of weaknesses that relate predominantly to **program outcome** measurement (rather than to broader context measures and indices) include:

1. Indicators that do not stretch sufficiently (i.e., are minimalist) or are not accompanied by higher level outcome-oriented indicators, such as:
 - Number of youth who complete leadership training or vocational training, but with no accompanying indicators tracking the use to which the training has been put and its effect on youth well-being.
 - Percent of target youth obtaining jobs defined as youth obtaining internships, when internships are generally short term and seldom convert into jobs.
 - Percent of youth improving their livelihoods defined as employment for six months in the 12 months following completion of training, or generating some income within one year of receipt of an employment “kit,” but with no follow-up after that.
 - Percent of youth completing leadership training that are active participants in the community, with “active” seemingly defined as youth voicing an opinion once at one event. This may represent a significant accomplishment in a culture where youth lack the standing to participate in community processes, but the use of the word “active” creates potential misunderstanding about program achievements and local dynamics.
2. Indicators that over-reach and are too ambitious for the program or do not fit the program direction, such as:
 - Percent of community members satisfied with access to services and resources, for an IR of “Increased access to quality services and resources”. The indicator fits the IR, of course, but the program did not address local government capacity and was instead a narrower effort that funded a limited number of participatory community development interventions. The program was not able to show progress on the indicator because the activities were insufficient to affect the sense of individuals in target communities that services and resources had improved all that much.
 - Perception of citizens on the effectiveness of local government in providing services for a program that was primarily supporting youth employment and engagement and was not strengthening municipal government per se.
3. Important results that go unmeasured due to M&E resource constraints, such as:
 - The effects (retention, belief, sharing, behavior change) of a program’s youth radio component, which reaches 600,000 youth.

4. Indicators that are weakly defined or not defined at all until it comes time to shape a survey questionnaire, at which point the survey question(s) may diverge substantially from the intent of those who specified the indicators:
 - When the indicator represents an abstract concept, such as the popular perception of the provision of justice, then the conversion of the indicator into one or more survey questions becomes critical to whether the indicator provides information on which programming decisions can safely be based or partial/inaccurate data upon which it would be risky to rely.
5. Data collection methods that do not appear to match the indicator, although these may simply be minor Performance Monitoring Plan mistakes that are caught later, when the need to acquire the data arises, such as:
 - Trying to elicit the views of religious leaders from a survey of youth and community members without over-sampling of religious leaders.
 - Percent of target youth saying their economic situation has improved when the survey question captures how they rate their situation on a five-point scale, and the percent reported is actually the percent who rate their economic situation as fair, good, or very good.
6. Higher level results that are not measured (e.g., the Garissa youth program does not try to measure youth empowerment, its major objective; PDEV I did not try to measure the overarching objective of mitigating terrorism and extremism in the Sahel, where the risk was in any case low).
7. Indicators at lower levels in the Results Framework are jumbled and a mix of output and outcome indicators combined.
8. Survey questions that are over-simplistic reductions of the indicator, such as:
 - Percent of citizens stating that violence is an effective way to resolve problems on a five-point scale from never to always with the indicator recording those who replied always, usually or sometimes. It is hardly surprising that most respondents found violence repugnant.
 - In some cases, the phenomenon being measured by the indicator may be so sensitive or so linked to a social norm that direct questions in surveys (and possibly in key informant interviews or FGDs) will draw a high proportion of “Don’t know” responses (a standard problem on surveys in Pakistan and Afghanistan) or will be strongly affected by social desirability bias, making the data unreliable. In some societies, for example, it might be socially preferable to indicate trust in others (whether one feels it or not) or to declare that tribes are united and it might well be safer not to admit that you would not go to the Taliban for help in solving problems or that you thought the Afghan National Army or Afghan National Police was the main force driving insecurity in your area. More indirect methods of questioning may be needed. Testing different formulations of how to ask delicate questions in a more thorough-going manner is one way to proceed.

These weaknesses are **not** particular to CVE/I monitoring systems but could be found in a random assortment of project- and program-monitoring frameworks drawn from different sectors. But they are mistakes that can increase data collection costs, expend scarce resources generating data of limited value, create confusion about changes that are happening, and fail to generate data needed to advance theories of change or adapt programming (for example, if those youth completing vocational training cannot obtain jobs or their small businesses fail with 12 months, are they left angrier than they were before the training, inadvertently producing a counter-productive CVE result?).

It is more difficult to draw conclusions about problems in the systems attempting to measure stability or stabilization progress, all of which are related to COIN environments, as these are still new and some are in flux. These systems will be compared later in this report.

In the two sections below, where we first address CVE and then COIN programs, we try to focus on the most common themes and issues across the sample. Those more interested in one kind of program may still find material of interest in the other section.

CVE Monitoring Systems

Four CVE M&E systems are reviewed here: those for PDEV I and II in the Sahel, and for the Garissa and Eastleigh youth programs in Kenya.

Monitoring the Environment or Context and Changing VE Risk

If we look at the “if-then” logic used by USAID to teach and test Results Frameworks, often measurement of the “then” portion is missing from these programs. Almost none of the programs reviewed formally tracked the larger environment or the changing level of VE risk via formal indicators and data collection processes. Sporadic attempts have been made (e.g., the PDEV I mid-term and endline surveys but with limited questions), but they seem thin and generally do not provide much usable data. A number do track the environment more informally and at varying intervals via staff, steering committees, media reports, periodic qualitative assessments, and the like. PDEV I staff (to the extent that they were in target communities, some of which were quite remote) kept an eye on the environment in those communities and reported changes in quarterly reports. OTI staff explained that as a matter of general approach, OTI uses quarterly reviews that assess program activities against the larger environment, and that staff shift objectives and activities in response to the changing needs of the environment.⁴

One important reason for limited formal data collection on the overall environment (and the rationale for these programs), apart from the burden imposed by the entire M&E system, is the fear of generating suspicion and thus hampering the ability of IPs to work effectively in the host country. If a program has been presented to the host country as a development program, active monitoring of VE activities, presence, and risks could generate distrust, particularly in environments where the US is already somewhat suspect if not actively disliked. As one implementer noted, at the start of the PDEV I project in the Sahel in 2006, it was clear that there were suspicions about US intentions and that in some communities the US was regarded as cruel and violent, leading individuals to view participation in project activities with caution. PDEV I staff were reluctant initially to admit to local staff and local IPs the rationale behind the project for fear that they would be unable to function. Even with the characterization of the project as a classic development project, it took time to overcome this distrust and persuade communities to participate in the project. This was true in the Garissa Youth Project as well; adults had many concerns that the Americans were trying to subvert Garissan youth and detach those youth from their traditional culture and religious values. This suggests then that IPs, in many environments, could be compromised if placed in the position of collecting data about the VE environment or even in some cases subcontracting directly for it from a third party research organization.

The USG might possess useful data on VE groups, activities, and risk levels, though such data are likely to be classified and not available to IPs. How frequent, substantial, and fine-grained these data might be is also open to question and likely to be highly variable. USAID staff should have access to these data and may be able to work with IPs in reshaping activities to fit a changing context.

Still, even where there is some informal tracking of the environment, there are some lingering questions about the rigor or consistency of assessments of changes in risk levels over time and how we learn what limits or

⁴ Interviews with 1) Stacia George and George Wilson of OTI/Washington; and 2) Galeeb Kachra of the Kenya Transition Initiative.

mitigates risk without more formally structured efforts. How do we really know whether our theories of change are correct, such that increased community resilience or social cohesion does in fact deter risk and make it harder for VE actors to take hold and advance their agenda? Do we learn enough in a systematic manner to make the right corrections and changes in our programs? It may be particularly challenging to keep up with environments in which there is considerable flux. If not keeping an eye on the context, how do you know you need to keep investing in a given locality or target group as opposed to some other locality or target group? OTI's programming is of course inherently more flexible than mainstream bilateral mission programming, affording it the opportunity to change objectives and not just activities over time. We do see some changes in other CVE programs; for example, in response to a VE risk assessment in Niger funded by the Africa Bureau in 2009, the PDEV I project shifted its emphasis some from Hausa to Tuareg communities and from rural to urban ones.

Another reason for limited formal monitoring is the cost of monitoring relative to the benefits of doing so.

An Innovation in Tracking the VE Environment

One interesting innovation to track the VE environment is OTI's use of Sensemaker in the Eastleigh youth program. An approach and software produced by Cognitive Edge (www.cognitive-edge.com), Sensemaker is currently helping OTI gain some understanding of whether youth are being recruited by VE groups like the Shabab, what that recruitment process looks like, and what the response of those approached is. Based on a sample of respondents telling a story about a particular subject, Sensemaker was first tested by asking 2,800 target youth about their needs, concerns, and aspirations. The results were used in planning activities and grants but were not a reliable guide to the VE environment or useful from a formal project monitoring standpoint. The second application, which took place in September 2012, asked 400 youth and 200 adults to tell a story about a family member or friend (if any) who was approached by a radical group about joining. The respondent then assumes the character of that person and answers questions related to the recruitment attempt and response. The story approach puts the incident at a remove from the respondent and so provides the respondent and the person about whom s/he is telling the story with anonymity. It is possible of course that the respondent knows 10 or 20 youth who have been approached by a VE group but the story will be about one person. If patterns emerge, this approach can provide useful qualitative data about recruitment pathways and approaches, kinds of responses, what characterizes vulnerability, and what characterizes resistance. The approach cannot quantify the amount of recruitment going on or the degree of success.

In lower risk environments (e.g., Burkina Faso, Chad), where development efforts are largely preventive, there is not much to uncover and informal, qualitative monitoring on an annual or semi-annual basis may be adequate. A decision to handle the monitoring in this way should be reassessed at intervals, if it is found that risk factors are beginning to escalate. When VE ideas and actors are not much present and have little traction, and where development assistance programs are largely preventive, it will also be difficult to determine whether the program itself is responsible for the risk remaining low or whether the risk would have remained low in any case – this is akin to the long-silent dog continuing not to bark.

Common and Emerging Results/Conditions and Indicators

In this section, we focus largely on more interesting indicators and efforts at measurement but also on a couple of common problems.

1. Social Cohesion and Community Resiliency

These concepts are cropping up more in programming and are as relevant to COIN programming as they are to CVE programs. Measurement is quite new and there is little certainty about how to capture such multi-faceted concepts with a limited number of indicators. PDEV II is in the early stages of developing indices for both social cohesion and community resiliency; this effort would be useful for USAID to track:

Goal Indicator 1: Social Cohesion Index (average score across target communities)

The index measures social cohesion in target communities. Social cohesion refers to the norms and networks that enable collective action. It encompasses institutions, relationships, and customs that shape the quality and quantity of a society's social interactions. The theory is that greater social cohesion leads to or produces greater resilience. As not all respondents will be PDEV beneficiaries, the index will capture spillover effects in communities.

The index will be divided into the following dimensions:

1. Groups and Networks - Key characteristics of formal groups to be measured include: density of membership, diversity of membership, extent of democratic functioning, extent of connections to other groups.
2. Trust - There are several types of trust to be measured; within established relationships and social networks; trust extended to strangers; trust in the institutions of governance.
3. Social Inclusion - Questions on this dimension are intended to find out who in the community is included in collective action, decision-making, and access to services.
4. Information and communication - Maintaining and enhancing social cohesion depends critically on the ability of the members of a community to communicate among each other, with other communities, and with members of their networks that live outside the community.

Specific questions that will be included in the baseline survey instrument to measure the social cohesion index will be developed in consultation with USAID and relevant stakeholders during baseline instrument development. A workshop process for instrument development is being considered.

The following two questions will be asked as part of the Social Cohesion index under dimension 3 (in keeping with PDEV I and Afrobarometer indicators): (1) Have you participated in local decision making processes in the last 12 months? (2) Do you think that you will take part in community activities that address community issues during the next 12 months?

Aspects from the current draft Niger baseline survey that could be included in this index include: the extent to which people are willing to work together, the extent to which they have gotten together to accomplish things, the extent to which the community is peaceful or violent, and the level of trust in different groups of people.

The World Bank's Social Capital (SOCAT) tool, found online, could be a useful resource for questions and approaches.⁵

Aggregation and weighting across survey questions and dimensions are pending issues for the index. The team is still contemplating how to express the index as an indicator but is currently considering the average score across target communities.

One issue of possible note with the concept of social cohesion is at which level of society social cohesion is most important. Most CVE programs seem focused on grassroots communities and neighborhoods and so the concept is more focused at that level. So if it is the community/neighborhood that needs to be more cohesive, is there also a concern about cohesion across surrounding communities and the larger society?

Social dilemma games (e.g., trust and trustworthiness game, public goods game) could be a more reliable way to investigate trust, the sense of reciprocity and the willingness to cooperate for the common good than surveys but would add considerably to the cost of monitoring. Still, it could be an interesting pilot for the West Africa regional mission's external M&E contractor to implement.⁶

Goal Indicator 2: Resiliency Index (average score across target communities)

The indicator is an index that provides a measure of target community resilience to violent extremism. The index is divided into two dimensions:

⁵ See

<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTSOCIALCAPITAL/0,,contentMDK:20193049~menuPK:418220~pagePK:148956~piPK:216618~theSitePK:401015,00.html>

⁶ See <http://www.socialdilemma.com/> for a basic explanation of these games.

- 1) Measure of vulnerability - measure of the factors (social, psychological, and physical) that make individuals vulnerable to participation in VE; and
- 2) Measure of resilience - measure of the factors (social, psychological, and physical) that enable vulnerable individuals to resist VE.

Some of the specific social, psychological, and physical aspects that will be examined include: education, financial stability, local partnerships, community leadership, external violence, community violence, attitudes to violence and tolerance, knowledge of different religions. Specific questions that will be included in the baseline instrument to measure the resilience to violent extremism index will be developed in consultation with USAID and relevant stakeholders during baseline instrument development.

There is much work to be done yet to flesh this index out. One element of uncertainty is whether this captures community or individual level resiliency, as the explanation appears to highlight both; the indicator is currently being rephrased to capture individual-level resiliency. The two kinds of resilience may be intertwined but are perhaps not identical. As noted on the summary form in Section II. Of this report, the team is now leaning toward gearing this index toward individual level resiliency, leaving the social cohesion index to cover the community environment. There is some suggestion in the literature that community resiliency may be more significant as a protective factor than individual level resiliency because individuals generally operate within the norms of a given environment.

For both resiliency and cohesion, it will need to be shown that increases can be linked to diminished VE risk. While the argument that fragile and fragmented communities are more at risk seems sensible, can we be as clear that cohesive, homogeneous communities, operating in certain environments where VE ideologies have appeal, are not?

2. Youth Empowerment

Youth empowerment is a common theme in CVE programs. Generally these programs include some mix of the following types of activities: education (formal and/or non-formal and efforts with religious schools), livelihoods, civic education and participation (including service learning), access to information, radio programming (including on messages of peace and tolerance), and social and cultural activities. Programs then grapple with measuring these different aspects and sometimes with the larger concept of empowerment. It should be noted that the USAID guide *Measuring Progress in Development Assistance Programs Countering Violent Extremism: an Introduction* deals extensively with youth programming and suggests indicators, some of which overlap with ones we see in these M&E systems.

2.1. Indicators for empowerment

This is a dense concept that may be defined differently in different programs but would generally include an ability to earn a living; the skills, knowledge and opportunity to participate in community decision making and activities; a positive outlook on the future or at least a sense of moving forward and having some ability to shape one's future; the ability to protect one's rights and interests through non-violent means; and a positive social network to provide support. Critical to empowerment might be the belief that you can take some action to bring about a decent future for yourself. This complexity suggests that empowerment cannot easily be measured through one or two indicators without losing important dimensions.

PDEV II in the Sahel is currently trying to assemble an index for "youth outlook in target communities," defined as their individual and collective vision of their future, which is being used to measure the Goal statement of Increased Resiliency against VE. This comes very close to the concept of empowerment. They intend to measure dimensions related to: economic outlook (attitude toward and practical vision of future career and economic potential); social network participation (actual participation and efficacy of participation); conflict outlook (attitudes toward existing and potential conflicts in their societies); and education outlook (attitudes toward and expectations regarding education). They are examining their

preliminary set of indicators from their Results Framework, developed before this concept was articulated, for fit. This is an interesting attempt and is worth tracking for its results and the associated analysis over time. One might question how essential “education outlook” is for youth 15-30, but remaining in school a year or two longer keeps youth occupied and could (though this may be questionable in these environments) improve their livelihood opportunities. The conflict outlook is an interesting addition and reflects efforts to convey the values of tolerance and moderation and to attach youth and their communities more firmly to traditional culture, which is moderate in this part of the world. This addresses neatly the potential pull factor of VE ideologies.

PDEV I and II both included indicators assessing the extent to which adults believe youth make a positive contribution to society. Seeking the views of those who might block youth participation and whose attitudes may need to be addressed as part of programming makes sense.

The Eastleigh youth program assesses the percent of youth possessing a strong sense of identity, the percent envisioning and working toward a better future, and the percent saying that mentoring/counseling had helped them overcome personal problems. Identity refers to traditional Somali culture, which is a moderating force, and strong ties to the community. The first two indicators are stronger measures than the third.

2.2. Indicators for livelihood

In the sample reviewed, many measures for capturing livelihood improvements seem weak because they are either short-term or un-ambitious, as some of the examples provided in the introduction to this section on indicators suggest. Youth livelihood programs often include vocational training, remedial education, access to credit or livelihood “kits,” small business training, internships and apprenticeships, and service learning. Granted, improving livelihood in many of these environments for substantial numbers of youth poses huge problems, but currently we are learning little about the extent to which investments in vocational training and entrepreneurship produce even a medium term return. What portion of assisted youth obtains and keeps a job? Is the job at or well below the level of expectation? What portion creates a small business, derives an acceptable income from it, and manages to sustain the business? What portion obtains credit, invests productively, and pays back the loans? Without answering questions like these, how do we know that we have not created expectations that remained unfulfilled, leaving youth more aggrieved than they were before? And how do we know what changes need to be made in livelihood programming in order to secure a better return? Some cohort of target youth need to be followed not just for the year after training is completed but across the project (and preferably into the future) to advance our learning. If indicators are weak and M&E resources scant, tracer studies could be put in place at the end of the project, as the Garissa Youth Project might do, to learn what has happened to youth beneficiaries subsequent to assistance. However, then good personal contact records need to be maintained, which is not something all projects in the sample managed to do. One difficulty with such tracking over time is that young adulthood is a period of time when life is in flux, and youth may migrate to areas with better employment opportunities.

PDEV I did capture on surveys the percent of youth rating their economic situation as fair, good, and very good on a five point scale, but an astonishing portion in Niger and Chad fell into the group that found their economic situation at least fair. Given objective economic conditions in the target groups/communities, the responses were hardly credible. One of the issues that crops up with scales is the definition of points on the scale, what respondents understood by “fair,” and how well trained the enumerators were to explain the differences. PDEV I staff also thought that cultural factors came into play and speculated that the heavy selection of “fair” as the response was because respondents felt that they should be content with what God had given them and should not show dissatisfaction. More specific questions can be better: possession of durables and assets, food intake, a roof over one’s head and type of house, etc. There are questions on the Afrobarometer and poverty surveys for these phenomena. But it might be easiest to match the indicators to the assistance in terms of jobs, income, small business activity, and the like.

PDEV I also assessed employer (i.e., those who accepted interns and apprentices) satisfaction with skills of vocational training graduates, an indicator that could provide useful feedback on technical and soft workplace readiness skills of vocational training graduates.

One aspect of the theory of change with respect to youth programs is that youth are less likely to participate in VE or anti-state activities if they have a job or decent means of livelihood. Youth programs are now more holistic than they once were and it is understood that a range of interventions is needed to strengthen youth resiliency and resistance to VE actors, and that employment by itself (particularly if part-time/intermittent or beneath the perceived skill level of the individual) may do little. Livelihood by itself is a difficult problem to solve in many of these environments such as in FATA and the Sahel, and we do not know how much vulnerability is reduced, if other interventions succeed but we fall short on livelihoods.

2.3. Indicators for participation and engagement in community life

How the indicators are constructed and targets set should depend on the age group targeted. A 15-year-old does not have the same standing to participate in local decision making that a 30-year-old might have. Useful indicators identified in the sample include:

- Percent of youth reporting an improved relationship with the community (G-Youth)
- Percent of youth stating their opinions are respected by community leaders (PDEV I & II)
- Percent of community leaders saying that youth participating constructively in community decision making (PDEV I)
- Percent of youth participating in civic activities (by type of activity) (PDEV I)
- Percent of youth stating positive changes in their communities as a result of their participation in civic activities (PDEV I & II)
- Number of community service hours donated by youth (it would also be useful to know the number of youth involved in donating). (G-Youth)

PDEV I had a good series of questions in its endline survey about participation that could be disaggregated for data from youth respondents (see the section below on community engagement/development). The various Afrobarometer surveys also have good participation questions.

3. Moderation and Non-Violence

Both the Eastleigh and PDEV I programs have tried to monitor attitudes toward the use of violence. PDEV I asked a general question, as noted above, and learned that not many people found violence an acceptable way of solving problems. The draft PDEV II questionnaire included this general question plus two others asking respondents to rate the major divisions in the community (ethnic, class, religious, or political beliefs) on a five-point scale from very small to a great extent and then asking if the differences had ever led to violence (yes/no with no time frame attached). These more specific questions could be followed by questions that probe whether violence was seen as an appropriate response to dealing with conflicts growing out of major divisions in the community. Still it is not clear what time series data over a five-year time frame might look like for such questions, especially in environments where the use of violence is low.

The indicator could perhaps be useful in some environments but it needs to be framed in specific situations – under what circumstances would violence be an acceptable tool for change? On the Eastleigh program, recognizing that a generic question about violence would yield little, the question (posed in FGDs) was framed in terms of the 2007 Kenya post-election violence and whether discussants would like to see a repeat of that violence. The effort to be specific is good, but the example is flawed. Émigré Somali youth might not relate strongly to a Kenyan event or even to elections, and while FGDs allow rich discussion, the topic seems almost designed to evoke a negative response, given how awful the post-election violence was. The question could be a good one for Kenyan youth, if phrased in terms of responses to the upcoming election, if it too

appears stolen. PDEV II may drop its survey questions on violence. More investment and experimentation would be needed to turn this concept into usable indicators or a measurement approach.

A more fruitful direction however could be to explore the flip side of violence by assessing support for moderation and tolerance. PDEV I had a number of useful measures for capturing particularly the results of its radio programming:

- Percent of citizens listening to radio programs on peace and tolerance or number of listeners of moderate radio programs
- Percent listening regularly
- Percent finding moderate voices attractive or appealing
- Percent sharing moderate messages
- Number of hours of CVE programming broadcast or rebroadcast
- Number of assisted radio stations producing own VE content
- Perception of trustworthiness of the different programs (PDEV endline survey only).

Asking trust questions and tolerance of those different from oneself could be a useful approach to the concept.

Other indicators in use include:

- Number of respected community leaders addressing VE quarterly (Eastleigh); # moderate leaders delivering more effective VE messages (PDEV I)
- Number of intra and inter-faith dialogues and activities; percent of participants saying they would be willing to engage in further dialogue (PDEV I)
- Percent youth reporting less frustration with/more satisfaction with government agencies (Eastleigh)

4. Community participation and development/strengthened local government

Common elements include community development programs with broad-based community participation around a development intervention (e.g., borehole, school, economic project etc.), enhancing community decision-making processes, including the marginalized, strengthening the capacity and resources of local governments, and improved access to information.

There are other sources available for local government indicators so we do not go into detail here. Some of the better measures found in the sample include:

- Percent participating in community decision making processes (PDEV I)
 - Several interesting survey questions were asked: Do you participate in decision making in your community? If yes, how often and why? If not, why not? If yes, by type of activity. If yes, are your opinions respected by community leaders? If yes, did positive changes result and of what kind? If yes, did you encounter any negative consequences from participation and of what kind? This is a rich stream of data that could be converted into a participation index. Even if not used in a formal indicator, it provides useful information for tailored programming.
- Percent expressing satisfaction with community decision making processes (PDEV I & II)

While the number of community development projects that are completed is of course captured, we have not seen indicators in the sample related to successful operation and maintenance of these projects over time (sustainability; e.g., the school is maintained, the well still produces water).

COIN Monitoring Systems

Seven COIN M&E systems were reviewed: the RC-South M&E system, two DSF/Afghanistan applications in the East, the Helmand M&E program (HMEP), the FATA/Malakand stability monitoring system, MISTP's M&E system, the Iraq Community Stabilization program (CSP) monitoring system, and the Yemen M&E program (YMEP). The following table compares the indicators used by the first five of these systems to measure the stability environment; in at least three cases (HMEP, FATA, MISTI), these indicators represent a fraction of what the M&E systems are measuring and in two (RC-South, HMEP), significant additional qualitative research (beyond what is required for the indicators) is conducted to understand the environment better. Two (MISTI and FATA/Malakand) are accompanied by significant formal program evaluation components. All five are operating in the very similar environments of Afghanistan and FATA and Malakand Division in Pakistan and thus unsurprisingly show similar features. These five are meta-systems meant to capture overall progress toward stability and to provide frequent guidance for planning. Two of the systems are new (the FATA/Malakand stability approach and MISTI) and are still working out and testing data collection and aggregation methods. USAID/Yemen's YMEP is not included in the table because political vicissitudes in that country interrupted the effort to develop a district-level "stability outlook" index, leaving the system at a very nascent stage with project measures linked to employment, service provision, local governance, and community participation. YMEP seemed to be moving in the direction of a DSF-type approach, though it might have been one tied to USAID program results. The Iraq CSP is also omitted from the table; its own stabilization indicator, number of insurgent incidents, seems like over-reach given that it was a formal S.O. indicator (thus suggesting within the program's manageable interest) and any number of variables external to the youth program might have impacted on violence. CSP's M&E system overall was weak even for the program it was implementing, and it seemingly undertook little effective effort to track the stability environment, a more common issue with CVE programs.

Readers should note that in this limited exercise we were unable to assess all the individual projects and project-level M&E data that fed into some of these meta-systems or are used in conjunction with them.

[Note that the table that follows needs to be printed on 11" by 17" paper to be read in hard copy.]

STABILIZATION/COIN INDICES

(Measures of the stability of the environment)

| Program → | RC-South Afghanistan | | DSF/ASI-East Afghanistan | | FATA/Malakand Pakistan | | Helmand Afghanistan | | MISTI Afghanistan | |
|--------------|---|--|--|---|---|---|---|--|--|-------------|
| Sector ↓ | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source |
| Security | <ul style="list-style-type: none"> • Survivability of Government Officials and Community Leaders, measured by the frequency and lethality of attacks on officials and leaders • Degradation of the Insurgency, measured by effectiveness of the insurgency, capacity of the insurgency, and will of the insurgency • Function and Effectiveness of Government Security Institutions, measured by popular perception of gov't security forces and institutions, and ANSF ability to defeat the insurgency • Functioning Community Security, measured by popular perception of local security | <ul style="list-style-type: none"> • SIGACTS/ Assassinations/ Night Letters; key leader interviews (GIRoA officials; USG officials; RC/RPs; Embassy and USAID; Shuras/Jirgas) • Key leader interviews (GIRoA officials; USG officials; RC/RPs; Embassy and USAID; Shuras/Jirgas) • Perception surveys; key leader interviews (GIRoA officials; USG officials; RC/RPs; Embassy and USAID; Shuras/Jirgas) • Perception surveys | <ul style="list-style-type: none"> • Civilian Security, measured by the percentage of people reporting: attacks against civilians in the past six months; fear of such attacks for self and family; perpetrators of the attacks (Taliban, gov't security forces, criminals, other civilians, security companies) • Perceptions of Security, measured through percentages of people reporting: family is safe or very safe in village; district has become safer in the last three months; gov't forces always effectively protect community; no sources of insecurity in district. • ANSF Presence, measured by percentages of people reporting ANA & ANP patrols at village level; checkpoints outside their district center, ANA & ANP provided security in their district; ANA & ANP effectively protected the community; ANSF was a source of insecurity (accounted for negatively) • Freedom of Movement, measured by percentages of people reporting never feeling safe leaving the house, village, and district; encountering anti-gov't elements on the road; "always" encountering trouble on the road; "never" leaving the district"; "always feeling safe leaving their house, village, and district (accounted for negatively) | <ul style="list-style-type: none"> • Household survey • Household survey • Household survey • Household survey; interviews with drivers | <ul style="list-style-type: none"> • Freedom of Movement, measured by reality and perceptions • Incidents of Violence, measured by type, location, and number by district • Territory Control/Presence, measured by presence of security forces and/or miscreant presence | <ul style="list-style-type: none"> • Monthly online surveys of implementing partner staff • Daily media monitoring of radio and newspapers • Monthly online survey of implementing partner staff | <ul style="list-style-type: none"> • Security, measured by perceptions of security • Police, measured by perceptions of police acting in the interest of the local people | <ul style="list-style-type: none"> • Head of household survey • Head of household survey | <ul style="list-style-type: none"> • Perceptions of Stability, measured by percent of Afghans reporting that their environment has become more stable • MISTI Stabilization Trends and Impact Evaluation Survey | |

| Program → | RC-South Afghanistan | | DSF/ASI-East Afghanistan | | FATA/Malakand Pakistan | | Helmand Afghanistan | | MISTI Afghanistan | |
|-------------------|--|--|--|---|---|---|--|--|--|--|
| Sector ↓ | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source |
| Governance | <ul style="list-style-type: none"> • Resiliency of Government Institutions, measured through officials' tenure and quality, the depth of government, and budgetary process • Public Confidence in Government Legitimacy and Effectiveness, measured through popular perceptions of government and popular usage of District Center • Predictability and Acceptability of Corruption, measured through popular perception of corruption | <ul style="list-style-type: none"> • Behavioral observation; perception surveys; key leader interviews (GIROA officials; USG officials; RC/RPs; Embassy and USAID; Shuras/Jirgas) • Perception surveys; behavioral observation • Perception surveys | <ul style="list-style-type: none"> • District Government Recognition, measured by the percentage of respondents who believe that the district government represents one's interest either always or most of the time; use of district government for justice services; percentage of the population visiting the district center • Perceptions of Governance, measured by the percentage of people: reporting positive trust in the district, provincial, and national level governments; and that they use the district government for justice services | <ul style="list-style-type: none"> • Household survey; Interviews of visitors to District Center • Household survey | <ul style="list-style-type: none"> • Basic Services, measured by availability and satisfaction • Civic and Political Participation, measured by citizen participation in civic and political events (political meetings, rallies and processions, protests, traditional jirgas, and official jirgas), number of civic events, and incidences of collective action | <ul style="list-style-type: none"> • Annual household survey; semi-annual focus groups • Daily media monitoring | <ul style="list-style-type: none"> • District Government, measured by perceptions that District Government listened and acted on behalf of the people rather than in their own interest • GIROA, measured by perceptions that GIROA listened and acted on behalf of the people rather than in their own interest • Taliban, measured by perceptions that it would not be a good thing for the people and the country if the Taliban were to return to power and govern Afghanistan | <ul style="list-style-type: none"> • Head of household survey • Head of household survey • Head of household survey | <ul style="list-style-type: none"> • Government Responsiveness, measured by percent of Afghans reporting that GIROA listens to their grievances • Government Visibility, measured by number of visits and community engagements by provincial and district GIROA officials • Service Delivery, measured by percent of targeted communities reporting availability of GIROA-delivered basic services • Government Capacity and Strength, measured by percent of tashkil (civil service positions) fill in targeted districts • Government Development Initiatives, measured by the percent of beneficiaries reporting GIROA development efforts • Anti-Government Sentiments, measured by percent of Afghans sympathizing with motives for anti-government violence | <ul style="list-style-type: none"> • MISTI Stabilization Trends and Impact Evaluation Survey • Review of sub-national GIROA reports/ travel logs, GIROA official survey • MISTI Stabilization Trends and Impact Evaluation Survey • DDP weekly reports • MISTI Stabilization Trends and Impact Evaluation Survey • MISTI Stabilization Trends and Impact Evaluation Survey |

| Program → | RC-South Afghanistan | | DSF/ASI-East Afghanistan | | FATA/Malakand Pakistan | | Helmand Afghanistan | | MISTI Afghanistan | |
|--|--|---|---|---|--|--|------------------------|-------------|--|---|
| Sector ↓ | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source |
| Economy | <ul style="list-style-type: none"> Resilience of Subsistence Economy, measured through sustainability and stability of subsistence economy | <ul style="list-style-type: none"> Key leader interviews (GIROA officials; USG officials; RC/RPs; Embassy and USAID; Shuras/Jirgas) | <ul style="list-style-type: none"> Market Activity, measured by the percentage of shopkeepers reporting improving business, and the percentage of shops that are newly opened in the district | <ul style="list-style-type: none"> Small scale survey of business owners | <ul style="list-style-type: none"> Livelihoods, measured by employment, income, resources and assets Market (Bazaar) Activity, measured by the number and types of stalls, the types of products available | <ul style="list-style-type: none"> Pakistan Poverty Scorecard, conducted by the Benazir Income Support Program (annual?); household survey; focus group discussions Field observations | | | | |
| Community Resilience/ Social Structures | <ul style="list-style-type: none"> Strength of Shura and Councils, measured through the regularity and predictability of their meetings, and perceptions of effective conclusions, representation, ability to export stability beyond local environs, and resiliency to INS attacks and intimidation Civilian Freedom of Movement, measured by population that has freedom of movement that is/is not limited in either time or space and is/is not dependent on security force presence | <ul style="list-style-type: none"> Behavioral observation Regression analysis on highway traffic patterns (satellite feeds); perception surveys; behavioral observation | | | <ul style="list-style-type: none"> Social Gatherings, measured through media reports IDPs, measured by numbers, flows, and location Status of Women, measured through male and female perceptions of women's well-being, security, livelihoods, inclusion, and social cohesion | <ul style="list-style-type: none"> Media reports OCHA and GoP disaster management agencies Semi-annual focus group discussions with men and women | | | <ul style="list-style-type: none"> Community Engagement, measured by percent of Afghans who participate in resolving community problems Perceptions of the Country's Future, measured by percent of Afghans reporting that the country is moving in the right direction Perceptions of Quality of Life, measured by percent of targeted communities reporting that their lives have changed for the better | <ul style="list-style-type: none"> MISTI Stabilization Trends and Impact Evaluation Survey MISTI Stabilization Trends and Impact Evaluation Survey MISTI Stabilization Trends and Impact Evaluation Survey |

| Program → | RC-South Afghanistan | | DSF/ASI-East Afghanistan | | FATA/Malakand Pakistan | | Helmand Afghanistan | | MISTI Afghanistan | |
|----------------------------------|---|--|-----------------------------|-------------|--|---|---|--|----------------------|-------------|
| Sector ↓ | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source | Indicator | Data Source |
| Rule of Law / Justice | <ul style="list-style-type: none"> Function and Effectiveness of Rule of Law, measured by the reliability and accessibility of rule of law (traditional and formal), and popular perceptions of the rule of law | <ul style="list-style-type: none"> Perception surveys | | | <ul style="list-style-type: none"> Justice, measured by what exists and perception of how it works | <ul style="list-style-type: none"> Justice as measured through a household survey of behavior (where disputes are taken) and rating of justice providers on key criteria like fairness; expert panel on justice and dispute resolution | <ul style="list-style-type: none"> Justice, measured by perceptions of satisfaction with statutory justice services | <ul style="list-style-type: none"> Head of household survey | | |

Conscious efforts have been made in some of these meta systems to simplify concepts and indicators due to the need to weed out potentially extraneous information that could send confusing signals and to mitigate the risk of users being overwhelmed by both the need to collect, analyze, review and apply voluminous data in high pressured, rapid change environments where decisions must be made quickly. Also as HMEP staff note, with frequent staff rotations and some personnel unaccustomed to interpreting and using social science data to make decisions, the risks of trying to present too much data are clear. One reason that a clear theory of change or model for stability is useful is that it reduces the universe of things that need to be examined. The DSF, as a general approach for programming (as opposed to the applications examined in Afghanistan) does not specify a model or theory of change, and as such has as voracious demand for data and analysis, at least in the initial stage when drivers of instability are first being determined. This approach can be compared to those approaches that have a clearer theory of change such as for RC-South, where the data needs have been substantially trimmed. The DSF has been narrowed for the specific environment in the East in the formulations presented here.

Unlike the other four systems compared in the table, MISTI uses a Results Framework as the scaffold for its indicators. It has also taken ten of its most critical indicators and placed them in a stability index (called the maturity model) in order to provide a coherent picture of changes in the level of stability, with five levels: normal, stable, fluid, unstable and insecure. HMEP and DSF/ASI-East have done something similar, and the FATA/Malakand stability index started with a strong mandate to keep it simple. While the concepts and indicators across the five highlighted stability measurement systems in the table line up reasonably well, readers should note that gaps in the table do not mean that the given system does not have indicators to capture those phenomena, only that the system has made decisions to highlight some indicators as more important for gauging stability than others. For example, HMEP has extensive service delivery and local economy indicators but has not selected these indicators as major ones illustrating changes in stability. In addition, a concept measured in one category on one index may be placed in another category on a different index; thus, freedom of movement is featured under security in the FATA/Malakand and DSF/ASI-East approaches but under community resiliency on the RC-South index.

Even where there have been attempts to simplify concepts, indicators and data collection, it should be noted that the meta systems are still not small systems in any sense, even though some of them have become more discriminating and manageable over time. The indicators are often complex with sub-components and some of the systems use both qualitative and quantitative data to feed into a scaled approach. For example, the RC-South system rates the presence of Taliban courts in an area on a scale of 1 to 5, with 5 meaning no presence and 1 meaning active throughout the district but substantial data collection and judgment feed into making that rating.

For the most part, for the sake of page length and because of the orientation of most of the meta systems to measuring the environment, we limit our comparison to the higher level stability measures included in the table. These systems are often measuring similar phenomena though not always in the same manner or with the same emphasis but there are differences in which aspects and measures they choose to highlight as their main stability “index” or cluster. The system templates provided in Section II include indicators that are used but not highlighted as key conditions for stability; some of these relate more directly to project efforts.

Perhaps because the RC-South monitoring efforts serve military needs, it places greater emphasis on the extent of Taliban control and governance, popular perception of the Taliban, and security. The other systems also look at the Taliban role but to a lesser extent. Examples include:

- Percent of heads of households who do not think it would be a good thing for the people and the country if the Taliban were to return to power and govern Afghanistan (HMEP)
- Satisfaction with the Taliban (versus the government) role in delivering security (HMEP)
- Percent of Afghans sympathizing with the motives for anti-government violence (MISTI)

- % of Afghans thinking it would be good/bad if the Taliban returned to power (DSF/RC-East)

1. Security

Security is naturally a major concern in these measurement systems. RC-South and DSF/RC-East are the only systems that appear to look at the “survivability” of local leaders and government officials (in the case of RC-South) and attacks on GIROA officials (RC-East), a major problem across the target areas given insurgent group assassinations. All systems look at the popular perception of security; HMEP, DSF/ASI-East and MISTI appear to rely primarily on perception.

- Three systems (DSF/ASI-East, RC-South,⁷ FATA/Malakand) formally measure freedom of movement and not just through household survey questions on perception of security but through analysis of highway traffic patterns; individual interviews with business owners; and on-line surveys with implementing partners traveling in and out of target areas.
- Two (DSF, FATA) highlight territorial control by government forces or ANSF presence; one has several indicators related to both degradation of the insurgency, community security and security redundancy (multiple forms of security provision) (RC-South). The FATA/Malakand Index makes a distinction between which forces control the area at night and which during the day.
- Two count incidents of violence⁸ and attacks against civilians (FATA/Malakand, DSF).
- Three count citizen perceptions of government security forces and police (which would be related to legitimacy of the government and confidence that insurgents can be defeated as well as to whether local people might be willing to share information with the police) (HMEP, RC-South, DSF). The ASI-East stability index has a number of measures of perception of the ANA and ANP.

RC-South and both uses of the DSF in Afghanistan have more indicators focused on aspects of security than the other systems. Four of seven categories in the DSF/ASI-East stability index relate to aspects of security. Most of these systems, despite somewhat different emphases and approaches, should derive roughly similar pictures of security, with the possible exception of those relying primarily on respondent perception (which could suffer from a time lag between changing reality and perception of that change as well as be at odds with reality). The team implementing the first household survey (November 2012) in connection with the FATA/Malakand Stability Index found that because data collectors could not travel to very insecure locations, the resulting survey findings likely over-estimated how secure residents felt and may have provided a biased (if more reassuring) picture of progress.

2. Governance

Governance is another element that looms large in stabilization monitoring. The theory underpinning this is that the government, often viewed by its citizens as feckless, corrupt and abusive, needs to demonstrate responsiveness and some level of accountability to citizens to gain legitimacy; legitimacy is critical to undermining support for insurgents. Some systems look at the threats in trying to provide governance in insecure environments. These systems measure perception of government performance variably:

- The sense that government listens to and acts in the people’s interests rather than the interests of office bearers (HMEP; MISTI assesses the extent to which citizens feel the government listens to their grievances).

Measuring Freedom of Movement FATA/Malakand Index

Data, drawn from an annual household survey and a monthly on-line survey of IPs, capture:

- The extent of freedom of movement on a four-point scale from inaccessible to fully accessible
- Which individuals have freedom of movement in the area: people of the area, people from a neighboring area, people from a distant area
- Reasons that freedom of movement is constrained (e.g., bad weather, army operations, etc.)

⁷ RC-South captures freedom of movement under community resiliency but the concept is the same.

⁸ Data on violent incidents could be difficult to interpret given that those using violence might eventually create such a fearful population that the need to deploy violence is much reduced.

- Perception that the district government represents one’s interests all or most of the time. (DSF/ASI-East)
- Perception and experience of corruption (and interestingly acceptability of corruption) (RC-South)
 - HMEP also assesses perception that corruption is a serious problem and that it is worsening outside the context of their narrower set of stability indicators
- Availability and satisfaction with basic services (MISTI, FATA/Malakand, RC-East)
 - Outside the context of its narrower index, as noted earlier, HMEP has an extensive set of service delivery indicators related to actual use of services (e.g., enrollment in schools) and satisfaction with them. MISTI also looks at availability of services and government-sponsored development efforts as well as civil service positions filled.
- Political and civic events and collective citizen action (FATA/Malakand)
- Popular recognition of district government through, citizens participating in district council meetings (DSF/RC-East); “use” or visits to district offices for services/help and government official visits to and engagement with communities (DSF, RC-South) (USAID/Pakistan/OTI has been focused in Pakistan’s FATA region on increasing interaction between government officials and citizens and has customarily monitored interaction via key informant interviews, FGDs, and rapid appraisal methods)
- The tenure and quality of government officials. Whether civil service positions are filled, government officials are visible and interact with citizens. (RC-South, HMEP, MISTI)
- Legitimacy and confidence in government are measured in some of the approaches and derivative in others (more a function of analysis) (e.g., RC-South versus FATA/Malakand)
- Two (RC-South and DSF/RC-East) measures perception of the desirability of a Taliban return to power
- MISTI assesses the percentage of Afghans sympathizing with motives for anti-government violence, which may not mean they also want the Taliban back in power but is related to central government legitimacy. HMEP captures perception that it would not be a good thing for the people and the country if the Taliban were to return to power to govern Afghanistan.

These systems are at least partially reliant on household surveys, which generally ask for experience and behaviors as well as opinion, but most go beyond this to behavior observation, review of government travel logs, and the like. The combination of indicators and the different kinds of data included in most of these systems should provide an adequate picture of the state of governance in the locality.

3. Revitalization of the Economy

Revitalization of the economy is a key element in three of the systems reviewed (FATA/Malakand, RC-South, DSF). One, RC-South, looks at resiliency of the subsistence economy via key informant interviews and does a scaled rating on a five-point scale:

| RC-South: Resiliency of Subsistence Economy <i>measuring distance from endemic subsistence sustainability</i> | |
|---|--|
| 5 | Economy thrives beyond subsistence, enabling trade for profit |
| 4 | Subsistence economy is sustainable and resilient |
| 3 | Subsistence economy is sustainable, but not resilient |
| 2 | Elements of a subsistence economy exist that are neither sustainable nor resilient |
| 1 | Elements of a functional, sustainable subsistence economy do not exist |

Two (FATA/Malakand and DSF/ASI-East) focus on market activity to gauge economic vitality; one (DSF/RC-East looks at the number of commercial districts using the main market road on a given day each week, number of non-agricultural businesses added along the main market road, and number of hectares irrigated and planted in perennial/annual crops (DSF/RC-East); and one (FATA/Malakand) draws on a well-tested 10-question Poverty Scorecard created a decade ago and used by the Benazir Income Support Program. The FATA/Malakand index will supplement Scorecard data with semi-annual FGDs on assets, resources, and income, and a survey question about whether respondents are better or worse off economically than they were previously. The FATA/Malakand survey also looked at how good a job respondents thought the government was doing in generating employment, bringing industry to the target area and building infrastructure. While HMEP does not include such indicators in its narrower stability index, it has measures in its overall system related to popular optimism about the economy, average household income and households above the poverty line.

4. Community Resiliency/Social Structures

Community resiliency/social structures are highlighted as a significant element in two of the approaches though relevant measures do appear under other categories in other systems. RC-South looks at the functioning of local shuras and councils, and under a category called the “resiliency of government institutions,” the survivability of community leaders, social cohesion in terms of village level heterogeneity, the present of Taliban courts, and freedom of movement. DSF/ASI-East looks at tribal cohesion, or the popular perception that the tribes were united (activities brought tribes together to work on common problems). The FATA/Malakand index captures IDP movements, social gatherings (a sign of normalcy), and the status of women; while MISTI considers participation in community decision making, the perception that the country is moving in the right direction, and perception that the quality of life has changed for the better (the latter two are more in the nature of aggregate measures that capture changes across the board). HMEP outside the context of its narrower stability index looks at the percent of Helmandis confident in future improvements, which also serves as a kind of over-arching measure of popular satisfaction. While optimism may be a sign of resilience, in very uncertain environments the most rational response would be that it is impossible to tell what the future might hold.

5. Justice and Dispute Resolution

Injustice is a powerful driver of grievance and conflict in Afghanistan and Pakistan. There is a multiplicity of

Capturing Justice in the FATA/Malakand Stability Index

Annual Survey:
Questions on:

- Satisfaction with aspects of justice (cost, speed, fairness, etc) on a four point scale with a don't know/refuse to comment category
- Favorability rating (same scale) for several justice linked institutions
- To which institutions they go for resolving specific kinds of conflicts and why

Annual Expert Panel:
Questions on:

- The institutions to which different kinds of legal problems/disputes are taken for resolution/justice.
- Why people seek out those institutions.
- What results and level of satisfaction they find.
- Extent and effect of justice reforms.
- Additional reforms needed.
- Whether people see any improvements or changes in how justice is delivered.

justice systems: modern, formal systems of courts and law; formal Shari’a law as sanctioned by the state; customary law and processes; the law as applied in the Frontier Crimes Regulation which governs FATA; and Shari’a law as applied by extremists. Measuring popular perceptions of these systems, preferences for and actual use of one or another of the systems for particular kinds of disputes; satisfaction with and use of the modern, formal system; extent of reach of any of the systems; functioning of these systems; and improvements in the licit systems presents a significant challenge. Aspects to be considered include: efficiency (speed of process and decision), fairness (bribery, bias), cost, and ability to execute judgments among others. Three of the stability indices (HMEP, FATA/Malakand, RC-South) flag justice as a major concern for stabilization and government legitimacy. All three measure popular perception and behavior (use and experience), and two consult with expert

opinion. The FATA/Malakand index's interesting attempt to grapple with justice sector measurement is discussed in the accompanying text box.

The DSF/RC-East also looks at dispute resolution via three indicators: the number of disputes resolved monthly by judges, the district governor and "legitimate" informal structures with formal system validation. It also does a qualitative rating on a 1-10 scale of the extent to which these institutions resolve disputes effectively for the local community. This approach appears to value quantity of cases over the quality/fairness of the decisions made and whether those decisions are enforced but it may be that a rising number of cases reflects a willingness to use one of these institutions (or at least a perception that no better alternative exists) and thus may serve as a weak proxy for fairness of judgments. This approach to counting cases is a much less expensive alternative to that of the FATA/Malakand Index's approach but it is not certain what it is communicating. Survey data do however provide alternative information on perceptions of dispute resolution institutions and processes.

Data Collection Approaches

Mixed Methods: The programs in our sample incorporate a mixed methods approach to data collection, which affords more rich and nuanced information for learning than reliance on any one method. All the programs include **objective, countable data** collected from direct observation, IPs, host country governments, USG military, international partners, or other sources. Much of the objective data measures outputs, such as number of leaders trained in moderation, but some measure outcomes, such as number of respected leaders addressing topic of violent extremism.

Household Surveys: All the programs in our sample also include **household survey data** with the exception of YMEP, which dropped its plan for doing so after USG evacuations from the country and the subsequent shift in programming focus. Eight of the programs conduct their own surveys; only RC-East and RC-South rely on outside surveys (specifically those conducted by the military). Outside surveys often lack adequate sampling or time series data, but using them lowers costs. PDEV II and Eastleigh are using treated and control groups in their survey, and MISTI is able to use quasi-experimental methods to gauge impact due to its large sample size, and may use randomized control trials (RCTs) if IPs are able to select villages based on RCT criteria. Since HMEP and FATA/Malakand are measuring the stabilization environment and stability conditions rather than the impact of activities, treated and control groups are not relevant.

Many of the programs using surveys experience problems with baselines and sampling. HMEP's ability to monitor progress is somewhat constrained by the lack of baseline data prior to 2009-10 when HMEP became operational. Baselines in PDEV I and II were delayed by a year partly due to struggling over results frameworks and indicators, and partly in the earlier program because of difficulties over the baseline survey. PDEV I also had difficulty with the local survey firm, which added greatly to the delay; it turned out that delay was the least of the problems as the resulting survey data were of poor quality and unreliable. Moreover, PDEV I also experienced inconsistent sampling: the baseline and endline surveys used different samples so could not be compared. The midterm evaluation had a limited sample size and handled some questions differently from both the baseline and endline, so it also could not be compared to these surveys either. Missing, delayed, or low-quality baselines and inconsistent sampling cut into the ability to measure performance and make determinations about the validity of theories of change.

In Afghanistan and Pakistan, household surveys end up biased toward men because of patriarchal norms within households, but in Garissa the bias is toward housewives, the unemployed, and students, who are at home during the main interview hours. There have been dramatic differences in responses between quarters for some questions on the Garissa community needs assessment (CNA), raising issues about the reliability of the data and sampling, though the IP believes that the results show more consistency in recent surveys.

In Afghanistan, the issue of over-surveying poses challenges on the reliability of the data. There is a high frequency of "don't know" responses on the military-funded household surveys, which is a probable

indication of survey fatigue or reluctance to respond to sensitive questions. Moreover, misdirection and manipulation may result from surveys due to social desirability bias, as Afghans familiar with the survey instrument have decided which answers it is safe or useful to provide. In the Tactical Conflict Assessment and Planning Framework (TCAPF) survey, for example, Afghans could surmise that if they said the national government should be responsible for fixing a problem, the district might get funds from the national government.

Poor security in Afghanistan also poses challenges on the quality of the data. Under HMEP, for example, interviewers may only go to cleared areas and grey areas, but not to no-go areas held by the Taliban. The accessible households may well view the government and the security situation more favorably than those households in Taliban-held areas (as noted earlier with respect to the recent FATA/Malakand Stability Index survey, for which data collectors could not travel to the most insecure areas, leaving findings that were rosier than they might have been – at an aggregate level – had travel been possible). Interviewers in Helmand were also supposed to carry GPS systems to verify that they went to the specified locations and did not make up data, but due to perceived risks of carrying a GPS some interviewers refused to do so. Even hidden devices, placed in a shoe heel for example, are seen as creating too much personal risk. Statistical techniques can be used for uncovering significant fraudulent data and triangulation can help to shore up gaps in survey data.

Another challenge for data collection posed by high-threat environments is the sensitivity of some lines of questioning. It would be unreasonable to ask direct questions about the police in Afghanistan and expect to get reasonable responses. Asking indirect questions may provide answers that minimize any social desirability bias.

The survey instruments developed under several of the programs provide a good resource for future monitoring efforts. The endline survey for the impact of radio messaging in PDEV I, for example, contains nuanced questions on moderation and tolerance. In the COIN realm, MISTP's household survey contains questions about respondents' perception of their personal experience to gauge stabilization trends. For example, it asks how secure respondents feel in their home or when traveling or how they compare their quality of life at present to one to two years ago, and to their expectations for the future.

There is interest in M&E circles about using SMS surveys for quick feedback from target groups on program effects. One program in our sample includes use of **SMS surveys** following program events. Eastleigh asks participants about the quality and utility of an event and obtains a low response rate of about ten percent. This could present an opportunity to track recipients' progress throughout time. However, the resulting data could be biased as some youth seem unwilling to provide cell phone numbers due to suspicion about whether they can be tracked; this could particularly be the case for émigrés from Somalia. The Garissa youth program planned to use SMS surveys to track the number of youth employed through income generating opportunities and number of youth accessing business loans, but it found SMS surveys too expensive and inadequate for its purpose: the cell phone service providers would not provide the sending phone number so Garissa could not tell who was responding and from what location.

In an interesting innovation, one program is conducting **interviews that capture a story** to explore VE concepts. KTI Eastleigh did not find the first application, which explored youth issues and not VE, useful from a monitoring standpoint but found it gave good insights about youth concerns, priorities, and relations in the communities and gave them good ideas for grants (see more on this storytelling analysis in the Eastleigh template and in the indicators section). The utility of the follow on effort to look at VE recruiting in not known at this writing as data analysis is still underway.

The Garissa program has conducted **before-and-after surveys** with grant recipients and participants in leadership trainings and civic education sessions and hopes to conduct a follow-up tracer study of some of its beneficiaries in September 2012, because most of its indicators do not show impact (e.g., what happens after a youth finishes vocational training or obtains a loan).

FGDs: A small portion of our sample conducted **Focus Group Discussions**. Eastleigh explored attitudes toward violence in 16 FGDs with youth groups, both beneficiaries and non-beneficiaries, and with local grantees. DSF/ASI-East used FGDs in both its impact and overall stability assessments. HMEP found that FGDs with women did not work in Helmand, and replaced them with individual interviews. The FATA/Malakand index used FGDs to discuss basic services, livelihoods, and the status of women. It conducted 32 FGDs and for reasons of safety and greater freedom of speech brought 271 men and women (half women) to Peshawar to participate in the FGDs (see textbox). USAID’s Monitoring and Evaluation project used Skype on a laptop as an unobtrusive way to monitor to the FGDs and to pass along, via a staff person in the room, suggestions to the facilitators to follow up on specific topics.

PDEV II is considering FGDs after its household survey.

Key Informant Interviews, Expert Panels and Qualitative Rankings: In addition, two programs use **key informant interviews to inform stability indicators**. RC-South collects information on five of its 11 stability indicators from interviews with GIROA officials, shuras/jirgas, and USG officials. Similarly, DSF/ASI-East made use of interviews with government officials, customary leadership, local business people, and truckers in both its impact and overall stability assessments. Another few programs collect in-depth qualitative data through **interviews as needed**. This kind of research provides context to quantitative data, analyzes the political context of localities, and provides cultural awareness to inform monitoring efforts. RC-South, HMEP, DSF/ASI-East and Eastleigh use local researchers and interviews to probe dynamics, trends, and issues, and PDEV II plans to conduct VE qualitative research in target communities where the risk of VE is increasing. For example, HMEP found on surveys that the majority show trust in the Afghan National Police (ANP) as an institution and think that it is capable of resolving disputes fairly and efficiently, but that only a small fraction of respondents say they are willing to take a crime to the ANP if victimized. Follow-up research to explore this discrepancy showed a strong preference for family-based dispute resolution and that people did not find the ANP approachable. In Afghanistan, the use of local researchers has been critical as they can go where USG personnel cannot. Security protocols are in place for the local researchers and to date, no compromise in security has resulted from this approach.

There is little use of **expert opinion** in our sample. The FATA/Malakand Stability Index is using an online survey of IPs on freedom of movement (monthly ratings online) and a rule of law expert panel to get at the provision of justice. DSF/RC-East is using the judgment of district stability teams to monitor security, governance, economic, and rule of law conditions. When experts are linked to program implementation, such as IPs and district stability teams, this data collection approach may render it open to bias of groups that want to show progress and that may not have the most reliable and up-to-date on-the-ground information.

Prior to its change of course, YMEP planned to identify IP panels and external experts to make assessments of progress towards stability for each district or theme (such as microfinance or participatory governance). IP panels would consist of the IPs currently operating in each district or theme, while experts would include academics, USG partners, USAID officials, or representatives from other donors or delivery organizations with specialized knowledge of the districts or thematic areas. They would enter their assessments in a decision assessment tool and load them onto the YMEP web-based information system. The decision assessment tool would prompt panelists to examine: the stability outlook (static, improving, or declining); estimates of time interval for stability improvement; relevance of existing instability drivers or triggers; effectiveness of activities in addressing stabilization; suggestions for more effective activities; usefulness of

FGDs in FATA/Malakand: Ensuring Female Participation

A local organization provided social mobilizers who recruited FGD participants and moved those participants from FATA and Malakand to Peshawar for a three-hour group discussion. The FATA/Malakand Index secured 50% female participation in the 32 FGDs. Local mobilizers were from the area in which they recruited; females were used to recruit female FGD participants. The mobilizers received training and remuneration, and the female mobilizers had male chaperones so that they could move about in public. Female FGD discussants brought chaperones from home. This approach to securing female participation worked well.

indicators for monitoring stability; and stabilization effects that were not measured and suggestions for how to measure them.

Scaling of Data: Two of the programs in our sample undertook significant efforts to scale the data in order to facilitate comparisons. In the RC-South approach, for example, a Rating Definition Level (RDL) is developed for each of the 18 indicators with assigned definitions for each point on a five-point Likert scale. This constitutes perhaps the chief innovation of this approach because the 5-point Likert scale for all the indicators enables a comparison of progress across all indicators. In MISTI, the maturity model provides standard definitions of five levels of district stability – Insecure, Unstable, Fluid, Stable, and Normal – utilizing key indicators for stability. MISTI will categorize districts according to one of these five standard definitions. The DSF/RC-East application makes more limited use of scaling with the District/PRT POC (main scorer) and the Platform Section Chief (validator) ranking a few complex qualitative indicators on a 1-10 scale. An example of this is one of the “Afghan-on-Afghan violence” indicators ranking the extent to which the population feels it is safe to travel and feels that the GIROA protects them from insurgents and criminals. Data from other indicators presumably feed into this analysis so in a sense this indicator functions as the conclusion one would draw in analyzing data from several indicators and in looking at the trend line over time. A ten-point scale is unusual – it certainly allows for differentiation and nuance but it may be very difficult to determine with precision whether a 6 or a 7 is the correct score, and it may be difficult to regularize the application of the scale across different District POCs so that scores are comparable across districts.

Periodicity of Data Collection: The COIN/stabilization programs collect data more frequently than the CVE programs do in response to their more volatile environment. DSF/RC-East is collecting data on its 30 indicators quarterly. HMEP collects some data on a rolling basis and carries out its household survey every quarter. The FATA/Malakand Stability index collects data for different elements of the index according to different schedules: the survey is done annually, FGDs semi-annually, bazaar observation quarterly, IP reports on freedom of movement monthly, and media monitoring (e.g., of violent incidents and social and political gatherings) is ongoing. The survey is annual due to its expense. In RC-South, data collection occurs almost continually and the streamlining of metrics has cut the assessment process from six weeks to a matter of days. Some of the DSF/Afghanistan data are collected (or aggregated and considered) weekly or monthly and might be based on daily records (clinic visits).

In non-volatile environments, data are collected less frequently. Eastleigh planned to conduct data collection quarterly but is shifting to semi-annually since it does not expect meaningful changes every quarter and since the data is of relatively little use in its programming. PDEV II is conducting a baseline, midline, and endline household survey and will of course collect data on routine project events and outputs more frequently. Garissa is conducting quarterly community needs assessments, which may be more often than necessary as the data have limited value for programming.

Data Aggregation: The challenge of communicating complex data, collected and analyzed at frequent intervals, can lead to over-aggregation of data, including dashboards for quick snapshots and the like. We did not find this risk much present in the sample reviewed. By and large, programs are intent on looking at localized data for planning purposes. With the exception of the earlier CSP, all of the COIN programs collected data at the district level, which allows for careful monitoring of micro environments. This is an emerging best practice in COIN M&E. PDEV II is still considering the level of aggregation for overall presentation of data (e.g., an average score across communities on a target index is a relatively high level of aggregation if there are many communities) but will be assessing data across indicators at the community level. The other two current CVE programs in Kenya both focus on relatively narrow geographic areas and target populations, reducing any risk of over-aggregation.

Use of Local Firms: Many of the programs in our sample use local firms (and sometimes the local affiliate of an international group) for data collection. Local firms can be useful for sensitive data collection because they are at one remove from the IP and they are likely to have an easier time identifying and hiring data

collectors representing the needed mix of ethnicity, gender and language. Using local firms in monitoring efforts can also help to build local monitoring capacity via training and quality control mechanisms. They do not represent a clear cost saving, however, as the capacity building and supervision costs can be substantial. Moreover, the need for methodological sophistication increases as VE/I monitoring efforts use increasingly abstract indicators and survey questions, and relying on inexperienced firms can undermine monitoring efforts. PDEV I provides an example of an under-supervised local firm that delayed the base line survey by a year and then produced unreliable data.

In an interesting innovation, RC-East personnel in its DSF application used local Afghan partners (e.g., officials, elders) to collect data for quantitative data of observable phenomena (e.g., clinic visits, commercial vehicles on the main market road) to build their interest in gathering and using data and their skills to do so.

Cost of Data Collection and Analysis: It is difficult to compare cost information for the monitoring systems in our sample. USAID and IP staff costs for M&E are often difficult to calculate. On implementation program, local partners often play a role in collecting data. MISTI, which covers all programs in the USAID Stabilization Unit's portfolio and does data collection across all four regional platforms and 83 districts, has a budget of \$5 million per year for three years, which includes a bi-annual household survey of 40,000 respondents in addition to 11 program evaluations, 1 experimental assessment, and 2 quasi-experimental impact assessments. HMEP's budget is \$2 million per year, which includes a quarterly household survey of 4,000 respondents, FGDs, and interviews, as well as one annual and up to four periodic reviews per year for Helmand province alone. PDEV II's budget includes \$1.26 million for the household survey for three iterations, but this does not include staff time or other data collection costs, including USAID's use of an external US contract to collect data in control communities and to do qualitative research in target ones. FATA's index application budget is roughly \$900,000 per annum (though this figure includes the index design and testing costs), which includes an annual household survey of FATA and Malakand Division as well as other forms of data collection. Costs should drop over time once the index has been finalized and data collectors are fully trained in implementing it. Because of the different types of data required, the difficulty of obtaining data in insecure environments, and the limitations on government and other sources of data, the mixed methods approach is costly and burdensome. Household surveys may be the most useful if carried out properly, but represent the most costly form of data collection. Efforts to undertake VE/I monitoring with less expense might focus on expert panels, FGDs, in-depth qualitative interviews, and SMS surveys, but aggregation will then pose more challenges.

Data Use and Learning

Data Analysis: In our sample, only HMEP, MISTI, and RC-South are conducting statistical modeling. They are using indicator data such as public service delivery and government legitimacy to assess impact on stability. HMEP, for example, uses regression analysis to investigate which forms of service delivery improvement are most tied to increasing government legitimacy. The findings allow planners to invest more in those services that increase government legitimacy (e.g., schools, agriculture support, or roads). At the program level, statistical analysis can also assess each level of the intervention logic (inputs, outputs, outcome, and impact). None of the CVE programs is using regression analysis.

Data Presentation: The fluidity of the environment and the need to track programs and stability at the district and sub-district level make digesting data from monitoring efforts particularly challenging for COIN programmers. HMEP and RC-South, in particular have devised shorter monitoring reports that make use of radar diagrams, heat maps, and dashboards – simple graphic summaries that show progress over time against the key indicators. This kind of data presentation highlights comparisons, key findings, and recommendations.

Data Review: Several programs model a best practice in the face-to-face review of data. HMEP holds two-three day planning sessions with leadership to discuss the data; they take place every three or six months.

Eastleigh also conducts program reviews quarterly, in accordance to OTI's model, and an external review annually. Similarly, MISTI recently held the first of a series of events for stabilization program M&E practitioners and interested parties designed to share best practices and lessons learned. This group will meet on at least a quarterly basis. DSF/ASI-East also undertook periodic reviews of data with program teams to ensure accuracy and determine causes of observed trends. The OTI team also reviews data on FATA on a frequent schedule, consistent with overall OTI M&E guidance.

While this kind of review is important, a constant need to educate new people on the data and how to use the data places a constraint on the this process. This pertains to environments with one-year rotations such as Afghanistan and Pakistan.

Data Access: Most of the programs examined have a data system that is accessible to the funder. MISTI and HMEP represent the more comprehensive data systems, which represent efforts to make up-to-date, consistent information available to a large number of relevant actors. MISTI is developing a knowledge management portal that will include evaluations and reports, raw datasets, maps, and GIS information. HMEP has a similarly robust website that is accessible to coalition government officials, their contractors, and UN agencies with authorization from the UK government. While most end users just read the HMEP reports they receive via email, some analysts in the survey and defense community have worked with the raw data. YMEP has developed a data system that enables USAID officials to submit queries and generate customized reports, though it is unclear how much they have used this function. For the FATA/Malakand index, USAID plans to make general statistics and its analyses accessible to stakeholders in Pakistan, but will not share the raw data. Among the CVE programs, Eastleigh uses the OTI database and PDEV II will have a database with a GIS link.

Data Impact: This inventory points to spotty impact of the monitoring efforts to date. PDEV I and CSP had weaker monitoring systems that made it hard to discern how well the programs were doing and how to adjust programming for optimal impact. While more robust and promising systems, PDEV II, MISTI, FATA/Malakand, and Eastleigh are in their early stages so do not provide examples of using data to guide programming. Garissa does show some impact from the monitoring effort: it has used process or output data to make adjustments in its activities and approaches, such as revising a training curriculum. According to interviews, the two programs most closely aligned with the military, HMEP and RC-South, suggest the strongest impact from monitoring but of the more robust systems they have also been in place longer. The fact that the Australian government is replicating HMEP in the adjacent Uruzgan province and ISAF is extending the RC-South approach to all of Afghanistan supports this assessment. DSF/ASI-East data and analyses were used to adjust programming and have informed the follow-on project, Community Cohesion Initiatives. They were also shared with other USG entities and donors to provide “situational awareness.”

RECOMMENDATIONS

This brief section on recommendations is oriented more toward the VE/I Steering Committee and to M&E challenges specifically than it is to individual Missions implementing CVE/I programs and developing M&E systems to track progress and impact.

Improving the State of M&E for CVE/I Programs

Establishing a Community of Practice: There is evidence that more collaboration is happening in the development of M&E systems for these complex programs. However, a more systematic approach is needed to share frameworks and logic models, indicators, approaches to complex measurement challenges, data collection tools, and attempts to simplify large M&E systems. Establishing a wider Community of Practice, with online posting of tools and discussions, as well as face-to-face workshops and “summits,” can help both discover and advance good practice. Because several of these attempts are new and have little or no time

series data behind them, it is difficult to discern at this juncture how they will work out in practice. Some of these systems, particularly the ones that deliberately simplify the amount of data they are trying to capture as well as those working with new concepts like social cohesion and resiliency, are worth following over time. Questions include:

- Did the combination of indicators yield useful information?
- How well did indicator definitions capture the concepts being measured?
 - Were adjustments for the culture and environment required?
- Was critical information missing?
- Was there indicator redundancy and could they have managed with fewer indicators?
- Was the information used and how effectively was it used (because if it is not used, the investment is hardly merited, even if usable information and analysis are produced)?

Training: As the new VE/I strategy goes into effect and more such programs and projects are developed, more training needs to be provided to USAID personnel, other elements of the USG implementing the equivalent of development assistance efforts to reduce CVE/I risk, and IPs on CVE/COIN M&E. A source of short-term technical assistance (STTA) probably needs to be made available to new projects and programs as they get underway, both to shape good theories of change as well as accompanying PMPs. This is particularly true for smaller, less well-funded programs that may lack access to high-level methodological and/or substantive expertise. Training may also be required in analysis and data use.

Solving Measurement Challenges: USAID/Washington might also invest in helping solve some of the trickier or more complex measurement issues:

- What might be the best ways to measure youth empowerment or community resiliency?
- What factors signal stabilization most strongly and should be part of a stability index, and how much does this vary from one district or country to another?
- How localized does data collection need to be in COIN environments?
- How and when does it make sense to aggregate across indicators and to scale data to provide a numerical level of accomplishment (currently, the RC-East effort appears to do this best)?

If USAID staff and IPs continue to struggle with the answer on their own, the costs will be greater and a heterogeneous assortment of indices and measures will result, some of which will be stronger than others. As an example, an important concept in youth empowerment could be the idea of “locus of control” or individual sense of agency, a concept drawn from Western psychological research. Individuals with a high internal locus of control believe that events result primarily from their own behavior and actions. By contrast, those with a high external locus of control believe that powerful others, fate, or chance primarily determine events. Western psychology has developed ways of measuring the locus of control, which may or may not translate well to other cultures.⁹ Some centralized effort on the relevance of the concept and ways it might be captured could be a boon to those working on programs promoting youth empowerment.

In addition, effort is needed on ways to capture perception, behavior and experience on sensitive topics in surveys. Many such questions are asked on surveys in COIN environments and they often draw a high percentage of “don’t know” responses or responses that seem counter-intuitive given the situation of the ground, so that the likelihood of social desirability bias affecting the responses is high.

Provision of Short Term Technical Assistance: STTA might also be provided to shape survey questionnaires (converting indicators into one or more questions) and sample frames. Just as USAID’s DRG Center is trying to collect good scopes of work for common outsourced tasks like evaluations, CMM or PPL

⁹ See Lynn Carter, *Measuring Progress in Development Assistance Programs Countering Violent Extremism*, USAID, September 2010, 18 for more on this topic.

might collect good surveys for CVE/I programs. There is a considerable need for quality control on CVE/I surveys – and a need to get baseline data in place much more quickly. This argues for the engagement of international survey research firms in many parts of the world where local research skills are minimal. USAID might also look into training a cadre of regional research organizations to implement such surveys as one cost-lowering measure.

Addressing a Few Gaps

Monitoring VE Environments: In the sample of programs reviewed, one gap appears to be in the lack of formal monitoring of the overall environment in CVE programs. This is in contrast to formal, rigorous measurement of the environment in insurgency contexts. It is difficult to tell with largely oral processes how much is actually being done to understand and adapt to a changing context. OTI does appear to do a better job of this than other parts of the Agency due to its well-articulated process. Still, it is more difficult for the reviewers to represent the value of such oral processes (formalized or informal) in this paper than it was for systems for which we had extensive written documentation.

Measuring Youth Livelihoods: As noted previously, programs are learning almost nothing about even the medium-term impacts from some forms of youth programming – employment, starting businesses, technical training. Youth are not tracked nearly long enough when they are tracked at all. We are much more likely to see indicators related to how many started new businesses or acquired a loan than what happened two or three years down the line. This means programs miss critical information that would allow them to adapt their interventions to improve their results. But more importantly, if the youth fail, it is possible that the assistance effort results in an elevated sense of injustice and grievance, suggesting that the investment has generated a negative return and not simply no return. USAID either regionally or centrally could invest in tracking some percentage of youth livelihood beneficiaries at least for a few years out post-project in order to learn.

Measuring the Connection between Improved Services and Government Legitimacy: One compelling hypothesis is that improved provision of services (if linked to the government) improves government legitimacy and undermines support for VE actors. This brings to mind a number of questions:

- Which services most enhance or undermine legitimacy?
- Do the relevant services vary across COIN or CVE environments?
- How does the level of perceived corruption affect the legitimacy calculus of citizens (some level of corruption is accepted and even expected in many of these environments)?
- Does it matter how the services are delivered or simply that they exist?

These are questions that are critical to programming and that seem largely unanswered. Data now being collected in some environments should help us answer them.

Monitoring the Extent of Data Utility and Use for Those Systems that Monitor the Environment: A lot of data are collected in COIN environments. HMEP staff have noted the difficulty of getting so much data into use. Some tracking of the more robust monitoring efforts and how those data are used, problems in using the data, and how the data affected decision making could be helpful in deciding how to put such systems together, how much data to collect and of what sort, and how to aggregate data in a useful fashion. This comes down to a question of knowledge management or the operating culture in high-pressured environments where decisions are made quickly.

ANNEX A: SCOPE OF WORK

Inventory of Monitoring and Information Systems for Programming against Violent Extremism and Insurgency

Scope of Work

I. Background

USAID has recently released a new policy on “The Development Response to Violent Extremism and Insurgency.” The policy builds on insights from years of actual program implementation across the globe, including a great deal of well-funded programs in Asia and the Middle East. The policy in particular calls for programs that focus on the drivers of violent extremism and insurgency and calls for focusing those engagements on those drivers likely to have the highest impact. Operationally, the policy calls for intense program management and innovation in the evaluation and review of programming. As a follow on to this policy, USAID has created a VEI Steering Committee to develop tools and guides for the implementation of this policy. As part of the distillation of lessons from previous efforts to capture information about VEI and adapt programming during implementation, the Steering Committee is seeking an inventory of past systems for real-time program feedback and adaptation in VEI programs as part of a larger effort to produce field handbooks and operational guides to assist USAID practitioners in the field.

II. Tasks

Prepare a report including the following:

1. Desk studies inventory of past monitoring and reporting systems or processes used by USAID to report on, learn about and adapt programs designed to address violent extremism and insurgency. The programs should include:
 - The Community Stabilization Program (CSP) in Iraq
 - Jeff Goebel (jgoebel@usaid.gov)
 - Will Wilhelm (williampwilhelm@gmail.com)
 - Jennifer Tikka (jtikka@usaid.gov)
 - Lynn Sauls (lsauls@usaid.gov)
 - The District Stabilization Framework (DSF) in Afghanistan
 - Jess Reitz (jreitz@usaid.gov)
 - Lauren Sweeney (lsweeney@usaid.gov)
 - David Ratliff (dratliff@state.gov)
 - Bea Covington (beacovington@hotmail.com)
 - Justin Richmond (jrichmond@state.gov, jrichmond@stabops.org)
 - District Stability Framework (DSF) in Pakistan
 - Andrew Hall (hallad@state.gov)
 - District Stability Framework (DSF) in South Sudan
 - Jay Singh (jsingh@usaid.gov or jsingh.usaid@gmail.com)
 - DSF worldwide
 - Eric Kotouc (ekotouc@usaid.gov; ekotouc@stabops.org)
 - Kristina Kempkey (kkempkey@usaid.gov)

- West Bank/Gaza – Sarah Borodin
- Yemen Monitoring and Evaluation Project (YMEP)
 - Jeff Ashley (jwashley@state.gov)
 - Caryle Commisa (caryle.cammisa@socom.mil)
- Colombia
 - Miguel Reabold (mreabold@usaid.gov)
 - Elizabeth Winger (ewinger@usaid.gov)
- Kenya (OTI Eastleigh)
 - Galeeb Kachara (gkachra@usaid.gov)
- Kenya (Garissa Youth- including partial application of DSF)
 - Dwaine Lee (dlee@usaid.gov)
 - Lucy Kithome (lkithome@usaid.gov)
 - Elena Vinogradova (evinogradova@edc.org)
 - Christina Ciak (cciak@usaid.gov)
- Niger and Chad (PDEV)
 - Moussokoro Kane
- Other Historical USAID Programs on a case-by-case basis after consultation with USAID/Asia/ME.

Non-USAID/USG programs for Comparison

- USIP, MPICE – countries of use?
- Coffey ID, HMEP, Afghanistan (British Foreign Office)

The desk studies should address the following questions:

- Are the M&E systems embedded in a clear theory of change?
 - What information were these systems designed to collect?
 - How well do the indicators/information match the desired changes?
 - How does that information differ across the cases?
 - How was this information to be collected?
 - Who has/had access to the information / database?
 - How user-friendly is the system? Was it likely to ease reporting for both the IP and USAID?
 - How was this information to be reviewed and used to adjust programming?
 - What were key constraints to utilization of the information?
2. An evaluation of the implementation of these same systems through a review of reports and data produced by them and through interviews with those who gathered the information and those who used or were supposed to use this data (including within USAID and IP organizations). This should answer the questions:
- What information was actually collected and what was its quality and reliability (in the estimation of the collectors and the end users)?
 - What information could not be collected and why?

- How was this data used to make adjustments to program design and implementation? How did it inform strategy, particularly in regards to VE/IP?
 - How did this system assist in countering VE or I efforts?
 - Were there unplanned discoveries from and uses for this data that should be captured?
 - Did the system in any way augment or undermine staff or IP security?
3. An inventory of concrete indicators and drivers which were considered to be most relevant in validating program design and adapting programs, both those that are not captured in the USAID *Guide to Drivers of Violent Extremism* and *Measuring Progress in Development Assistance Programs Countering Violent Extremism*.
- What specific VE/I indicators were used to measure program or project outputs or outcomes? How were VE/I indicators woven in to general sector programming?

III. Deliverables

- The report described above.
- A presentation of the report and discussion with the VEI Steering Committee and other interested USAID staff.

IV. USAID Support

USAID will facilitate access to:

- Relevant program and system design documents, quarterly and annual reports of IPs and reports from data monitoring systems as needed.
- Access to actual monitoring and evaluation databases where practical.
- USAID, former USAID and IP contacts for interviews, including introductions where needed to facilitate full cooperation.
- Access to other inputs as needed within the scope of these tasks.
- Access to the actual data collection systems would be pertinent – would want to look at how the data is stored, cross-tab functions for comparing data, built-in systems for linking the Mission’s CDCS, etc.

ANNEX B: INTERVIEW GUIDE

Inventory of Monitoring and Information Systems for Programming against VE/I

MSI for the AME Bureau

Draft May 20, 2012¹⁰

We envision the collection of data using the following three phases. In Phase I we will email relevant individuals explaining the task and requesting key documents. We propose to start with a document review in order to spare USAID staff time in answering questions that the documents might address. We will then move to interviews to flesh out the picture that the documents present and to address concerns like cost that the documents are unlikely to speak to. We propose two levels of interviews, described below.

Phase I will include an email of introduction to the inventory project, along with a request for specific documents and reports including, but not limited to, written documentation of the monitoring and information system design, documents explaining VE/I strategy (Results Framework and an accompanying narrative), COIN or CVE specific documents used to influence the M&E system and/or programming, project logframe(s), relevant project descriptions, and/or project/program annual reports and evaluations.

MSI will review these documents to gain a broader understanding of strategic approaches to VE/I efforts; to review the set(s) of indicators being used to measure VE/I efforts; to become familiar with the design and intention of the monitoring system in each country; and to compare approaches and systems across the various Missions.

Phase II: VEI Interview Guide

The interview guide is designed for the purposes of conducting a phone or in-person interview. The purpose of Phase II is to ask questions left unanswered by the document review. The following list is illustrative and should be tweaked based on what information each Mission is able to provide. Interviews could be conducted with implementers as well as USAID personnel.

Name (from Phase I)

Position (from Phase I)

Role

Programs/Projects of Relevance

Are there logframes or a Results Framework that incorporates stabilization/peacebuilding/COIN/CVE considerations at the project or program level?

What risks and assumptions underpin the RF/logframe/project logic?

Are their explicit theories of change associated with the program, logframe or RF? What are the key determinants of stability in this environment? Are these M&E systems embedded in a clear Theory of Change?

Are there specific indicators that measure stabilization/peacebuilding/COIN/CVE?

¹⁰ Prepared by Lynn Carter and Jennifer Ulman.

Do indicators for the said programs/projects also measure the impact on the environment (the conflict situation) and/or do they measure the impact of the situation on the program/project?

Describe the Mission's/Office's/Region's monitoring practices overall.

What specific monitoring systems and databases exist, if any? Describe how this system is used internally and/or externally with IPs/other USG partners/donors? Who has access to the information/database? How user friendly is it?

Was the system designed to incorporate measurement of COIN/stabilization/CVE objectives? What else is the database/monitoring system used for (GIS capability, Security info, etc.)?

Were there monitoring efforts that were attempted but then were dropped or failed due to labor or cost requirements, insecurity and/or overall complexity?

Based on the results of these interviews, we may then request access to the database or samples of data in the database, or other relevant documents mentioned in the interviews.

Phase III: VEI Interview Guide

The purpose of Phase III questioning is to delve deeper into how the more interesting and/or successful monitoring systems and explore more fully how the information/data coming out of them were actually used. Phase III questioning is simply a continuation of Phase II, and will be disregarded in cases where the system was not actualized or appears of limited replication value.

What type of measurement tools/approaches (surveys, focus groups, etc) are/were generally used? What is deemed most appropriate given the specific context in which these programs are taking place? If a database exists, what approaches are most challenging to capture?

Have RCTs or quasi experimental impact assessment design techniques been used? With what results?

How are data collected? By whom? Are there different practices for the various projects?

What information/data is actually collected? How would you describe its reliability and quality?

Are written analyses generated out of the system? If no system, then how is analysis done and captured? Is there analysis done internally within the Mission? Is there a process in place?

How extensively have survey data been analyzed? What are the perceived limits of surveys in that environment (e.g., respondent fatigue and canned answers, security/supervision of data collectors, extreme social desirability bias due to insecurity, etc)? Have survey experiments been implemented where social desirability bias is a serious risk?

Are data aggregated across programs/projects? How?

Are data aggregated across districts/provinces/regions and at what cost to understanding specific micro-climates?

How is analysis used? For what (deciding on/adjusting project activities, changing activities, revising indicators, revisions to strategy)?

How was this information/data used to make adjustments to program design and implementation? How did it inform current or future strategy, particularly in regards to VE/I?

How did this system assist in countering VE or I efforts?

How is the data or system used to influence interagency efforts?

Were there unplanned discoveries from these data that should be captured?

What information/data collection was planned but not attained and why?

Did the system in any way augment or undermine staff or IP security?

What is the approximate cost of the monitoring system design and for data collection and analysis? Given experience to date – how would you describe the utility of the system versus its cost? What they might do differently now? Any lessons learned?

ANNEX C: INTERVIEWS AND MATERIALS REVIEWED

| Community Stabilization Program (CSP) - Iraq | |
|---|---|
| Documents Reviewed: | <p>IRD, Monitoring and Evaluation Plan for the Iraq Community Stabilization Program, July 1, 2007</p> <p>IRD, Community Stabilization Program Final Report, January 2010</p> <p>IBTCI, Data Quality Assessment, Iraq, December 2006</p> <p>IBTCI, Evaluation of USAID's Community Stabilization Program (CSP) in Iraq: Effectiveness of the CSP Model as a Non-lethal Tool for Counterinsurgency, July 2009</p> <p>IBTCI, USAID Community Stabilization Program (CSP) Counterinsurgency (Coin): Rapid Programmatic Assessment, April 2008</p> |
| Interviews: | Will Wilhelm, former USAID Senior Advisor in M&E, 11/06/12 |

| District Stability Framework (DSF) – Afghanistan | |
|---|---|
| Documents Reviewed: | <p>District Stability Framework Quick Reference Guide</p> <p>District Stability Framework Powerpoint, COIN Training Center-Afghanistan and USAID/OTI</p> <p>New RC-East Campaign Assessment Review Board Governance-Development Metrics, 25 May 2012</p> <p>Overall Stability Assessment in Twelve Districts in Eastern Afghanistan: Final Report, Afghanistan Stabilization Initiative-East, 2012</p> <p>RC-East Campaign Assessment Review Board Governance-Development User's Guide, 25 May 2012</p> <p>Summative Impact Assessment of ASI-East In Three Districts In Eastern Afghanistan: Marawara, Shirzad, And Barmal Districts, Afghanistan Stabilization Initiative-East, August 2012</p> |
| Interviews: | <p>Molly Byrne, OTI/Afghanistan, 11/13/12</p> <p>Alexa Courtney, Caerus Associates, 08/14/12</p> <p>Bea Covington, former FPO in Afghanistan and DSF Coordinator for RC-East from 2006-February 2012, 08/16/12</p> <p>Kristina Kempkey, CMC, 11/16/12</p> <p>Justin Richmond, former DSF Coordinator for RC-East, 11/19/12</p> <p>Lauren Sweeney, OTI, 08/16/12</p> <p>Email communication with Andrew Rathmell, Coffey International Development, 08/12</p> <p>Email communication with Lynn Carter, MSI, 08/12</p> |

| Eastleigh – Kenya | |
|--------------------------|---|
| Documents Reviewed: | <p>Performance Monitoring Plan</p> <p>Survey Questionnaire</p> <p>Focus Group Discussion Questionnaires</p> |

| Eastleigh – Kenya | |
|--------------------------|---|
| | Cognitive Edge Signification Framework |
| Interviews: | Abdimalik, Chemonics M&E Officer, 07/12/2012 |
| | Galeeb Kachra, USAID Kenya Transition Initiatives COR, 08/16/2012 |

| FATA/Malakand Index – Pakistan | |
|---------------------------------------|---|
| Documents Reviewed: | IMEC, Stability Index Statement of Work, April 2012 and May 2012 versions |
| | Powerpoint presentation on concepts and proposed measures/discussion with IMEC and USAID teams, March 2012 |
| | <i>Stability Index: Initial Findings – Draft Report</i> , undated but from November 26, 2012 |
| Interviews: | Jill Tirnauer, IMEC, various dates, via email and Skype, from May – Nov. 2012 |
| | Andrew Hall, OTI, USAID, 11/12/12 |
| Notes: | Lynn Carter provided input into the conceptual development of the index in November/December 2011 and in September 2012 |

| Garissa Youth Program – Kenya | |
|--------------------------------------|--|
| Documents Reviewed: | Performance Monitoring Plan, revised 2010 |
| | Risk Assessment, 2008 |
| | Project Design Document, 2008 |
| | Community Needs Assessment survey instrument |
| | Community Needs Assessment survey data for youth related questions for five rounds |
| | Quarterly report from 2009 |
| | Quarterly report for April-June 2012 |
| Interviews: | Conference call on 08/03/12 with Jacqueline Glinn (COP, EDC); Caroline Riungu (M&E Officer); Nancy Taggart (EDC home office) with follow up Q&A by email |

| Helmand Monitoring and Evaluation Programme (HMEP) – Afghanistan | |
|---|--|
| Documents Reviewed: | HMEP Annual Review, 2011 |
| | Coffey ID, Innovative Approach to Evaluating Interventions in Fragile and Conflict-Affected States: The Case of Helmand, 04/11 |
| | HMEP Draft Periodic Review, First Quarter 2012 |
| | HMEP Annual Review 2011 Key Findings Powerpoint |
| Interviews: | Andrew Rathmell, Technical Director of HMEP at Coffey International Development, 07/3/12 |

| Measuring Impacts of Stabilization Initiatives (MISTI) – Afghanistan | |
|---|--|
| Documents Reviewed: | Draft PMP, 07/26/12 |
| | MISTI Stabilization Trends and Impact Evaluation Survey, 08/12 |
| Interviews: | John Roscoe, COP of MISTI, 08/16/12 |

| PDEV I – Niger and Chad | |
|--------------------------------|---|
| Documents Reviewed: | PMP, original |
| | PMP, revised 2010 |
| | PIRs for revised PMP |
| | Midterm Amex evaluation survey instrument, 2010 |
| | Midterm Amex evaluation survey data, 2010 |
| | Baseline survey questionnaire |
| | Baseline survey questionnaire data |
| | End line survey questionnaire |
| | End line survey questionnaire data |
| | Final program report |
| | Encompass Peace through Development Program Assessment, March 2011 |
| Interviews: | Anne Salinas and Anna Mecagni, FHI 360, 07/07/12 |
| Notes: | Lynn Carter also drew on workshop notes from a CVE seminar she facilitated in Dakar, Senegal in September 2009. |

| PDEV II – Niger, Chad, & Burkina Faso | |
|--|---|
| Documents Reviewed: | RFP |
| | Preliminary PMP |
| | Preliminary Survey Instrument |
| | Endline survey for PDEV I |
| | Revised PMP (late August 2012) |
| Interviews: | Allison Poyac-Clarkin and Chung Lai of IRD, various dates via email and telephone |
| | Rob Kevlihan, senior conflict advisor for USAID/West Africa and COR for PDEV II, various dates, via email and Skype |

| Regional Command (RC) South – Afghanistan | |
|--|---|
| Documents Reviewed: | Caerus Powerpoint Presentation, “Experimentation and Learning: Crafting Meaningful Measurements of Stability”, 05/03/12 |
| | “Communicating the Here and Now: A Way Forward for Operational Assessments in Conflict and Post-Conflict Environments” by William P. Upshur (Caerus Associates) and Major Jonathan W. Roginski (US Army), 08/2011 |
| | AAG-S Campaign Assessment Rating Definition Levels |
| | Indicator Datasheet for Stabilization Trends Analysis |
| | MSI, Measuring the Impact of Stabilization Initiatives: Desk Review of Stabilization Resources and References, 07/13/12 |
| Interviews: | Alexa Courtney, Caerus Associates, 08/13/12 |
| | Erin Simpson, Caerus Associates, email communication, 11/2/12 |

| Yemen Monitoring and Evaluation Project (YMEP) - Yemen |
|---|
|---|

Yemen Monitoring and Evaluation Project (YMED) - Yemen

| | |
|---------------------|---|
| Documents Reviewed: | Indicators, IRs, AOs for CLP |
| | Indicators, IRs, AOs for RGP |
| | Yemen PMP, 02/11 |
| | YMED, Third Party Activity Monitoring of the Community Livelihoods Project, 10/11 |
| | YMED Projects Application User Manual |
| Interviews: | Robert Van Heest, IBTCI, 09/5/12 |

Other Sources

| | |
|-------------|--|
| Interviews: | Marcie Moburg, PPL, USAID, in person, 07/30/2012 |
| | Stacia George, OTI, in person, 07/30/2012 |
| | George Wilson, OTI, in person, 07/30/2012 |