NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

IMPACT ASSESSMENT IN SPECIAL WARFARE

by

Garrett M. Searle

December 2017

Thesis Advisor:	Robert Burks
Second Reader:	Douglas Borer

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704–0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE December 2017	3. REPORT TYPE AND DATES COVERED Master's thesis

4. TITLE AND SUBTITLE IMPACT ASSESSMENT IN SPECIAL WARFARE	5. FUNDING NUMBERS
6. AUTHOR(S) Garrett M. Searle	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number DD-N-2017.0083.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

The impact of operations that build partner capacity to counter extremist groups is difficult to measure. Assessments usually focus on the outcomes of training (was capacity built?) but not the ultimate effect of that new capacity (was a violent extremist group degraded as a result?). To address this shortfall, this thesis argues that a method known as impact assessment, used by the development and law enforcement communities to evaluate countering violent extremism (CVE) programming and policing strategy, can be applied to assess the social impact of military capacity-building efforts with similar goals. To demonstrate that utility, the author examines the case of U.S. engagement in Niger and develops a theory of change to describe the logical path from capacity-building activities to their intended effect of countering extremist groups. Then, to test impact assessment in practice, the author conducts an ex post facto, quasi-experimental assessment of the treatment effect of U.S. engagement in Niger. The substantive results of this study identify impact that was otherwise hidden in observational data and highlight the need for more rigorous assessment. The author recommends the application of impact assessment methods to improve the theoretical understanding of cause and effect, identify real impact, learn from unexpected results, and motivate adaptation and innovation.

14. SUBJECT TERMS special warfare, irregular warfare, impact assessment, impact evaluation, theory of change, countering violent extremism, foreign internal defense, measures of effectiveness	15. NUMBER OF PAGES 105
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540–01-280-5500 Standard Form 298 (Rev. 2–89) Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

IMPACT ASSESSMENT IN SPECIAL WARFARE

Garrett M. Searle
Major, United States Army
B.A., Wheaton College, 2006

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN DEFENSE ANALYSIS

from the

NAVAL POSTGRADUATE SCHOOL
December 2017

Approved by: Robert Burks
Thesis Advisor

Douglas Borer
Second Reader

John Arquilla
Chair, Department of Defense Analysis

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The impact of operations that build partner capacity to counter extremist groups is difficult to measure. Assessments usually focus on the outcomes of training (was capacity built?) but not the ultimate effect of that new capacity (was a violent extremist group degraded as a result?). To address this shortfall, this thesis argues that a method known as impact assessment, used by the development and law enforcement communities to evaluate countering violent extremism (CVE) programming and policing strategy, can be applied to assess the social impact of military capacity-building efforts with similar goals. To demonstrate that utility, the author examines the case of U.S. engagement in Niger and develops a theory of change to describe the logical path from capacity-building activities to their intended effect of countering extremist groups. Then, to test impact assessment in practice, the author conducts an ex post facto, quasi-experimental assessment of the treatment effect of U.S. engagement in Niger. The substantive results of this study identify impact that was otherwise hidden in observational data and highlight the need for more rigorous assessment. The author recommends the application of impact assessment methods to improve the theoretical understanding of cause and effect, identify real impact, learn from unexpected results, and motivate adaptation and innovation.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACM	civil-military actions
BPC	building partner capacity
CA	civil affairs
CME	civil-military engagement
CMSE	civil-military support element
COIN	counterinsurgency
CVE	countering violent extremism
C-VEO	counter-violent extremist organizations
DD	difference-in-differences
DOD	Department of Defense
FAN	Armed Forces of Niger
FM	field manual
FSATT	feasible sample average treatment effect on the treated
GRAP	Global Research and Assessment Program
ISAF	International Security Assistance Force
IW	irregular warfare
JCET	joint combined exchange training
JP	joint publication
MEDSEM	medical seminar
MISO	military information support operations
NATO	North Atlantic Treaty Organization
SC	security cooperation
SFA	security force assistance
SME	subject matter expert
SOCAFRICA	Special Operations Command Africa
SOF	special operations forces
SW	special warfare
TSOC	theater special operations command
USAID	United States Agency for International Development
VEO	violent extremist organization

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank my advisors, Dr. Robert Burks and Dr. Doug Borer, for their clear guidance, thoughtful feedback, and dedication to the success of their students. I would also like to thank Dr. Camber Warren, whose support and assistance was critical to the success of this research effort. I am grateful to Dr. Heather Gregg, Dr. Jon Czarnecki, Dr. Anna Simons and other members of the exceptional faculty of the Naval Postgraduate School and Naval War College, who contributed to this effort through their teaching, encouragement, and feedback.

I would also like to thank John Wishart, whose support made this project possible. Additionally, I am deeply indebted to the many leaders, colleagues, and friends of mine who gave their time, conversation, and ideas to assist with this effort.

Lastly, I would like to thank my wife, Tegan, who inspires me every day with her compassion, generosity, and strength.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The impact of operations designed to build partner capacity to counter extremist groups and insurgencies is difficult to measure. After years of U.S. forces advising and assisting partners in Africa, South-East Asia, and the Middle East, the connections between those efforts and their intended effect are not well understood.

In 1962, Robert McNamara was developing a system to measure the progress of the military assistance effort in South Vietnam. He had a list of measurements that he considered vital for the assessment of that campaign: control of lines of communication and critical infrastructure, the number of pacified villages, and, most importantly, the number of dead Viet Cong. When Brigadier General Edward Lansdale appeared in his office, McNamara asked him what he thought of the list. Lansdale had been the architect of the successful Philippine counterinsurgency effort against the Hukbalahap rebellion and had spent a considerable amount of time in Vietnam in the early years of American involvement there. Looking at McNamara's list, he concluded that measuring those factors would lead to misplaced confidence in the progress of the campaign. Lansdale felt that the most important factor was missing—what he called the "X factor," or "the human factor."[1] To him, the most vital component was "What the people out there on the battlefield really feel; which side they want to see win and which side they're for at the moment. That's the only way you're going to ever have this war decided."[2] McNamara was disappointed by Lansdale's answer and erased his note about the "X factor," dismissing Lansdale and his suggestion as "too unconventional."[3]

Almost 50 years later, when General Stanley McChrystal assumed command of the International Security Assistance Force (ISAF) in Afghanistan in 2009, appreciation for the "human factor" in counterinsurgency had grown considerably. In his initial assessment, McChrystal noted that "Our strategy cannot be focused on seizing terrain

[1] Howard Jones, *Death of a Generation: How the Assassinations of Diem and JFK Prolonged the Vietnam War* (New York: Oxford University Press, 2003), 185.

[2] Jones, *Death of a Generation,* 185.

[3] Cecil B. Currey, *Edward Lansdale: The Unquiet American* (Washington: Brassey's, 1998), 2.

or destroying insurgent forces; our objective must be the population. In the struggle to gain the support of the people, every action we take must enable this effort."[4] Despite his emphasis on the population and its role in the counterinsurgency fight, McChrystal observed that assessments tended to measure progress primarily based on "kinetic events," which prevented "an accurate assessment of the insurgents' intentions, progress, and level of control of the population."[5] This incongruity continues to plague U.S. irregular warfare efforts, particularly those that are indirect, where the focus is on building the capacity of a host-nation military to address an extremist or insurgent threat. The special operations forces engaged in those efforts lack the necessary tools and expertise to assess the specific impact of their efforts on Lansdale's "X factor." Focusing specifically on capacity-building efforts in Africa, this thesis explores that deficit and identifies the method known as impact assessment as a means of addressing it.

A. TWO SIMILAR ATTACKS, TWO DIFFERENT OUTCOMES

As dusk fell on June 3, 2016, soldiers from the Armed Forces of Niger looked south from their small outpost across the dry scrubland and the outskirts of Bosso town, in the remote southeastern corner of Niger. Beyond the town was the Komadougou River and Nigeria's Borno State, home to the violent extremist group Boko Haram. As the day's heat gave way to a cooling breeze from the south, it appeared to bring with it small fingers of rising dust, barely visible in the fading light. In an instant, the serenity was shattered by the clatter of automatic weapons and rocket-propelled grenades screeching out of the bush and exploding on the makeshift dirt and sand-bagged battlements of the Nigerien army. Groups of Boko Haram fighters, crammed into cut-down Toyota pickups and mounted two and three at a time on cheap Chinese motorcycles, emerged from the darkness and swarmed towards the small outpost, intent on driving out the forces there and capturing the town. The fight was over quickly, the Army was forced to withdraw

[4] Stanley McChrystal, "COMISAF Initial Assessment (Unclassified)," August 30, 2009, published by *The Washington Post*, September 21, 2009, http://www.washingtonpost.com/wp-dyn/content/article/2009/09/21/AR2009092100110.html.

[5] McChrystal, "COMISAF Initial Assessment."

from their position with 32 soldiers killed and 67 wounded. Boko Haram spent the night torching buildings and looting the town, leaving early the next morning before the Nigeriens could organize a counter-attack.[6]

Nine months later, a large force of Boko Haram fighters was once again moving from their stronghold in Nigeria for a raid across the border in Niger. Their target was the small Nigerien outpost in Gueskerou, 20 miles southwest of Bosso. This time, however, they met a different fate. The local population observed and reported the movement of these forces toward Gueskerou, and the Nigerien Army moved a rapid response battalion, a partner force of U.S. Special Forces advisors, into position for an ambush. The Boko Haram fighters were caught flat-footed and exposed—pushed back into Nigeria with 57 killed and many more wounded.[7] They had no response for a Nigerien force that seemed to anticipate their moves. The flow of information that facilitated this rapid and effective response was a result of efforts by the Nigerien Army's civil-military affairs (referred-to by the French acronym ACM) forces, partnered with U.S. Army Civil Affairs (CA) teams, to build more productive relationships with at-risk populations.

The outcome seen in this second example illustrates the intent of this and other capacity-building efforts in Africa: to counter violent extremist organizations like Boko Haram and reduce their hold on vulnerable populations on the continent.[8] Ideally, some of that effect is lethal: killing and capturing extremists; but much of it is non-lethal: building a more competent, professional, and engaged military force that can garner the support of the population and reduce the influence of extremist groups. Achieving the latter goal means increasing perceptions of national government legitimacy and influence in areas impacted by an insurgency or non-state group and building capability to produce relationships that support security force action against those groups—like the relationships that supported the Nigerien counter-attack. In many conflicts where the U.S.

---

[6] "Boko Haram attack in Niger 'kills 32 soldiers'," *Aljazeera*, June 4, 2016, http://www.aljazeera.com/news/2016/06/boko-haram-attack-niger-kills-32-soldiers-nigeria-160604125912477.html.

[7] "Niger Forces Kill 57 Members of Boko Haram: Defense Ministry," Reuters, April 10, 2017, http://www.reuters.com/article/us-nigeria-security-idUSKBN17C29B.

[8] Donald C. Bolduc, *Commander's Narrative: Ongoing Discussion on Special Operations in Africa*, (Stuttgart, DE: U.S. Special Operations Command Africa, 2016).

military is engaged, responsibility for this non-lethal component falls on Civil Affairs forces, but each component providing advice and assistance is involved because the forces they train and equip will inevitably interact with civilian populations and impact the relationship between them and the state. These relationships solidify gains made through combat operations and strengthen the competitive advantage of the supported state as it consolidates control in areas of contested sovereignty.

Unfortunately, anecdotes like the one above do not provide the empirical means to understand how and why the partner capacity being built by U.S. forces is impacting the relationship between governments and their population. This problem is not unique to Civil Affairs forces—in general, capacity-building operations suffer from an inability to connect that new capacity to impacts on the critical relationship between partner forces and populations whose loyalty is necessary in the fight against extremist groups. There presently exists no mechanism to learn and adapt based on evidence of success or failure, making it difficult to aggregate those effects in support of broader political-military objectives. A number of conceptual frameworks have been proposed to measure the effectiveness of stability operations and this type of capacity building.[9] However, their recommendations were specific to the direct approach of counterinsurgency in Iraq and Afghanistan or focused on training program outcomes, rather than their specific impact or effect, thus leading to the type of misplaced confidence described by General Lansdale. Certainly, commanders need to understand the effectiveness of a training program in producing capable forces and establishing a productive bi-lateral relationship, but they also need to understand what effect those forces are having on the relative strength of the enemy or within the populations they serve. Ultimately, investment in building the capacity of our partners should be measured based on these higher-order goals. Joint and service doctrine and existing military practice do not provide sufficient information or adequate tools for this purpose.

[9] Jan Osburg, Christopher Paul, Lisa Saum-Manning, Dan Madden, and Leslie Adrienne Payne, *Assessing Locally Focused Stability Operations*, (Santa Monica, CA: RAND, 2014); Christopher Paul, Brian Gordon, Jennifer D. P. Moroney, Lisa Saum-Manning, Beth Grill, Colin P. Clarke, Heather Peterson, *A Building Partner Capacity Assessment Framework: Tracking Inputs, Outputs, Outcomes, Disrupters, and Workarounds*, (Santa Monica, CA: RAND, 2015).

B. RESEARCH FOCUS

To help address the assessment shortfall demonstrated above, this thesis looks beyond the Department of Defense (DOD) assessment paradigm and identifies potentially useful assessment mechanisms from other social science disciplines. Programs implemented by the development and law enforcement communities to counter violent extremism (CVE) and evaluate policing strategy use a form of evaluation known as *impact assessment* to understand their effect.[10] Impact assessments differ from traditional measures of effectiveness by constructing a logic model of cause and effect, called a theory of change, and then using that theory to guide experimental or quasi-experimental assessment design with a defined counterfactual. This thesis argues that these methods are effective tools to assess civil-military capacity building and other components of special warfare (SW) campaigns, and demonstrates that utility through their application within the context of an ongoing special warfare effort. Since these methods are already in wide use by interagency and other unified action partners, their application by SW practitioners would aid in what Melton and Holshek call the "incorporation of civilian and military assets under a coherent, strategic civil-military conceptual framework that addresses the gulf between people and their system of governance."[11]

In much of the recent literature about assessment in irregular war, and consistently in my conversations with participants in these conflicts, the question seems to be about metrics—"How can we find the right metrics to measure effectiveness?"; or, "What is the right set of metrics to link tactical actions to effects at the operational or strategic level?"[12] I want to be clear that the goal of this thesis is NOT to identify some new set of metrics that will definitively answer those questions. Rather, it proposes a

10 The development sector refers to assessment as "monitoring and evaluation," and thus uses the term "impact evaluation" to describe the same processes that are referred to here as impact assessment. These terms are synonymous. See US Agency for International Development, *USAID Evaluation Policy* (Washington, DC: U.S. Agency for International Development, 2011 (Updated 2016)), www.usaid.gov/evaluation/policy.

11 Kevin Melton and Christopher Holshek, "Symposium Workshop Report," in 2016-2017 Civil Affairs Issue Papers: Leveraging Civil Affairs, ed. Christopher Holshek (Carlisle, PA: Peacekeeping and Stability Operations Institute), xxi.

12 See Jonathan Schroden, "Why Operations Assessments Fail: It's Not Just the Metrics," *Naval War College Review* 64, no. 4 (2011): 89–102. His note 4 offers a comprehensive list of metrics critiques and recommended structures for measuring progress in irregular war.

philosophical choice to more diligently pursue knowledge of the real impact of operations at the tactical level, so that the impact can be interpreted and aggregated at higher levels with much greater confidence. The purpose of this thesis is not to make specific recommendations about *what* to measure, but rather about *how* to measure, in order that we can gain a greater understanding of the specific impact of actions taken within the conflict environment.

C. APPROACH

This thesis employs a mixed-methods approach, using literature review, input from subject matter experts, and a quantitative study of existing survey data to both frame the problem and identify and test a potential solution. The impact assessment demonstration focuses on special operations engagement in Niger, in northwest Africa, where U.S. Special Operations Forces (SOF) have been advising and assisting the Nigerien Armed Forces (known by the French acronym FAN) for several years.[13] The U.S. Special Operations Command Africa (SOCAFRICA) strategy there includes efforts to address root causes of extremism, counter extremist narratives, and improve the links between populations and the legitimate government.[14] Based on these characteristics, the SOF engagement effort in Niger is a good fit to test the utility of impact assessment methodologies.

D. FINDINGS

The primary contribution of this thesis is demonstrating the application of rigorous impact assessment methodology to indirect special warfare campaigns, as an effective means to measure their impact on violent extremist groups. The demonstration has two primary components. The first is the application of the method known as *theory of change*, to describe a logical pathway from non-lethal capacity building to the desired effect of countering violent extremist organizations (C-VEO). I argue that this method,

[13] The former SOCAFRICA commander, BG Donald Bolduc, often said that "We're not at war in Africa, but our African partners are." Donald C. Bolduc, "A View from the CT Foxhole: Brigadier General Donald C. Bolduc, Commander, Special Operations Command Africa," Interview by Brian Dodwell, *CTC Sentinel* 9, no. 5 (May 2016): 8.

[14] Bolduc, "View from the Foxhole," 7-10.

which is commonly used as the foundation for impact assessment in other fields, has wide utility within the special warfare context and would greatly aid in the alignment of special operations with the activities of other government agencies and civilian organizations working in parallel. The second component of the demonstration is the application of impact assessment methods to real data from a special operations capacity-building exercise in order to test their utility in the special warfare context. By restructuring this existing data into a quasi-experimental design and applying analysis methods from the social sciences, I show that impact assessment is productive in identifying effects that were otherwise hidden in observational data—confirming its value for evaluation of special warfare activities.

Although it is not the intention of this thesis to illuminate new information about the efficacy of SOF engagement vis-à-vis the population, the substantive results of the impact assessment are significant. The assessment identifies a nearly equal and opposing impact of capacity-building activities on confidence in U.S. and Nigerien institutions, indicating that civilian populations can respond selectively to the component elements of a coalition, and that an increase in the visible presence of foreign forces may impact attitudes toward the host nation government. This result is both encouraging and frustrating and should provide motivation for further study of the social impact of special warfare engagements and exercises. Better assessment is needed to provide the necessary information to adjust the common understanding that guides these operations, so that the intended effect can be achieved. If the goal is to strengthen the competitive advantage of legitimate governance against violent extremist groups, then capacity-building activities should be focused on increasing military capability and improving popular confidence in that governance. This thesis demonstrates that theory of change methodology and quasi-experimental impact assessment are productive means for evaluating the real effect of special warfare activities and adapting our approach to achieve that focus.

E. THE WAY AHEAD

In Chapter II, I frame the problem in terms of the recent literature on assessment within the DOD generally, and within the context of capacity building and special

warfare more specifically. I then look outside of the defense bureaucracy to other foreign service agencies and the law enforcement community to identify impact assessment as a best practice for understanding program/policy effects.

In Chapter III, I demonstrate how the concept of a theory of change, widely used within the fields listed above, can serve as a foundation for impact assessment. Based on the paradigm of "ungoverned spaces," commonly used by special warfare practitioners to describe the environment where insurgencies and terrorists proliferate, I develop a theory of change that describes the logical pathway from capacity building to desired effect within those spaces.

In order to test the viability of impact assessment methods within the special warfare context, in Chapter IV I conduct an ex post facto, quasi-experimental assessment of the treatment effect of SOF engagement in Niger during the Flintlock exercise in 2015. The assessment uses a difference-in-differences (DD) design to compare the change in institutional confidence among populations in areas that were exposed to SOF engagement (or its descendent effect) to that in areas that were not exposed to the treatment. In order to improve causal inference, I use matching to create treatment and control groups that are similar on a range of potentially disruptive variables.

In Chapter V, I interpret the results of that impact assessment and their implications for the theory of change developed previously. This shows how impact assessment would create a cycle (Figure 1) of improved theory, improved methods and actions, more focused (targeted) effort, and improve the assessment process itself. I then conclude with a series of recommendations based on my findings and suggestions for future research and action.

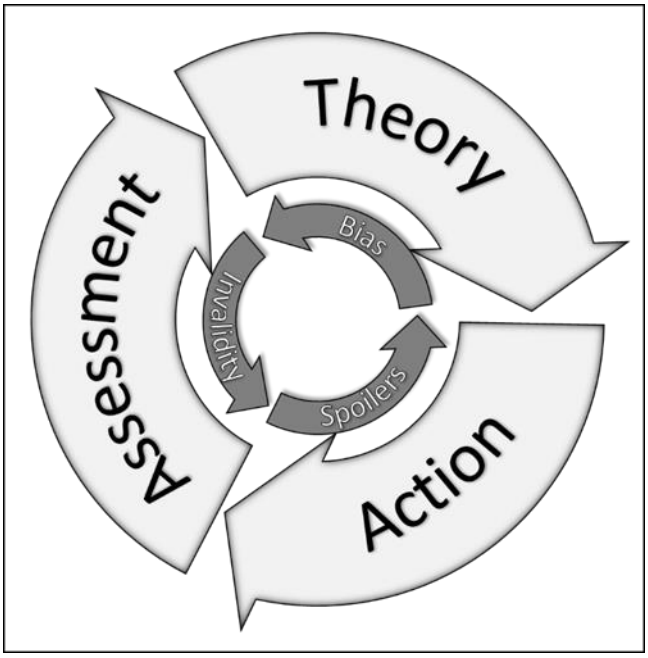


Figure 1. Cycle of theory, action, and assessment

THIS PAGE INTENTIONALLY LEFT BLANK

II. FRAMING THE PROBLEM AND POTENTIAL SOLUTIONS

A. OPERATIONAL ASSESSMENT STRUGGLES IN COUNTERINSURGENCY

The ideas of assessment and evaluation have been a part of the DOD culture for decades and in many ways, the DOD served as a type of clearinghouse for evaluation frameworks and philosophies that expanded into the civil sector.[15] Much of that evaluation was internally focused, related to efficiency for procurement, maintenance, and logistics. More recently, the defense literature on operations assessment has been motivated by the complex operating environments encountered by U.S. and NATO forces during the counterinsurgency campaigns in Iraq and Afghanistan. There is a mountain of writing about the right way and the wrong way to assess these operations and the inadequacies of the military and the larger national security enterprise in assessing operations and campaigns for progress toward established goals. Common culprits for these failures include inadequate doctrine, disengaged or uninterested leaders, a lack of training for planners and practitioners, unclear or overly-ambitious goals, environmental complexity, and even a lack of resources designated for assessment.[16]

The lack of consistent and substantive doctrine on the concept of operational assessment is a common theme among critiques of the efforts in Iraq and Afghanistan.[17]

---

[15] Harry P. Hatry, *Performance Measurement: Getting Results* (Washington, DC: The Urban Institute Press, 1999), xiii.

[16] These concepts will be explored in the following paragraphs. For writing on doctrine, see note 17 below; for training, see note 22; for discussion on the impact of overly-ambitious goals, see Christopher Paul, "Foundations for Assessment: The Hierarchy of Evaluation and the Importance of Articulating a Theory of Change," *Small Wars Journal* 9, no. 7 (2013): 4, and Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken," *Naval War College Review* 64, no. 4 (2011): 103–124; for disinterested leaders, see Schroden, "Why Operations Assessments Fail," 89–102; environmental complexity is a common theme, but a good example is Jason Campbell and Michael O'Hanlon, "Measuring Progress in Iraq," *Wall Street Journal*, 13 July 2007.

[17] Christopher R. Rate and Dennis M. Murphy, *Can't Count It, Can't Change It: Assessing Influence Operations Effectiveness*, (Carlisle Barracks, PA: U.S. Army War College, March 14, 2011); William Gregor, "Military Planning Systems and Stability Operations," Prism 1 (June 2010): 99–114; Jan Osburg, Christopher Paul, Lisa Saum-Manning, Dan Madden, and Leslie Adrienne Payne, *Assessing Locally Focused Stability Operations* (Santa Monica, CA: RAND, 2014); Ben Connable, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency* (Santa Monica, CA: RAND, 2012), http://www.rand.org/pubs/monographs/MG1086.html.

Schroden describes both joint and service doctrine as favoring definitions of assessment categories over prescriptive information on how to actually go about designing and implementing an assessment methodology.[18] He describes joint doctrine as being very thorough in its admonition of the need to conduct assessment and the potential benefits but providing very little guidance on how it could or should be accomplished. His assessment of service doctrine is a little more promising in terms of volume, but not in terms of its quality or utility. He describes FM 5–0, *Operations Process*, as including a significant contradiction by at once discouraging the commitment of valuable time and energy to assessment schemes and then later admonishing the need to "devote the time, effort, and energy needed to properly uncover connections between causes and effects."[19] That level of understanding about the environment and the impact of a unit's actions upon it certainly requires a rigorous assessment methodology.

It is encouraging that there seems to be some recent progress on the inclusion of more rigorous assessment content in joint and service doctrine. Paul et al. cite a number of ongoing efforts to encourage better assessment practice through improved doctrine. This includes the integration of more thorough assessment guidance in Joint Publications 3–0, *Joint Operations*, and 5–0, *Joint Operational Planning*, and a new chapter on assessments in the recent revision of JP 3–13, *Information Operations*.[20] Directly acknowledging the weaknesses of current joint doctrine, the Joint Staff J-7 published a handbook for assessment planning and execution in 2011, much of which has now been included in the most recent revisions of Joint Publication 5–0.[21]

18 Schroden, "Why Operations Assessments Fail."

19 Schroden, "Why Operations Assessments Fail," 92. Quoted text from Department of the Army, *The Operations Process*, FM 5-0 (Washington, DC: Department of the Army, 2010), 6-7.

20 Christopher Paul, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, (Santa Monica, CA: RAND, 2015), www.rand.org/t/RR809z1.

21 Joint Chiefs of Staff, *Commander's Handbook for Assessment Planning and Execution*, (Joint Chiefs of Staff, 2011).

In addition to doctrine, a lack of training is often cited as a reason for failures in operational assessment.[22] This includes confusion about who should be doing assessment and what training they need to do it well. Pre-deployment training for units conducting stability operations and counterinsurgency is wholly lacking on assessment-related training, so capability is completely dependent on the personality and educational background of the individuals involved.[23] At the operational level, the task of assessment is often given to operations analysts. These individuals typically have a strong background in quantitative methods, but may have little understanding of the effects of friendly actions on the environment and how to measure them.[24] They also may lack a strong background in the qualitative approaches that are often required to gain a full picture of what is happening on the ground.[25]

Another argument regarding the difficulties of assessment in Iraq and Afghanistan is that assessments in these conflicts were overly centralized and overly quantitative.[26] Connable argues that this represents a perplexing departure from counterinsurgency doctrine and theory that emphasizes a complex and decentralized operating environment that necessitates a localized approach based on a mission command philosophy. He argues that "incongruity between decentralized and complex [counterinsurgency (COIN)] operations and centralized, decontextualized assessment has led military staffs to rely on ad hoc assessment methods that leave policymakers and the public dissatisfied with U.S. COIN campaign assessments."[27] From these critiques, a number of recommendations have emerged for more balanced approaches to assessing the progress of counterinsurgency and stability operations, where metrics are combined

[22] Schroden, "Why Operations Assessments Fail," 95; Jan Osburg, et al., *Assessing Locally Focused Stability Operations*, 32-34; Rate and Murphy, *Can't Count It, Can't Change It*, 9.

[23] Osburg, et al., *Assessing Locally Focused Stability Operations*, 34.

[24] Schroden, "Why Operations Assessments Fail," 97.

[25] Osburg, et al., *Assessing Locally Focused Stability Operations*, 34.

[26] Connable, *Embracing the Fog of War*.

[27] Connable, *Embracing the Fog of War*, xv.

with localized commander-centric qualitative assessment to gain a more complete understanding of the situation and the progress that is being made.[28]

The approaches and recommendations in these references deal with assessing the over-all progress of a massive and complex campaign. None of them deal with the problem of assigning causality to specific actions taken by the counterinsurgent. Leaders within the irregular warfare (IW) campaigns in Iraq and Afghanistan have often included surveys and social science method in campaign assessments, primarily as a means to measure and aggregate diffuse effects into an assessment of over-all campaign progress.[29] However, the results were not particularly useful because, as Blanken and Lepore observe, the assessments tended to "devolve into simply gauging the environment, with little understanding as to how changes in that environment are the result of military activities."[30] Making that link between components of the campaign and their real impact in the specific social context where they are employed requires the use of more rigorous social science methods and the identification of a counterfactual.

Additionally, the majority of the literature presented to this point is focused on situations where the organization doing the assessment is directly involved in the fighting in a designated theater of armed conflict. What about scenarios in which an indirect, special operations-centric approach is being implemented? These approaches, referred to broadly as special warfare, have been highlighted as a small-footprint, low-cost alternative large-scale military intervention.[31] They rely heavily on special operations

---

[28] Downes-Martin, "Operations Assessment in Afghanistan Is Broken"; Osburg, et al., "Assessing Locally Focused Stability Operations."; Connable, Embracing the Fog of War; Morgan L. Courtney, In the Balance: Measuring Progress in Afghanistan (Washington, DC: Center for Strategic and International Studies, 2005); Gregor, "Military Planning Systems and Stability Operations," 111; Campbell, Jason, Michael E. O'Hanlon, and Jeremy Shapiro, "How to Measure the War," Policy Review 157 (2009), http://www.hoover.org/publications/policy-review/article/5490.

[29] Alejandro S. Hernandez, Julian Ouellet, and Christopher J. Nannini, "Circular Logic and Constant Progress: IW Assessments in Afghanistan," in Assessing War: The Challenge of Measuring Success and Failure, ed. Leo J. Blanken, Hy Rothstein, and Jason J. Lepore (Washington, DC: Georgetown University Press, 2015), 222.

[30] Leo J. Blanken and Jason J. Lepore, "Principles, Agents, and Assessment," in Assessing War: The Challenge of Measuring Success and Failure, ed. Leo J. Blanken, Hy Rothstein, and Jason J. Lepore (Washington, DC: Georgetown University Press, 2015), 8.

[31] Linda Robinson, "The SOF Experience in the Philippines and the Implications for Future Defense Strategy," PRISM 6, no. 3 (2016): 151.

forces as advisors and trainers, and typically include direct support to combat operations conducted by the partner nation in the form of intelligence and logistics support. U.S. military doctrine refers to this support as foreign internal defense (FID), defined as "participation by civilian and military agencies of a government in any of the action programs taken by another government…to free and protect its society from subversion, lawlessness, insurgency, terrorism, and other threats to their security."[32] In most FID operations, foreign forces do not participate in combat operations directly but support the efforts of their host through a variety of security cooperation (SC) tools.[33]

Existing literature includes a number of studies that describe or recommend an assessment framework for components of FID. These studies recommend frameworks for improving the management and assessment of security cooperation and capacity-building activities. Their recommendations largely deal with these functions at the operational and sometimes strategic level to identify how SC and capacity-building programs achieve theater campaign plan and theater security cooperation plan objectives.[34] Others deal exclusively with the organizational structure necessary within the DOD to create a comprehensive approach to security cooperation, recommending the creation of a permanent force structure within the U.S. military (other than special operations forces) that is dedicated to the conduct of FID and other SC missions.[35]

---

[32] Joint Chiefs of Staff, *Foreign Internal Defense*, JP 3-22 (Joint Chiefs of Staff, 2010) ix.

[33] Robinson, "The SOF Experience," 152.

[34] Ross Meyer, "SOF Regional Engagement: An Analysis of the Effectiveness of Current Attempts to Shape Future Battlefields," (master's thesis, Naval Postgraduate School, 2003) http://hdl.handle.net/10945/950; Christopher Paul, Brian Gordon, Jennifer D. P. Moroney, Lisa Saum-Manning, Beth Grill, Colin P. Clarke, and Heather Peterson, *A Building Partner Capacity Assessment Framework: Tracking Inputs, Outputs, Outcomes, Disrupters, and Workarounds* (Santa Monica: RAND, 2015).

[35] Theresa Baginski et al., *A Comprehensive Approach to Improving U.S. Security Force Assistance Efforts* (Carlisle, PA: Strategic Studies Institute, U.S. Army War College, 2009); Scott G. Wuestner, *Building Partner Capacity/Security Force Assistance: A New Structural Paradigm* (Strategic Studies Institute, 2009), 58–59. In 2017 the Army formally created that new force structure by standing-up the 1st Security Force Assistance Brigade.

There have also been several studies that describe and recommend an assessment framework for a specific component of security cooperation or special warfare.[36] An example of this work is the thesis of Leuthner and Cabahug that prescribes a framework for evaluating Joint Combined Exchange Training (JCET) events. Their analysis is telling, because it highlights the total absence of effective measurements for these special warfare components: "From the time of the JCET program's inception, there is no evidence that a deliberate effort has been attempted to evaluate JCET effectiveness. … No evidence yet has been marshaled to support the claim that an executed JCET was successful based on any measure of objective analysis."[37] The framework they recommend is output-based, measuring the performance of the training itself and the specific goals of the authority, in this case that both the U.S. and partner unit receive appropriate and useful training benefit, that participants are properly vetted, and that the training is institutionalized by the partner nation.[38] These measures are important, but they do not address the larger goals that a JCET is being used to achieve.

Within the context of an indirect approach, special warfare campaign, an individual JCET represents a component of a patchwork of training, exercises, and authorities that are stitched together by the Theater Special Operations Command (TSOC) or a special operations task force to create a campaign narrative.[39] Regional exercises often serve as a type of culmination event for partner forces being advised by

---

[36] Scott D. Leuthner and Emmanuel Cabahug, "Joint Combined Exchange Training Evaluation Framework: A Crucial Tool in Security Cooperation Assessment," (master's thesis, Naval Postgraduate School, 2015) http://hdl.handle.net/10945/47809; and Charles Michael Johnson, Jr., *Counterterrorism: DOD Should Enhance Management of and Reporting on Its Global Train and Equip Program*, GAO-16-368 (Washington, DC: Government Accountability Office 2016).

[37] Leuthner and Cabahug, "Joint Combined Exchange Training Evaluation Framework," 55.

[38] For an articulation of this official purpose, see *Examining DOD Security Cooperation: When It Works and When It Doesn't: Testimony Before the Armed Services Committee, House of Representatives*, 114th Congress (2015) (statement of Douglas Fraser, General, USAF (Retired) Former Commander, United States Southern Command), http://docs.house.gov/meetings/AS/AS00/20151021/104083/HHRG-114-AS00-Wstate-FraserUSAFRetD-20151021.pdf.

[39] *Examining DOD Security Cooperation: When it Works and When It Doesn't: Testimony Before the Armed Services Committee, House of Representatives*, 114th Congress (2015) (statement of Christopher Paul, RAND Office of External Affairs), http://docs.house.gov/meetings/AS/AS00/20151021/104083/HHRG-114-AS00-Wstate-PaulC-20151021.pdf. For an example of a SOF-centric indirect approach see David P. Fridovich and Fred T. Krawchuk, "The Special Operations Forces Indirect Approach: Winning in the Pacific," *Joint Forces Quarterly* 44, (2007): 24-27.

U.S. SOF under different authorities, and these exercises themselves are designed to have their own specific impact on extremist groups.[40] The most notable examples of this type of campaign are the efforts of SOF to support Philippine security forces against transnational terrorist groups in the region of Mindanao and a similar effort in Columbia to assist the Columbian armed forces in their fight against the *Fuerzas Armadas Revolucionarias de Columbia* (FARC).[41] SOF took a lead role in both of these FID missions, but they included support from a variety of military and civilian government entities and non-governmental organizations, all operating based on their individual legislated authorities and mandates.[42] Therefore, an understanding of the implicit legislated goals of each component is important, but they do not address the ultimate objectives of the campaign (like countering a violent extremist organization or insurgency) once they are stitched together.

1. Assessment Practice for Capacity Building in Africa

For special warfare efforts on the African continent, many, if not all, of the above conditions and factors are involved in the difficulty of producing consistent and useful assessments of the social impacts of capacity building and other engagement. The official assessment effort focuses almost completely on partner force assessments, with the theoretical goal of an eventual transition once the unit has reached a designated level of proficiency.[43] But assignment of responsibility for the conduct of these assessments makes them unreliable. Generally, individual teams and companies are held responsible for assessing their own performance during a deployment, typically covering a 5 to 9-month period. Predictably, these assessments present almost universal success, are highly subjective, and lean heavily on measures of performance. The generally weak performance of African partner units, and a tendency to rotate forces out of these units

[40] Author discussion with former commander of the Joint Special Operations Task Force – Trans Sahel (JSOTF-TS), July 27, 2017.

[41] For a more detailed examination of the campaign in the Philippines, see Fridovich and Krawchuk, "The Special Operations Forces Indirect Approach;" for Columbia, see Mark Moyar, Hector Pagan, and Wil R. Griego, *Persistent Engagement in Columbia*, (Tampa: JSOU Press, 2014).

[42] Robinson, "SOF Experience," 152.

[43] Author discussion with SOCAFRICA staff member, 25 July 2017.

after they receive U.S. training and equipment, creates a perpetual cycle within the capacity building effort.[44] Constant turn-over of forces creates a patchwork of assessment formats and no long-term connectivity between them.

Policy documents that guide operations, like the civil-military engagement (CME) program that provides authority and funding for Civil Affairs teams deployed on the continent, support this structure by only requiring assessments from rotating teams. The result is years of work with no unifying narrative that connects unit actions with specific outcomes or changes in the environment, and a heavy reliance on anecdotal evidence of effect. These stories, which were repeated by several current and former practitioners I talked with, are generally similar to the one included at the beginning of this thesis, and describe a successful partner unit action that was enabled by popular support. These anecdotes are quite useful and persuasive (great for thesis introductions) and should not be discounted. They should, however, be married with more substantive information that attempts to connect U.S. advice and assistance to the observed effect through a causal pathway, thus creating a more complete assessment and a mechanism to learn and adapt when the effect is different than expected.

B. IMPACT ASSESSMENT IN THE GOVERNANCE, DEVELOPMENT, AND LAW ENFORCEMENT SECTORS

If the current options available in DOD doctrine and assessment practice are insufficient, it is useful to look at other disciplines for alternative methods. In the public sector outside of defense, the trend in performance measurement has shifted in the last several decades from aggregated, jurisdiction-wide data collection and presentation, to a more nuanced study of the results, or outcomes, of specific services or programs.[45] The spheres of development, public diplomacy, community policing, and others have begun to apply the concept of impact assessment to the complex problems they face in order to better understand the linkages between action and impact, program and outcome. These

[44] Lesley A. Warner, *The Trans Sahara Counter Terrorism Partnership: Building Partner Capacity to Counter Terrorism and Violent Extremism* (Alexandria, VA: CNA Corporation, 2014), 74-76.

[45] Hatry, "Performance Measurement," 4.

studies are being done even in conflict environments, making a strong case for potential utility for measuring indirect special warfare efforts.[46]

The design and implementation of these studies have two distinct characteristics. First, the programs they assess are generally designed based on a theory of change or another form of logic model that aids in both the development of relevant interventions and the assessment methodology to evaluate them.[47] According to Paul, "The theory of change for an activity, line of effort, or operation is the underlying logic for how planners think elements of the overall activity, line of effort, or operation will lead to desired results. Simply put, a theory of change is a statement of how you believe the things you are doing are going to lead to the objectives you seek."[48] The process of developing a logic model illuminates the assumptions that are inherent in the process—typically links between cause and effect.[49] These assumptions become hypotheses that can be tested through assessment.

Second, these social science-based studies generally incorporate an experimental or quasi experimental design. Both of these formats test causal hypotheses where the action or intervention is viewed as a treatment that is tested based on the measurement of pre-determined indicators.[50] Such studies are structured to include both a treatment group, which is exposed to the activity or intervention, and a control group (also called a comparison group), which is not. Inferences are made about program impact based on

---

[46] Marie Gaarder and Jeannie Annan, *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*, Policy Research Working Paper No. 6469 (New York: World Bank Independent Evaluation Group, 2013); Christopher Blattman and Jeannie Annan, "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State," *American Political Science Review* 110, no. 1 (2016): 1-17.

[47] Christopher Paul, "Foundations for Assessment: The Hierarchy of Evaluation and the Importance of Articulating a Theory of Change," *Small Wars Journal* 9, no. 7 (2013): 1-7; Jane Reisman and Anne Gienapp, "Theory of Change: A Practical Tool for Action, Results and Learning" prepared for the Annie E. Casey Foundation by Organizational Research Services, 2004; North Atlantic Treaty Organization, *A Framework for the Strategic Planning and Evaluation of Public Diplomacy* (Lisbon, Portugal: Joint Analysis and Lessons Learned Centre, 2013); Sue C. Funnell and Patricia J. Rogers, *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models* (San Francisco, CA: Jossey-Bass, 2011).

[48] Paul, "Foundations for Assessment," 2.

[49] Reisman and Gienapp, "Theory of Change."

[50] Howard White and Shagun Sabarwal, "Quasi-Experimental Design and Methods," *Methodological Briefs: Impact Evaluation 8* (Florence: UNICEF Office of Research, 2014).

comparisons of pre- and post-test measurements. The difference between the two categories is that in experimental designs, the members of the treatment and control groups are chosen randomly. In quasi-experimental designs, the membership in these groups is either selected by the administrator or members are self-selected based on need or location. The control group is selected so that it is as similar as possible to the treatment group so that their comparison identifies the counterfactual—what would have been the outcome if the program or intervention had not been implemented.[51] Table 1 identifies the categories of assessment and evaluation used by other disciplines and compares their characteristics to those used within the DOD. There is no equivalent to impact assessment within the defense assessment doctrine.

Table 1. Comparison of civilian and military evaluation and assessment categories[52]

Civilian Evaluation Categories	**DoD Equivalent**
Impact Evaluation / Impact Assessment	No equivalent
Measures the change in outcome or environment that is attributable to a defined intervention - Based on models of cause and effect - Credible and rigorously defined counterfactual - Randomization creates the strongest evidence	
Performance Evaluation	Measures of Effectiveness
Similar to impact evaluation but without the counterfactual – measures change that is associated with intervention/action	Assess changes in behavior, capability, or environment that indicate achievement of objective or effect (JP 3-0)
Performance Monitoring	Measures of Performance
Ongoing and systematic collection of performance indicator data (inputs, outputs and outcomes)	Assess friendly actions that indicate task accomplishment (JP 3-0)

[51] White and Sabarwal, "Quasi-Experimental Design and Methods," 1.

[52] Adapted from USAID Evaluation Policy, 2011 and Joint Chiefs of Staff, *Joint Operations*, JP 3-0 (Joint Chiefs of Staff, 2017).

1. Evaluating Countering Violent Extremism Effects

The development community has successfully applied impact assessment for CVE programs in ways that may transfer to population-centric approaches in foreign internal defense.[53] According to the U.S. Agency for International Development (USAID) policy on impact evaluation, any new untested development approach must be evaluated before it can be expanded in scale or scope.[54] USAID frequently assesses its programs using studies based on a theory of change linking the interventions to specific outcomes and implements quasi-experimental designs to study those linkages.[55] The results of these assessments are used to adjust programming—focusing on those interventions that are assessed to be working as designed, and rethinking theory or program design where they are not.

The USAID evaluation policy document lays out a number of reasons why rigorous methods are used. Chief among the justifications are the program adjustment mentioned above and the necessity to provide adequate accountability for the public resources expended on these programs. At the time of the policy's release, there had been a sharp decrease in substantive evaluations of the agency's development programs and an increase in public skepticism regarding their utility or effectiveness.[56] It is not unreasonable to think that a similar skepticism could grow surrounding the real impact of the defense department's indirect approach to low intensity conflict. Demonstrating that impact through sound assessment would limit the threat from those critics.[57]

---

[53] US Agency for International Development, *Mid-Term Evaluation of USAID's Counter-Extremism Programming in Africa*, (Washington, D.C.: US Agency for International Development, 2011); Daniel P. Aldrich, "Mightier than the Sword: Social Science and Development in Countering Violent Extremism," (Washington, DC: US Agency for International Development, 2012), http://pdf.usaid.gov/pdf_docs/PNADY453.pdf

[54] *USAID Evaluation Policy*, 9.

[55] Steven E. Finkel, Reynaldo T. Rojo-Mendoza, Cassilde L. Schwartz, Chris A. Balasco, and Anne Kreft, *Impact Evaluation of Peace through Development II (P-DEV II) Radio Programming in Chad and Niger* (Washington: US Agency for International Development, 2015).

[56] *USAID Evaluation Policy*, 1.

[57] For an example of this type of skepticism see Eliza Griswold, "Can General Linder's Special Operations Forces Stop the Next Terrorist Threat?" *New York Times Magazine*, 13 June 2014, https://www.nytimes.com/2014/06/15/magazine/can-general-linders-special-operations-forces-stop-the-next-terrorist-threat.html?_r=0.

In practice, USAID has found that for governance and CVE programming, impact evaluation can be an effective tool for gaining insight into how specific program components are contributing to the ultimate objective. In West Africa, where the agency has a number of large CVE programs in place, they found that impact evaluation is most useful when it is conducted intentionally on a narrow range of activities or on a single program type. Compared to conducting an evaluation of an entire project, which could have a budget of tens of millions of dollars and span multiple countries, a small assessment can incorporate more rigorous controls and ultimately produce a result that provides greater confidence in the implementation of that component in the larger program.[58] Currently, USAID's West Africa mission is designing a study that looks specifically at a type of CVE radio program used to amplify moderate voices. The assessment will include randomized assignment to treatment and include a control group that is not exposed to the radio programming. This will allow the agency to understand the specific impact of those programs on extremist sentiment. Although that result is only truly valid for the specific sample and specific context, the result will either provide greater confidence in the broader effect, or, if the impact is different than expected, the impetus to make changes to content or delivery.

The impact assessment being developed in West Africa is one of a number of similar evaluations supported by the agency's Center of Excellence on Democracy, Human Rights and Governance at their headquarters in Washington.[59] That organization created a learning agenda of 12 primary research questions and sponsors impact evaluation with both funding and academic expertise.[60] The Center's research question about accountability between local government and civil society organizations motivated an impact assessment of a training program for traditional leaders in Zimbabwe. The goal

[58] Author discussion with Director, Regional Peace and Governance Office, USAID West Africa, 19 July, 2017.

[59] U.S. Agency for International Development, "Center of Excellence on Democracy, Human Rights and Governance," accessed September 12, 2017, https://www.usaid.gov/who-we-are/organization/bureaus/bureau-democracy-conflict-and-humanitarian-assistance/center.

[60] U.S. Agency for International Development, "USAID Democracy, Human Rights and Governance Center Learning Agenda" (Washington, DC: U.S. Agency for International Development, 2016), https://usaidlearninglab.org/library/usaid-democracy,-human-rights-and-governance-center-learning-agenda.

of the program was to reduce violent extremism at the community level, but the specific links between program activities and violence were not well understood. With the institutional and funding assistance from Washington, the team designed a randomized control trial using two different training procedures and a control group that did not receive any training. The results showed that the training variant where traditional leaders were trained alongside community leaders was more effective at improving governance indicators. However, the study also found that this arrangement increased social tension and the incidence of political intimidation—an unexpected negative consequence.[61] The information provided by the impact assessment allows policy-makers and program managers to make informed decisions—looking at both the positive and negative impacts to determine whether the benefits outweigh the risks. This productive relationship, for the purposes of evaluation expertise, reach back, and funding, between the operational units in the field and the institutional component in D.C. is one that could be applied to the special warfare community.

2. Impact Assessment and Law Enforcement Strategy

In his book, *The Tipping Point*, Gladwell examines the crime epidemic in New York City in the 1980s, looking specifically at crime on the subway system as representative of that epidemic, and identifying the actions that led, in part, to its end in the 1990s.[62] He gives credit for the dramatic drop in crime to the implementation of a policing strategy known as the "broken windows" approach, first proposed by Kelling and Wilson in a highly influential 1982 article in the *Atlantic Monthly*.[63] In the New York subway system, that strategy involved a number of seemingly minor environmental changes, like the complete removal of all graffiti from the subway cars and the strict enforcement of fare-beating. Gladwell argues that these minor conditions, when

[61] Kate Baldwin and Shylock Muyengwa, *Impact Evaluation of Supporting Traditional Leaders and Local Structures to Mitigate Community-level Conflict in Zimbabwe* (Arlington, VA: Social Impact, 2014).

[62] Malcolm Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference* (Boston: Little, Brown and Company, 2000) 139.

[63] George Kelling and James Q. Wilson, "Broken Windows: The Police and Neighborhood Safety," *Atlantic Monthly*, March, 1982, https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/.

aggregated and unchecked, created a very powerful descriptive norm that had an escalatory effect; according to the police commissioner, "the graffiti was symbolic of the collapse of the system."[64] According to proponents of the approach, the dramatic effort to remove it and end fare-beating changed the environmental context and had a dampening effect on the entire range of criminal activity. This style of policing disorder was then adopted by many police forces across the country because of its reported success in ending the crime epidemic in New York.

The apparent effect of the broken windows approach has useful corollaries to the environmental context of irregular conflict that I explore later in the thesis, but for now, the example provides the opportunity to demonstrate the utility of impact assessment to improve understanding of cause and effect. This is true for two reasons: first, the study of law enforcement is a close a parallel to insurgency and irregular warfare, requiring both coercive and persuasive mechanisms to overcome disorder and violence. Second, much like COIN and IW, policing is often hyper-localized, where methods should be dictated by the social and political context. Despite the need for localized approaches, strategies in both fields are often applied blindly based on apparent success elsewhere. The rapid spread of broken windows policing strategies despite limited empirical evidence of their effectiveness is a good example of this phenomenon.

Gladwell's description of the rapid drop in violent crime in New York is convincing, but his anecdotal evidence is nothing more than an interesting correlation unless it is subjected to a more rigorous assessment. Partly on account of the difficulty of assessing causality, the debate about this type of policing has become highly polarized, and critiques of the strategy are many.[65] Opponents argue that these policies do not address the social and structural root causes of crime and disproportionately target

[64] Gladwell, *Tipping Point*, 142.

[65] Randall G. Shelden, "Assessing 'Broken Windows': A Brief Critique," Center on Juvenile and Criminal Justice, www.cjcj.org/uploads/cjcj/documents/broken.pdf; and R. Panzarella, "Bratton Reinvents 'Harassment Model' of Policing," *Law Enforcement News*, June 15-30 (1998) 13-15; and John E. Eck and Edward R. Maguire, "Have Changes in Policing Reduced Violent Crime?: An Assessment of Evidence," in *The Crime Drop in America*, ed. Alfred Blumstein and Joel Wallman (Cambridge: Cambridge University Press, 2000), 226.

minorities and the disenfranchised.[66] Moreover, the strategy has spawned a whole host of approaches that seem far removed from the original concept: a hyper-focus on misdemeanor arrests, random "stop and frisk" tactics, and aggressive "zero tolerance" policing that breeds distrust between communities and police forces and actually appears to make things more difficult for the police.[67] Relying mostly on observational data, studies have shown that there is no evidence of a link between the misdemeanor arrests that are associated with a broken windows strategy and a decline in serious felony crime.[68]

On the other side of the debate, police officers who are convinced of the effectiveness of the broken windows approach will often structure their responses to these critics from a position of experiential authority, claiming, for example, that "police don't have time for these virtual-reality theories; they do their work in the real world."[69] This kind of response is unproductive and undermines, rather than supports, the credibility of those who are honestly trying to understand what policies will actually prevent violent crime and disorder. Proponents have also published studies to support their claims about effectiveness, but these studies have difficulty assigning causality because of their reliance on observational crime and policing data.[70]

A more useful and reliable means of understanding the impact of these methods is being pursued by criminologists Anthony Braga and Brenda Bond, who summarized the issue this way:

---

[66] Shelden, "Assessing 'Broken Windows,'" 8.

[67] Sarah Childress, "The Problem with 'Broken Windows' Policing," *Frontline*, June 28, 2016, https://www.pbs.org/wgbh/frontline/article/the-problem-with-broken-windows-policing/.

[68] Bernard E. Harcourt and Jens Judwig, "Reefer Madness: Broken Windows Policing and Misdemeanor Marijuana Arrests in New York City, 1989-2000," University of Chicago Public Law and Legal Theory Working Paper, no. 142 (2006): 2; and Mark G. Peters and Philip K. Eure, *An Analysis of Quality-of-Life Summonses, Quality-of-Life Misdemeanor Arrests, and Felony Crime in New York City, 2010-2015* (New York: Office of the Inspector General for the New York Police Department, 2016), http://www1.nyc.gov/assets/oignypd/downloads/pdf/Quality-of-Life-Report-2010-2015.pdf.

[69] William Bratton and George Kelling, "There Are No Cracks in the Broken Windows," *National Review* (February 28, 2006), http://www.nationalreview.com/article/216913/there-are-no-cracks-broken-windows-william-bratton-george-kelling.

[70] George L. Kelling and William H. Sousa, Jr., *Do Police Matter? An Analysis of the Impact of New York City's Police Reforms* (New York: The Manhattan Institute, 2001).

> Given the widespread popularity of broken windows policing, considerable need exists to conduct additional rigorous evaluations of its crime-control effectiveness and to develop some much needed empirical evidence on the key elements of the approach that generate observable preventive benefits.[71]

Their study did exactly that. They created an experimental design and applied broken windows-style policing strategies to randomly selected treatment locations and compared the change in crime and disorder calls for service with control locations where no change in policing strategy occurred. The results of the impact evaluation showed a dramatic improvement in outcome, and comparison of the various component strategies indicated that "the strongest crime prevention gains were generated by situational prevention strategies rather than by misdemeanor arrests or social service strategies."[72] Their experimental research design and rigorous methodology make their causal claims quite strong. Based upon their results, they recommend a problem-oriented policing strategy and a co-design process that incorporates local input, rather than a hyper-focus on misdemeanor arrests.

This example of experimental evaluation of policing strategy is a useful illustration for the measurement of results in special warfare campaigns. Much like the policing example, the impacts of capacity building activities are often highly subjective, and accurate reporting about impact can quickly fall victim to bureaucratic politics and entrenched ideas about cause and effect. For this reason, it is critical that practitioners, who are rightly convinced of the utility of their methods, refrain from *ad hominem* attacks on their critics, and instead adopt an approach similar to that of Braga and Bond.[73] You could replace the activity and the outcome in the quote from the previous paragraph, and it would be thoroughly accurate: *since building partner capacity is a popular component of irregular warfare campaigns,* "significant need exists to conduct

---

[71] Anthony A. Braga and Brenda J. Bond, "Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial," *Criminology* 46, no 3 (2008): 582.

[72] Braga and Bond, "Policing Crime," 578.

[73] Others have adopted this approach as well. Braga recently compiled an inventory of quantitative studies of the broken windows strategy: Anthony A. Braga, Brandon C. Welsh, and Cory Schnell, "Can Policing Disorder Reduce Crime? A Systematic Review and Meta-analysis," *Journal of Research in Crime and Delinquency* 52, no. 4 (2015).

additional rigorous evaluations" *of its effectiveness in countering VEO,* "to develop some much needed empirical evidence on the key elements of the approach that generate observable preventive benefits."[74] All available means should be used to draw specific causal links between those observable impacts and the specific interventions and actions. Studies like the Zimbabwe impact evaluation and the policing study demonstrate that randomized experimental design is possible even in areas where it was thought to be difficult or impossible. In the Zimbabwe study, randomization was achieved by phasing implementation into two years and randomly assigning villages that would receive the training in the first year. Before and after measurement was bracketed around the first year of implementation, creating a natural control group from those villages that were slated to receive the training in the second year.

Even when randomized experimental assessment design in the context of a special warfare campaign is unrealistic, quasi-experimental designs, when combined with modern statistical methods, can simulate the randomized assignment into treatment and control and improve understanding of cause and effect. Lyall, Blair, and Imai demonstrate this possibility with a survey experiment that studies the effects of violence on popular support for ISAF and the Taliban in Afghanistan.[75] Rather than asking respondents directly about their attitudes toward the combatants, they use an endorsement experiment to elicit those attitudes indirectly. Violence (treatment) cannot be controlled by these researchers and is decidedly non-random. So they use a natural experiment structure and matching to approximate that level of control.[76] If the military applied these methods to its own assessments, it would have the additional benefit of a very nuanced understanding of when and why it acted within the environment, increasing the strength of the resulting findings.

To that end, a team of analysts and authors at the RAND corporation have published a series of papers, articles, and monographs that strive to bridge the gap

[74] Italicized text added by the author, quoted text from Braga and Bond, "Policing Crime and Disorder Hot Spots," 582.

[75] Jason Lyall, Graeme Blair, and Kosuke Imai, "Explaining Support for Combatants During Wartime: A Survey Experiment in Afghanistan," *American Political Science Review* 107, no. 4 (2013).

[76] Lyall, Blair, and Imai, "Explaining Support," 689.

between the rigorous, quasi-experimental assessments being employed by the development and civil sectors, and the struggles of the DOD to find adequate methods for assessing complex conflict environments.[77] The most significant of these publications is a comprehensive study that applies the lessons and techniques of academic evaluation of public communication and public diplomacy to defense-sector efforts designed to inform, influence, and persuade. Its prescriptions include the use of logic models and the application of social science experimental design to demonstrate cause and effect.[78] While specifically targeting evaluation of information operations and military information support operations, their recommendations have broad utility for evaluating indirect counterinsurgency, foreign internal defense, and civil-military operations, where campaign success is largely dictated by relationships of trust and legitimacy. This thesis validates that utility through the use of a special warfare impact assessment demonstration.

Prior to conducting that demonstration, the following chapter will introduce the theory of change concept and apply it to the special warfare operating environment.

[77] Paul, "Foundations for Assessment"; Paul, et al., *Assessing and Evaluating: Desk Reference*; Christopher Paul, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners* (Santa Monica, CA: RAND, 2015).

[78] Paul, et al., *Assessing and Evaluating: Desk Reference*.

III. ASSESSMENT CONTEXT: WHAT TO MEASURE AND WHY

In a dusty, West African conference room in 2014, a handful of U.S. special operations advisors described the situation in Niger as "a war taking place in ungoverned spaces… where the Westphalian state can't project power."[79] They described their role in that war as supporting the reinforcement of the foundations of society: governance, development, and security. Their interlocutor, *New York Times* columnist Eliza Griswold, calls the concept of ungoverned spaces a "shibboleth," an in-group identifier whose real meaning is irrelevant for members of the counterterrorism community. In doing so, she casts doubt on the term's validity in describing the political context that leads to insurgency and extremism.[80] But for special warfare practitioners, the concept of ungoverned or under-governed space retains its relevance because it describes the absence of the vital components of the state, a gap that allows sovereignty to be questioned and ultimately challenged—a phenomenon that is becoming more prevalent as civil wars and violent non-state challengers proliferate across the globe.[81]

This thesis explores the utility of impact assessment as a means to better understand how improved capacity within that context affects the relationship between a military force and the civilian populations it is charged with protecting. In order to do so, it is important to first take a step back to examine the context where capacity building occurs, identify some theory that can serve as a framework for how that new capacity impacts the environment, and develop a logic model, also called a theory of change, to describe the path from action to effect and guide assessment of impact. This chapter combines existing research, the unclassified foundational documents from special operations forces, and discussions with current and former practitioners to develop these

---

[79] Griswold, "General Linder's Special Operations Forces."

[80] Griswold, "General Linder's Special Operations Forces."

[81] Arquilla and Ronfeldt describe how the information age favors a dispersed, networked form of warfare that they call "netwar"—a form that gives non-state groups an advantage as they challenge the sovereignty of nation-states that still function using industrial/bureaucratic structures: John Arquilla and David Ronfeldt, "The Advent of Netwar (Revisited)," in *Networks and Netwars: The Future of Terror, Crime, and Militancy,* ed. John Arquilla and David Ronfeldt (Santa Monica, CA: RAND, 2001), 6.

components—providing the necessary theoretical and philosophical foundation for the demonstration of impact assessment methods.

I will start by exploring the idea of ungoverned space, which the practitioners interviewed in Niger identify as a key concept that shapes their understanding of the operating environment. An understanding of the critical functions of the state and its ability to project them into under-governed areas will aid in the development of an assessment strategy for those that are supporting states in the struggle against insurgent, extremist, or other internal threats.

A. THE SOVEREIGNTY GAP: COMPETING FOR CONTROL

Ghani, Lockhart, and Carnahan describe the issue of state control in terms of a "sovereignty gap" between the legal recognition of a state and its de facto ability to exercise that sovereignty within its borders.[82] They propose a set of primary functions that a state should perform within its territory as a type of sovereignty metric. The list is headed by a legitimate monopoly on violence and also includes a number of legal, administrative, economic, and social responsibilities that, when performed in an integrated way, create a powerful reinforcing effect that builds trust in the system as a whole.[83]

These arguments support the need for a focus on areas of contested sovereignty, but doing so in a way that reinforces the sovereignty of the state rather than undermining it. Meierhenrich offers a similar perspective by emphasizing that the authority of states rests in their functional utility and the confidence that it creates in individuals and groups.[84] Meierhenrich identifies six primary functions of what he calls a "usable" state. In addition to concrete concepts like security and resource control, his list includes more abstract ideas such as "displaying resolve" and "lending credibility."[85] Although these concepts might seem unclear, ultimately it is a combination of both real and

[82] Ashraf Ghani, Clare Lockhart, and Michael Carnahan, "Closing the Sovereignty Gap: An Approach to State-Building," *Overseas Development Institute*, Working Paper 253, (September 2005).

[83] Ghani, et al., "Closing the Sovereignty Gap," 9.

[84] Jens Meierhenrich, "Forming States after Failure," in *When States Fail: Causes and Consequences*, ed. Robert I. Rotberg (Princeton: Princeton University Press, 2004).

[85] Meierhenrich, "Forming States," 156.

psychological factors that would lead an individual or group to support an alternative to the state, so these ideas must be part of the equation. States must have the capability to project security, administrative, and economic functions within their borders and create the perception of confidence, predictability, and credibility that ultimately ensures the loyalty of people and groups that are inherently self-interested. Due to these cognitive factors, a state must leverage localized sources of influence and political representation in order to consolidate power in areas of contested sovereignty. The capability to develop, encourage, and ultimately influence populations through civil society organizations is an important component of the "usable" state, particularly when its capability to provide essential services and economic incentives is limited.

The concept of sovereignty and consolidation of state power has been particularly challenging in post-colonial settings. In his book, *States and Power in Africa*, Herbst identifies a historical precedent in Sub-Saharan Africa and other post-colonial settings with low population density, whereby states have developed a strategy of gaining control over core urban or economic zones and then governing peripheral areas based on the availability of resources and security forces to do so.[86] This strategy inherently leaves much of the land-mass in an ungoverned or under-governed status based on a calculation of costs and benefits to the state. In this way, in many post-colonial settings but particularly in Africa, "power still radiates outward from the core political areas and tends to diminish over distance."[87] Despite the apparent fragility of this construct, borders remained fairly static during the first decades of the post-colonial period, and conflict centered on political control of the urban core.

However, the rising tide of state failure and the increasing prominence of successful violent non-state groups demonstrates that the old paradigm is no longer

[86] Jeffrey Herbst, *States and Power in Africa: Comparative Lessons in Authority and Control* (Princeton: Princeton University Press, 2000). Regarding the frequency of state collapse, Herbst states that "although state failure does occur elsewhere, it occurs most often and most dramatically in Africa." (p 262).

[87] Herbst, *States and Power*, 252.

functioning.[88] Herbst identifies a shift that occurred in the 1980s, where post-colonial African states began to be threatened by rural-based insurgencies. The first of these was led by Museveni in Uganda, who drew upon Maoist principles to build a base of support in the hinterland. As a result, these states can no longer count on the assumption of de facto control of the area within their established territorial boundaries.[89] The steady rise of the threat of state-sponsored insurgencies and non-state groups demands a shift away from the traditional notions of security against external threats and a renewed emphasis on areas that were previously left alone due to the high relative cost of their integration.

Kilcullen describes this competitive internal environment and identifies a changing irregular warfare context that is characterized by the growth of massive urban sprawl—at a scale that is quickly outpacing the capability of states to maintain control.[90] While his analysis of the changing conflict environment and the characteristics of what might be called ungoverned space is important, it is Kilcullen's portrayal of the competition among groups within that space that is most relevant to this thesis, challenging the traditional understanding of power dynamics within irregular war. His theory of *competitive control*, based on the writing of Fall and Kalyvas, states that "in irregular conflicts…, the local armed actor that a given population perceives as best able to establish a predictable, consistent, and wide-spectrum normative system of control is most likely to dominate that population and its residential areas."[91] This theory applies to

---

[88] A recent United Nations report identified a three-fold increase in major civil wars in the last decade, many of which have become internationalized, protracted, and complicated by the presence of jihadist groups; Sebastian von Einsiedel et al., *Civil War Trends and the Changing Nature of Armed Conflict* (occasional paper 10, United Nations University Centre for Policy Research, 2017), https://cpr.unu.edu/civil-war-trends-and-the-changing-nature-of-armed-conflict.html.

[89] Herbst, *States and Power*, 254.

[90] David Kilcullen, *Out of the Mountains: The Coming Age of the Urban Guerilla* (Oxford: Oxford University Press, 2013). Exploration of the implications of this changing environment is beyond the scope of this thesis, but Kilcullen's analysis challenges the traditional notions of ungoverned space as a rural phenomenon and describes the growth of a new type of urban hinterland. In these ungoverned spaces, non-state groups, empowered by a vastly increased level of local and transnational connectivity, will compete with the state for control. The implication is that although the default for practitioners is to look for ungoverned spaces in remote areas, these spaces will increasingly be found in the midst of core political zones, surrounded by areas where power is fully consolidated. This may greatly impact the functions of the state that are most critical for expansion of control.

[91] Kilcullen, *Out of the Mountains*, 126; Kilcullen references Bernard B. Fall, "The Theory and Practice of Insurgency and Counterinsurgency," *Naval War College Review*, Winter 1998 [1965]; and Stathis N. Kalyvas, *The Logic of Violence in Civil War* (New York: Cambridge University Press, 2006).

a wide range of non-state elements, from drug cartels to Islamic extremist groups, all competing against the state or simply filling-in where the state has ceded control. Each actor that is competing for control of an area where state sovereignty is in question will create some form of normative system, and whichever group or government does this most effectively is likely to gain de facto control of that space.

In the context of this discussion on consolidation of state control in contested areas, the role of the state must be to outcompete its rivals "across the full coercion-persuasion spectrum, allowing it to establish an uncontested normative system over a given population or territory."[92] In reality, this looks much like the usable state concept presented earlier. These competitive systems must have components that are attractive to the affected population, like conflict resolution mechanisms and the provision of services that are seen as essential in the given context. More importantly in contested areas prone to conflict, however, are the more coercive and persuasive mechanisms that create a sense of security and predictability when the population follows the rules established by the dominant actor.[93]

B. CREATING COMPETITIVE ADVANTAGE: INFLUENCING NORMATIVE CONTEXT AND BEHAVIOR

The concept of ungoverned or under-governed spaces certainly has value to the special warfare practitioner supporting a partner nation and seeking to develop a framework to assess that support. However, it is critical that foreigners who are wading into these spaces acknowledge that they may not be seeing the whole picture—that in the absence of visible state presence, some alternative normative system has likely developed in its place.[94] If the goal of a capacity building effort is to build and support the components of a normative system that can outcompete potential rivals, alternatives, or adversary regimes, what should be the focus of the state and its external sponsors? To answer that question, I argue that there are two primary ingredients for the creation of

[92] Kilcullen, *Out of the Mountains*, 133.

[93] Kilcullen, *Out of the Mountains*, 133.

[94] Zachariah Mampilly, *Rebel Rulers: Insurgent Governance and Civilian Life during War* (Ithaca: Cornell University Press, 2011), 21.

competitive advantage: first, the development of effective institutions and systems in the primary areas of concern, and second, the use of influencing activities and engagements to "bootstrap" the advantage created by those institutions and systems.[95] The focus of the state and its external sponsors should be on the development of institutions and processes that create a sense of security and predictability for affected populations. The emphasis should be placed on the creation of effective security forces that develop strong relationships of mutual trust with civilian populations and support the essential bureaucratic and judicial components for the specific economic and social context—all of these engaging with existing civil society and social structures.

In order to succeed in this effort and build competitive advantage, states and their external sponsors must adopt what Grynkewich calls a "strategy of displacement."[96] His analysis primarily describes the specific scenarios in which a violent non-state group is providing social services, and the elimination of the group will result in the expansion of popular grievances now associated with those services. However, his formulation could be expanded to include the various normative elements of social control, both coercive and persuasive, to include the critical functions of security and justice. It is not sufficient to simply build a suitable security force and eliminate the security threat posed by an adversary group through direct targeting or isolation from the population. Instead, the state must displace the coercive and persuasive functions that allowed the group to gain de facto control or that were creating a stream of recruits and material support.

Apart from the demonstrable construction of responsive and accountable institutions, influence activities can be used to overcome environmental and social obstacles that give non-state competitors an inherent advantage. Cragin and Gerwehr define influence activities as those that "attempt to influence the perceptions, cognitions, and behavior of foreign governments, organizations, groups, and individuals."[97] Although their study focuses on influence campaigns at the strategic level, the spectrum

---

[95] Craig Cohen, "Measuring Progress in Stabilization and Reconstruction," *United States Institute for Peace Stabilization and Reconstruction Series* 1 (March 2008): 8.

[96] Alexus G. Grynkewich, "Welfare as Warfare: How Violent Non-State Groups Use Social Services to Attack the State," *Studies in Conflict & Terrorism*, 31 (2008): 351.

[97] Kim Cragin and Scott Gerwehr, *Dissuading Terror: Strategic Influence and the Struggle Against Terrorism* (Santa Monica, CA: RAND, 2005), 14.

of psychological objectives that they describe is applicable at the operational and tactical levels that are the focus of this thesis. Although there is an inclination to associate influence with psychological operations (now called military information support operations, or MISO), that tendency is misplaced—any time foreigners show up to partner with a host nation force, they are engaged in an activity that will have an influencing effect, whether they intend to or not.

The influence spectrum described by Cragin and Gerwehr is a range of potential effects defined by three primary categories: compliance, conformity, and conversion.[98] The objectives of compliance and conformity are likely to be the primary focus for influence activities in the context of irregular warfare. They align well with the coercive and persuasive components of competitive control: the coercive mechanisms of control result in popular compliance based on demonstrated punitive impacts of defiance, while the persuasive elements change the context so that the desired behavior or allegiance is advantageous to the individual.

This form of persuasive influence is explored by Goldstein and Cialdini, who describe social norms and their influence on human behavior. They describe "the focus theory of normative conduct" as a way to better understand how and when social norms will affect behavior.[99] Focus theory has two central components. The first is the need to separate the two different types of behavioral norms and their distinct effects on behavior: descriptive norms identify things that people *are doing*, while injunctive norms identify what people *ought to do*. The second component of focus theory is that a particular norm only has traction for changing behavior to the extent that it is readily available consciously. The authors argue that in order to change behavior, an audience must be convinced or reminded that a certain behavior is normative—that it is the default mode for the group and the right thing to do—and make that information salient or readily available when a decision to act is made.[100] This suggests that the most powerful

[98] Cragin and Gerwehr, *Dissuading Terror,* 15.

[99] Noah J. Goldstein and Robert B. Cialdini, "Using Social Norms as a Lever of Social Influence," in *The Science of Social Influence: Advances and Future Progress*, ed. Anthony R. Pratkanis (New York: Psychology Press, 2007), 177.

[100] Goldstein and Cialdini, "Using Social Norms," 177.

way to influence individual behavior is to impact the social or communal norms. For example, for individuals who identify strongly with a particular group, or generally in societies where identity is constructed primarily based upon group membership, emphasis on group norms will result in greater adherence to the normative behavior.[101] If a particular society is organized around traditional, familial, or ethnic ties, and positive norms relating to security forces can be identified, then drawing attention to these behaviors could have positive impacts on individuals that are inclined to deviate from that norm.

The dramatic changes made by the New York subway system and police force in the late 1980s and early 1990s, described earlier during the discussion on assessment of law enforcement strategy, demonstrates the link between normative context and individual and group behavior. Gladwell validates the power of descriptive norms—what he refers to as "the Power of Context"—in determining the behavior of individuals and groups.[102] The example shows that for military forces building the capacity of partners fighting irregular conflicts, there should be an emphasis on changing the normative behaviors of their partner forces towards the populations with whom they interact, and vice versa. Advisors should look at communities that are identified as at-risk and identify the normative response to security forces and how those forces behave vis-à-vis the population, its key leadership, and other influential actors or groups. If these interactions are counterproductive, adversarial, or even just apathetic, the focus of the influence campaign and advisory effort should be on reversing that trend. Additionally, the demonstrated power of descriptive norms should motivate practitioners to seek out interventions that acknowledge or leverage their effect. An example of this type of thinking can be seen in a series of recent U.S. government-funded efforts in Niger and Nigeria, where small grants funded brush clearing around villages and roads in areas

[101] Goldstein and Cialdini, "Using Social Norms," 170.

[102] Gladwell, *Tipping Point*, 142.

threatened by Boko Haram.[103] Like the graffiti on the subway, the overgrown brush creates a tunnel effect—a descriptive norm that creates the context of insecurity, whether or not these areas are actually being used by the extremists for concealment. Changing that descriptive norm could have influential effects and actually change the attitudes and behavior of the population in that area. If these are included in a theory of change, their impact can then be tested through evaluation.

As part of the process of assessing operations designed to influence a particular population, practitioners must be aware of the potential for behavioral and descriptive norms to play a spoiling role in those operations. The same awareness will lead to opportunities to identify and leverage normative behavior that can positively impact the link between actions and desired effect. The discussions on logical pathways and the demonstration of impact assessment methods in the next section and chapter add greater weight to these connections.

C. A THEORY OF CHANGE FOR CIVIL-ENGAGEMENT CAPACITY BUILDING

The analysis in the previous sections identified the sociopolitical and environmental context for special warfare campaigns against violent non-state groups. In order to demonstrate the utility of the impact assessment method in that context, I will first identify the specific logical pathway, or theory of change, that will be assessed. In this section, I propose a theory of change that outlines the logical connections from capacity building to the desired effect on populations vulnerable to VEO influence and recruitment. Those connections then form the basis for the impact assessment. The theory is based on the stated goals of SOCAFRICA engagement in the command's unclassified foundational documents, discussions with current and former participants in these efforts, and the analysis conducted in the previous sections. The theory of change identifies how

[103] U.S. Agency for International Development, "Fact Sheet: Office of Transition Initiatives, Niger," August, 2015. https://www.usaid.gov/political-transition-initiatives/niger. The following Facebook post from USAID/OTI provides an example of this type of intervention in Nigeria: https://www.facebook.com/USAIDOTI/photos/a.10151469771197441.1073741827.110642982440/10154217137102441/?type=3.

capacity building activities might lead to the consolidation of control, and what should be measured to demonstrate progress.

In current practice, U.S. forces often build their efforts around binary task-and-purpose statements that do not adequately account for the complexity of the environment or the logical progression from *a* to *b*. For example, a component of non-lethal capacity building efforts might be oriented around a statement like the following:

SOF forces build host nation civil-military engagement capacity [in order to] *counter violent extremist organizations.*

This statement contains several assumptions that are hidden by the phrase "in order to,"—what Paul refers to as "a huge assumptive gap."[104] In his writing he demonstrates that taking this binary statement at face value suggests that you only need to measure the activity (capacity building) and the outcome (VEO strength).[105] This flaw aligns with the existing DOD framework, which calls for measuring performance (activities) and measuring effect (observed change). But what if capacity is built and there is no apparent change in the VEO's level of influence? This seems to be the situation in many places were special operations forces are building partner capacity. Fortunately, just like social interventions of all kinds, any action taken by a military force to affect a change within a complex human environment is based on some theoretical understanding about how that action will lead to the desired effect, whether or not that theory is stated explicitly.[106] So it is possible to expand the task and purpose statement into a clearly articulated theory of change that identifies the appropriate inputs, outputs, and outcomes to measure and also enables the identification of assumptions to test.[107] A basic theory of change for this example could be laid out as follows.

---

[104] Paul, "Foundations for Assessment."

[105] Paul, "Foundations for Assessment." Paul's example is a similar binary statement: "*Training and arming local security guards will lead to increased security."*

[106] C.H. Weiss, "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families," in *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*, ed. J. Connell, A. Kubisch, L. Schorr, and C. Weiss (New York: Aspen Institute), 65-92.

[107] Organizational Research Services, "Theory of Change: A Practical Tool for Action, Results and Learning," 2004, As of March 19, 2014: http://www.aecf.org/upload/publicationfiles/cc2977k440.pdf.

IF we build partner security force capacity to conduct effective and targeted civil-military operations ***AND*** this capability is employed in vulnerable areas, ***THEN*** these vulnerable populations will view their military as more effective, legitimate and capable ***AND*** support and information from these populations for the military force will be increased ***AND*** support for extremist organizations will be reduced ***AND*** VEO recruitment will be reduced ***THEN*** VEO strength and influence is degraded.[108]

Essentially, two additional steps and more specific intermediate outcomes have been added to the task and purpose statement listed above to describe a logical connection between the capacity building and violent extremism. Some of these outcomes would take place simultaneously, so the visual depiction of this theory of change found in Figure 2 is useful in demonstrating this process logically. The progression shows how capacity building that is absorbed and applied in the right areas creates competitive advantage for the state and strengthens their position of normative control. This theory is consistent with Lansdale's view of the effect of what he called "civic action," which was designed to fundamentally change and strengthen the relationship between the people and the government, using the military as a conduit for this change.[109]

[108] This theory of change is sourced from an unpublished manuscript written by the author based upon my own experience working with partners in East and West Africa: Garrett Searle, "Partner Nation Civil-Military Capacity Building: Exploring the Counter-Violent Extremism Effect," unpublished manuscript, 2015. The "If..Then.." logical structure used in this theory of change is a common method of presenting or simplifying the theory into a logical pathway: see Daniel Stein and Craig Valters, *Understanding Theory of Change in International Development* (London: Justice and Security Research Programme, 2012), 4.

[109] Edward Geary Lansdale, *In the Midst of Wars: An American's Mission to Southeast Asia* (New York: Fordham University Press, 1991), 70.

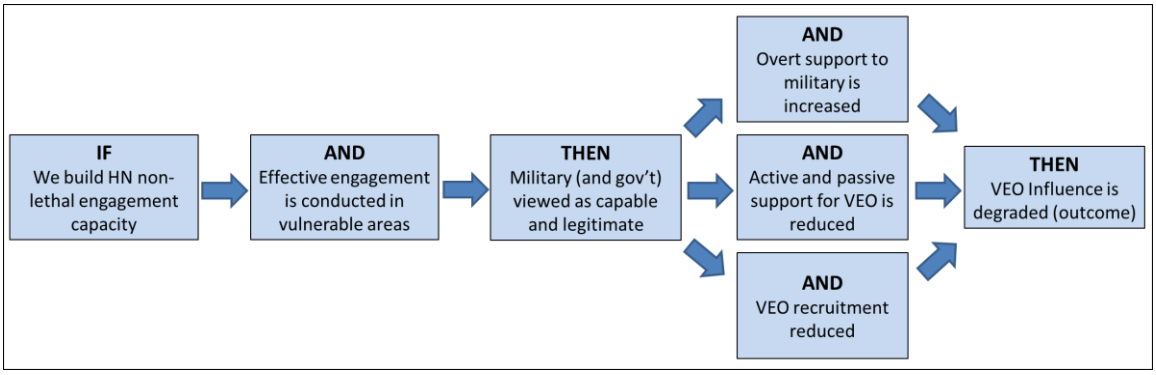


Figure 2. Example theory of change for civil engagement capacity building

It is important to note that any theory of change, including the one presented here, still includes a number of assumptions about the relationships between action and effect, and between the various steps in the model. In each particular case there are a multitude of factors that affect decisions about loyalty, support, or acquiescence.[110] For example, we assume that a population that views their military as effective and capable will be more likely to provide overt support for that military in the form of information and recruits. Another step assumes there is a connection between an individual's evaluation of military competence and governance competence generally. So, it is important to recognize that these are only assumptions, and should be tested through assessment in each specific context.

In order to make the theory of change useful for that purpose, the identified assumptions are then converted into hypotheses for the purposes of evaluation. For the example presented here, I identified six hypotheses that could be tested:

1. Positive and productive engagement by trained military forces in areas vulnerable to VEO influence will produce an increase in positive assessments of military forces as legitimate and capable.

2. Positive assessments of military/police effectiveness are associated with lower levels of extremist sentiment or support.

[110] Kalyvas explores the range of factors that include political, economic, social, and pragmatic justifications for joining or supporting government or insurgent forces: Kalyvas, *The Logic of Violence*, 95-97.

3. Positive assessments of military/police effectiveness are associated with a positive view of legitimate governance institutions more broadly.

4. An improved assessment of military capability and legitimacy will increase a person's likelihood of actively or passively supporting the military (i.e., providing information about VEO activity).

5. An improved assessment of military capacity and legitimacy will lower a person's likelihood of actively or passively supporting a VEO.

6. An improved assessment of military capability and legitimacy will decrease the likelihood of an individual joining or actively supporting a VEO.

None of the statements listed here seem controversial, but a diligent operational approach requires practitioners to acknowledge the fact that they are, indeed, hypotheses, and then go about the business of testing them. The impact assessment demonstration presented in the next chapter will provide an example of how this could be carried out. The demonstration looks specifically at the first hypothesis—testing the connections between the first three "If…And…Then" statements in the theory of change.

The theory of change presented here is not designed to be comprehensive, but merely as a constructive example for the purposes of the demonstration to follow. In fact, the analysis from the previous two sections of this chapter indicate that additional caveats might be necessary or useful to make this theory of change more accurate. For example, it might be necessary to add another "AND" statement in step two to acknowledge the influencing effect of descriptive norms. This might be stated this way: *AND the context of insecurity is reduced*. Also, it is clear from the previous analysis that military capacity must be responsive to specific contextual needs, indicating that an additional "AND" statement might be required. In this case, I include that as part of what would be considered "effective" in the existing model. If that kind of contextual need is ignored, it would certainly serve as a spoiler for other, seemingly productive, engagements.

It is also important to note that this theory of change is not limited to forces, like Civil Affairs, that specifically train, advise, and assist partner forces in civil engagement. Any partner force that interacts with or works alongside of these populations is included in the creation of potential impact. Partner nation counterterrorism forces often look and function like advanced infantry units and lean heavily on human intelligence and other forms of overt support from populations in areas where they are deployed for internal security and counterinsurgency missions. Their interactions with those populations can be critical to success and an important component of the competitive advantage equation. As such, the special operations forces advising them should be cognizant of that important relationship and their potential role in fostering it.

Further research could expand this theory of change to incorporate a full exploration of its potential branches and spoilers. Some authors and organizations conceptualize theory of change as a process design tool that guides participants through the identification of a long-term goal and the incremental conditions necessary to achieve it.[111] Those conditions then become desired outcomes that can be aligned with actions. Future research could adopt this approach and apply design thinking to conceptualize a complete theory of change for capacity building and influence operations to counter violent extremist groups in a specific context.

As is stands, the theory of change articulated here is sufficient to demonstrate the utility of the method as the foundation for impact assessment. The hypotheses developed during the formulation of the theory of change guide the necessary measurement to test their individual validity and the cogency of the entire logical progression. The following chapter will demonstrate how a quasi-experimental impact assessment can be used effectively for that purpose.

[111] Dana H. Taplin and Helene Clark, *Theory of Change Basics: A Primer on Theory of Change* (New York: ActKnowledge, 2012), 1.

IV. IMPACT ASSESSMENT DEMONSTRATION

The theory of change developed in the previous chapter provides the necessary foundation to test the utility of impact assessment as a measurement tool for special warfare capacity-building activities. In this chapter, I demonstrate the application of quasi-experimental assessment design by testing a component of that theory of change in the context of ongoing special warfare activities. For that demonstration, I use real data from U.S. engagement in Niger, where U.S. SOF have been working for several years to strengthen the military's capability to counter VEO. The demonstration uses an ex post facto, quasi-experimental design to assess the treatment effect of SOF engagement during Exercise Flintlock 2015. Specifically, the assessment studies the impact of Flintlock activities in light of the first and third hypotheses from the theory of change: positive and productive engagement by trained military forces in areas vulnerable to VEO influence will increase assessments of military forces and national government as legitimate and capable.

The assessment follows the example of Lyall (2009 and 2010) by using a combination of both matching and difference-in-differences design to identify treatment effects.[112] The assessment compares the change in institutional confidence among populations in areas that were exposed to SOF engagement (or its descendent effect) to those in areas that were not exposed to the treatment. In order to improve causal inference, I use matching to create treatment and control groups that are similar on a range of potentially disruptive covariates. The results of this study clearly demonstrate the utility of these methods by illuminating impacts that were not revealed in the original assessment and improving confidence in the connections between action and effect. Apart from demonstrating the viability of these methods within the context of special warfare, the substantive results of the assessment are valuable in their own right, because they indicate that additional components may be necessary within the theory of change

[112] Jason Lyall, "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya," *Journal of Conflict Resolution* 53, no 3 (June 2009); and Jason Lyall, "Are Coethnics More Effective Counterinsurgents? Evidence from the Second Chechen War," *American Political Science Review* 104, no. 1 (2010).

proposed in Chapter III. The results provide evidence that descriptive norms are spoiling the anticipated effect of the capacity building event and preventing the desired improvement in the state's competitive advantage.

A. BACKGROUND—EXERCISE FLINTLOCK 2015

Exercise Flintlock is an annual Joint Chiefs of Staff-sponsored multinational military exercise, with participants from both African and Western partner nations. The primary goals of the exercise are to foster regional cooperation, develop the counterterrorism capacity of African partner militaries, and counter VEO in North and West Africa.[113] The event is planned and executed every year by SOCAFRICA and hosted on a rotational basis in one of several West African nations. In 2015, the exercise was hosted by Chad, with other training locations spread throughout the region. These other locations, including Niger, were essentially just small surges for existing FID missions.

For the purposes of an impact assessment demonstration, the case provides several useful advantages. First, there is an existing assessment for the exercise based on before and after survey data, covering a two-month period that includes all exercise-related events and other engagements that took place during that time. The existing assessment provides both available data and a useful comparison: by using the same data to draw much stronger conclusions, I am able to clearly demonstrate the utility of the methods. Second, as an initial application of these methods to a special warfare context, the surge of activity associated with the exercise and the relatively short period of time covered by the data provides greater confidence in the estimation of causal relationships. Third, the author was personally involved in the planning and execution of the exercise, and participated in civil-military engagements and other events that occurred in Chad. This personal experience provides additional clarity regarding the objectives and intent of the exercise. It also places the author in a position of personal culpability for any negative

[113] Bardha Azari, "Flintlock '15 Wraps up in N'Djamena, Chad," *United States Africa Command Media Room*, March 09, 2015, https://www.africom.mil/media-room/article/25269/flintlock-15-wraps-up-in-ndjamena-chad.

findings, hopefully making the study feel less like a critique and more like a constructive example of a potentially useful practice.

B. ASSESSMENT COMPARISON AND IMPROVEMENT

1. Original Data

Contracted through the U.S. Special Operations Command's Global Research and Assessment Program (GRAP), ORB International conducted both pre- and post-intervention surveys (Wave 1 and Wave 2, respectively) in Niger, with the first survey in mid-February and the second in mid-April, 2015.[114] ORB distributed the sample for both surveys equally among urban areas in the regions of Niamey, Diffa and Agadez, with a minimum of 450 interviews in each region. The total sample size included 1,368 respondents in wave 1, and 1,371 respondents in wave 2. The surveys were conducted through face-to-face interviews by local interviewers familiar with the customs and language of the respondents.[115] The purpose of the surveys was to assess knowledge of and impact of the Flintlock exercise, as well as media exposure, institutional confidence, extremism, and other social instability factors. The survey responses also include demographic information and the geospatial location of each interview. I used the R statistical program to process the original survey data, filter for non-responses, and create the indexes used as independent and dependent variables in this study.[116] Additionally, I used R to create the geospatial elements and other visuals in this chapter. R has unique advantages over other statistical programs because it is completely open-source and can support a huge variety of statistical and geospatial analysis tasks. Appendix A has more specific information on the construction of the index variables in this study.

---

[114] ORB International, *Flintlock 2015 Survey – Niger Analytic Report*, May 2015.

[115] ORB International, *Flintlock 2015*, 3.

[116] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, (Vienna, 2017), https://www.R-project.org/; Gary King, Christopher Lucas, and Richard Nielsen, "Matching Frontier: R Package for Computing the Matching Frontier," R package version 1.0.0, 2015.

2. Existing Assessment

The existing assessment, based on the ORB surveys, consists of the original survey report and an additional product, created by DigitalGlobe, which combines the survey data with other geospatial information to create a polished report.[117] Unfortunately, the assessment largely misses the mark, focusing heavily on awareness of Flintlock, rather than the impact of its specific activities. The methods used in the assessment do not identify any impact that can be tied directly to the partnered training events and civil-military engagement that was taking place during the period between survey waves. The survey includes questions about Nigerien institutions, but direct impact on trust in those institutions is not assessed (or cannot be assessed using comparisons of descriptive statistics only). The following represents an example of the identification of a potential impact, weakened by the inability to assess causal inference (emphasis added by the author):

> Opinions of the U.S. government and military remained fairly static in all regions except Agadez where opinions of both institutions rose ten points post-Flintlock (U.S. government - 55% to 65%; U.S. military - 57% to 67%). *While it's difficult to account exactly for the positive increase in Agadez from the survey data alone*, it is important to note that a Medical Assistance Activity (MEDSEM) as part of the Flintlock Exercise occurred in Agadez on February 26th.[118]

This example shows the difficulty in assessing causality without the necessary experimental or quasi-experimental structure in place to make those assessments meaningful.

3. Potential Improvements

So what can be done to improve this assessment? My hypothesis is that it is possible to learn much more about the impact of activities like those that occurred during this exercise by associating the observational data—in this case a survey instrument, but it could be a behavior change, event count, or other observable—with specific geospatial information about where those activities took place. This creates a quasi-experimental

---

[117] DigitalGlobe, *Exercise Flintlock 2015 Population Study: Analysis of public awareness and support in Niger regarding Exercise Flintlock 2015*, (Tampa, FL: DigitalGlobe, 2015).

[118] ORB International, *Flintlock 2015 Survey*, 6.

design. The term "quasi" is used because the assignment to treatment and control is not random, as it is in a true experiment, and therefore subject to potential treatment bias. That treatment bias can then be overcome using statistical methods like matching, which I explore more thoroughly later.

Figure 3 shows survey respondent locations in green and blue for wave 1 and wave 2.[119] The locations of capacity building and civil-military engagement that took place between the two survey waves are identified by red circles (exaggerated so that they can be seen on the map). These were geolocated from unclassified reporting from the exercise. In Niger, Flintlock-related events and concurrent training and engagements occurred in Agadez and Diffa regions. In Agadez, SOF teams conducted training with Nigerien military forces, partnered with Nigerien military and civilian entities to conduct a medical outreach event, and conducted key leader engagements with civilian government and traditional leaders. There was also a U.S. Agency for International Development (USAID)-sponsored event at the mayor's office in Agadez designed to improve transparency between local governance and the population. In Diffa region, U.S. SOF teams were training and advising Nigerien forces who were actively fighting cross-border attacks by the extremist group Boko Haram. Also in Diffa region, SOF Civil Affairs teams partnered with Nigerien counterparts during the exercise period to conduct engagements with civilian and traditional leaders in Diffa town, Maine-Soroa, and N'Guigmi.

Zooming into the city of Agadez, Figure 4 shows the geospatial assignment of treatment and control based on the locations of training and engagement events in the treatment window. I chose a one-kilometer radius for assignment to treatment. This choice is somewhat arbitrary, but that distance identifies respondents who have a high likelihood of personal exposure to the exercise activities and engagements. The survey data was not intentionally aligned with exercise event locations, so the one kilometer radius is necessary to incorporate clusters of survey respondents into the treatment group.

[119] All figures were created using the R statistical software. Imagery imported using ggmap function: D. Kahle and H. Wickham, "ggmap: Spatial Visualization with ggplot2," *The R Journal* 5, no. 1 (2013): 144-161, http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Reducing the size of this treatment assignment window did not change the direction of the treatment effect but did reduce the statistical power of the results. Increasing the size of the assignment radius eliminates local control populations and threatens the internal validity of the result. The map in Figure 5 clarifies the assignment to treatment and control—showing the treatment group highlighted in red and the control group in black.

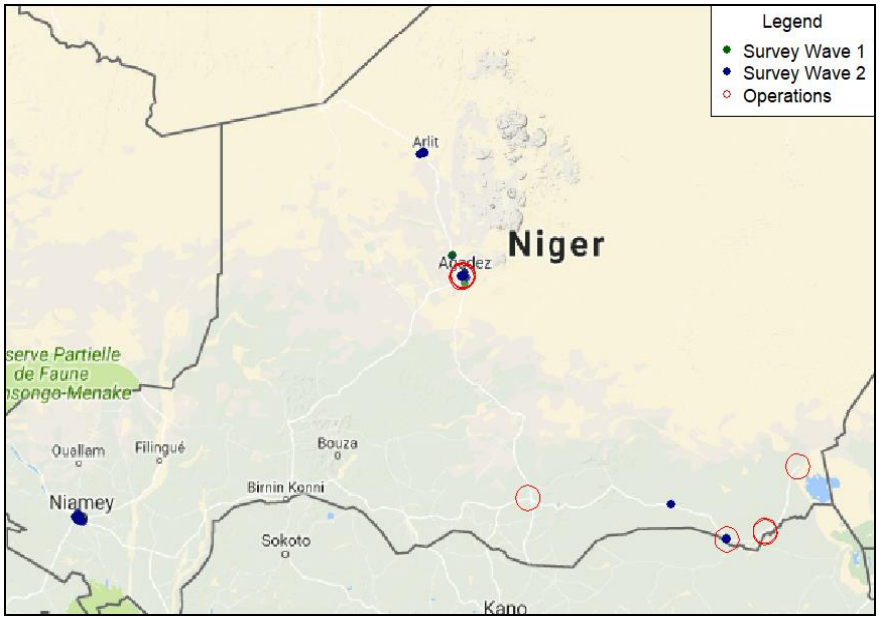


Figure 3. Survey respondent locations and Flintlock events

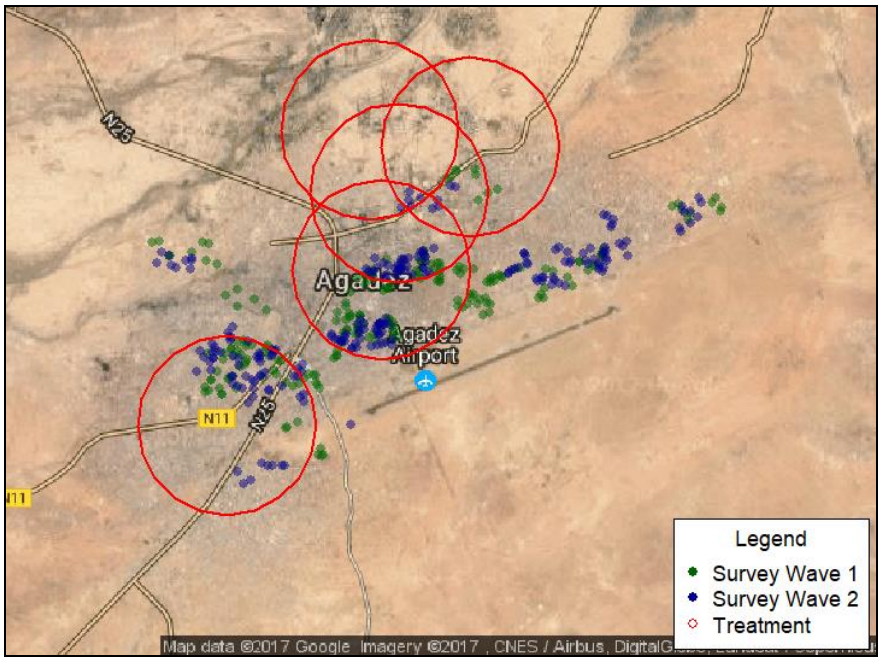


Figure 4. Respondent and Flintlock events detail: Agadez

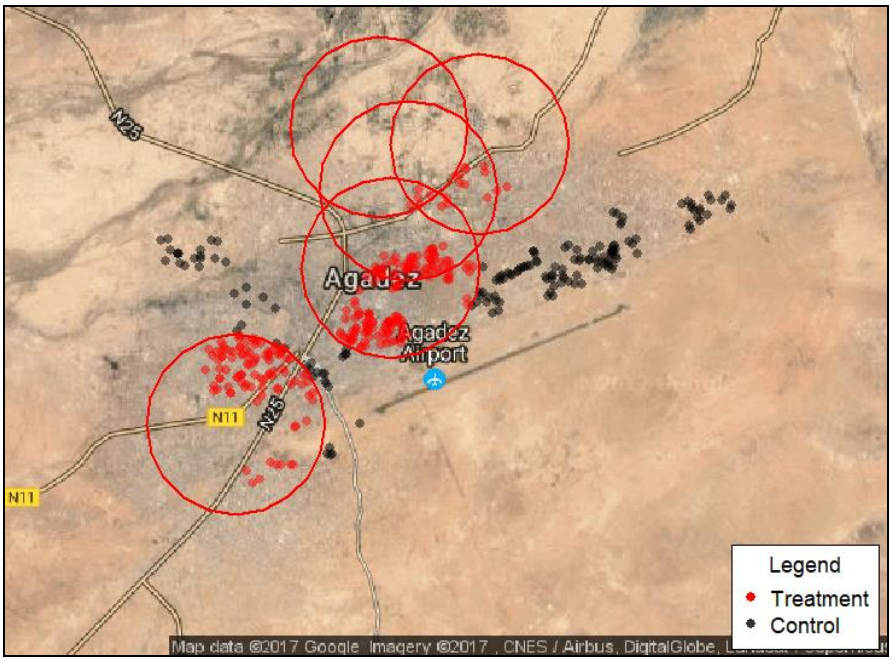


Figure 5. Treatment and control assignment: Agadez

4. Dependent Variables

The demonstration in this chapter assesses the impact of Exercise Flintlock engagements on popular confidence in government institutions, in accordance with the theory of change constructed in Chapter III. That theory posits that if capacity is being built to engage more effectively with vulnerable populations, and that capacity is exercised in areas where those populations live, then the military and government will be viewed as more capable and legitimate. In this assessment, popular confidence in Nigerien governance institutions serves as a measure of the strength of the state's competitive advantage.[120] I also consider the impact of the exercise on confidence in U.S. institutions, because that was highlighted as a "measure of success" in the original assessment.[121] Both of the dependent variables in this study are index variables, consisting of the sum of responses to several questions about confidence in various government institutions, both Nigerien and foreign. The appendix provides more information about these indices.

C. METHODOLOGY

1. Difference-in-Differences

In order to assess the treatment effect, I used a statistical technique known as difference-in-differences—a form that comes from the field of econometrics and is frequently used to assess the impact of various kinds of social intervention.[122] Difference-in-differences is a version of fixed effects estimation for aggregate observational data, and uses the change in outcome for the control as representative of the unobserved counterfactual among the treated population.[123] In this study, since it is impossible to measure the outcome for the treated population if the exercise did not

[120] This philosophical choice is covered more thoroughly in Chapter 3, based on the Kilcullen's theory of competitive control, Kilcullen, *Out of the Mountains,* 126.

[121] ORB International, *Flintlock 2015 Survey*, 6.

[122] For an example of this method in development intervention see U.S. Agency for International Development, *MISTI Stabilization Trends and Impact Evaluation Survey Analytical Report, Wave 5: Sep 28 – Nov 3, 2014* (Arlington, VA: Management Systems International), 330.

[123] Joshua D. Angrist and Jorn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton: Princeton University Press, 2009), 228.

occur, the change in the dependent variable among those that were not exposed to exercise events is assumed to be representative of the change that would have also occurred, in the absence of exercise events, among the entire population. According to Angrist and Pischke, the critical identifying assumption in the difference-in-differences method is that the trends would be the same for both treatment and control in the absence of treatment.[124] Treatment causes a change from this common trend. Although the pretreatment values for treatment and control may be different, the continuation of this difference is accounted for by the treatment group unobserved effect, which accounts for fixed effects among the entire population.[125] This parallel trend assumption is pictured in Figure 6.[126] The short time-frame between treatment and the measurement of outcome in this study makes it particularly well-suited for the use of difference-in-differences causal estimation. Treatment effects demonstrated through difference-in-differences estimation are most reliable when there is a close temporal link between treatment and effect.[127]

[124] Angrist and Pischke, *Econometrics*, 230.

[125] Angrist and Pischke, *Econometrics*, 230.

[126] Columbia University Mailman School of Public Health, "Difference-in-Differences Estimation," accessed June 5, 2017, https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation.

[127] Lyall, "Indiscriminate Violence," 348.

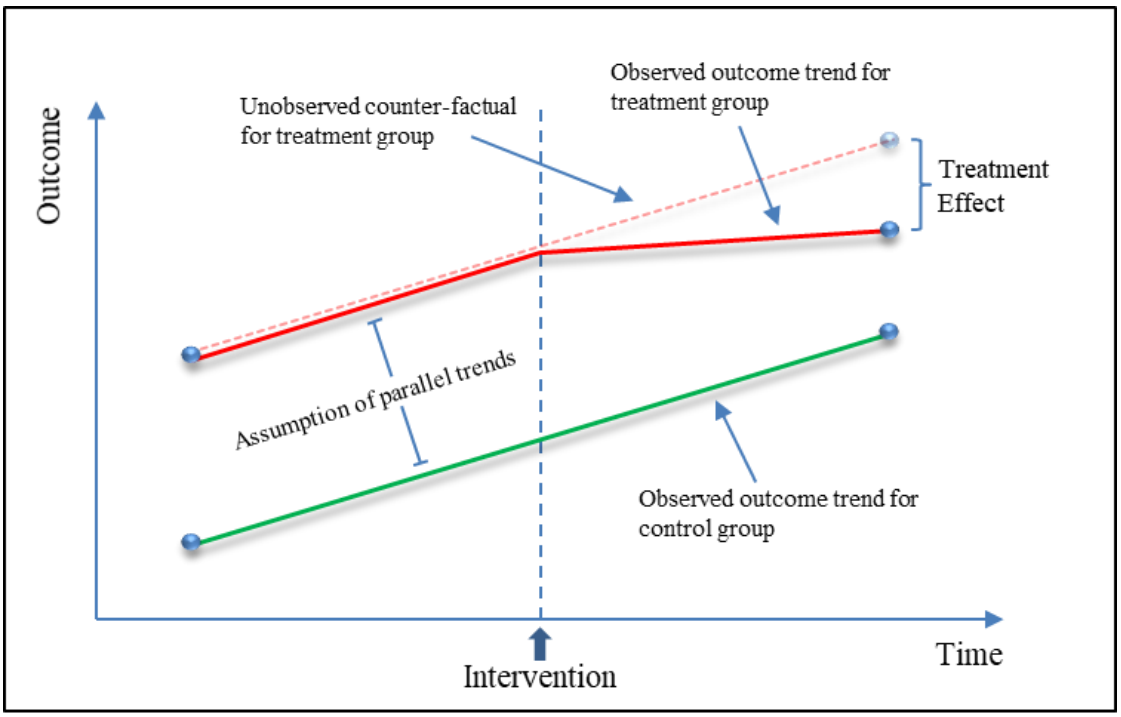


Figure 6. Difference-in-differences concept[128]

In its simplest form, it is possible to calculate difference-in-differences treatment effect using arithmetic, by subtracting the differences in before and after means of both the control group and treatment group, and then subtracting those differences from each other.[129] However, the use of a simple regression formulation allows the addition of covariates and an estimate of standard error and confidence intervals.[130] This formula makes use of one dummy variable (0 or 1) that represents whether a respondent was in the treatment geography and another that indicates whether the respondent was surveyed before or after the treatment. A third variable represents the interaction between these two, and indicates those respondents that are both within the treatment geography and

[128] Adapted from Columbia University Mailman School of Public Health, "Difference-in-Differences Estimation."

[129] Michael Lechner, "The Estimation of Causal Effects by Difference-in-Difference Methods," *Foundations and Trends in Econometrics* 4, no 3 (2010): 172.

[130] Lechner, "Estimation of Causal Effects," 194.

surveyed after treatment. The following represents the difference-in-differences regression model:[131]

$$\mathbf{Y = \beta_0 + \beta_1 * [Treatment] + \beta_2 * [Time] + \beta_3 * [Treatment * Time] + \beta_4 * [Covariates] + \varepsilon}$$

where:

Y is the outcome (dependent) variable.

β_0 is the intercept.

β_1 is the coefficient of the treatment dummy variable and estimates the mean difference between the treatment and control groups prior to the intervention.

β_2 is the coefficient of the time dummy variable that indicates whether the measurement was taken before or after the intervention. It estimates the mean change in outcome during the interval between these measurements among the control group. It serves as a type of fixed-effects control for any change that would have occurred in the absence of the intervention.

β_3 is the coefficient for the interaction term between the time and treatment dummy variables—this is the DD coefficient and estimates the treatment effect. It identifies the difference in the mean change in outcome between the treatment and control groups. This coefficient estimates the impact of the intervention on the dependent variable.

β_4 represents the effect of matched covariate controls. This study employs a total of nine variables in the matching process for age, gender, poverty level, media consumption, sense of security, and the four dominant ethnicities.

2. Matching Frontier

Even with the controls provided by the DD regression design, the very non-random assignment into treatment and control creates a significant threat to internal validity and limits confidence in causal inference. Without randomized assignment, there

[131] Regression equation is adapted from Columbia University Mailman School of Public Health, "Difference-in-Differences Estimation."

is a significant risk of confounding, where the difference in observed outcomes is attributed to the intervention, but that difference is actually caused by another unobserved factor.[132] One can only make an assessment of a causal relationship between treatment and outcome if assignment to treatment is independent of all other factors—as is the case in randomized experiments.[133] Since random assignment of treatment will usually be impractical in the context of special warfare activities, alternative methods must be sought to overcome these threats to validity and gain greater clarity of cause and effect. In natural experiments and designed or ex post facto quasi-experimental studies, like the one described in this chapter, the statistical method known as matching can assist in the creation of treatment and control groups that are essentially identical on all observed covariates—a characteristic known as balance.[134] Matching replicates randomization and reduces the potential impact of confounding by producing treatment and control groups that are only randomly different from one another for the observed covariates.[135] It also greatly reduces model dependence in parametric regression analysis and moderates the large numbers of assumptions necessary in those models and the external information necessary to make those assumptions.[136]

In order for matching to be considered a success, it must achieve greater balance between treated and control groups and also retain a sufficiently large and representative sample useful for estimating effects. King, Lucas, and Nielsen have proposed a method that optimizes the process of achieving balance between these two mandates—what they

---

[132] Elizabeth A. Stuart, Eva DuGoff, Michael Abrams, and David Salkever, "Estimating Causal Effects in Observational Studies Using Electronic Health Data: Challenges and (some) Solutions," *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 1, no 3 (2013): 3, http://repository.edm-forum.org/egems/vol1/iss3/4.

[133] Paul W. Holland, "Statistics and Causal Inference," *Journal of the American Statistical Association* 81, no 396 (1986): 945-960.

[134] Luke Keele, "The Statistics of Causal Inference: A View from Political Methodology," *Political Analysis* 23, no. 3, (2015): 31.

[135] Stuart, et al., "Estimating Causal Effects,"

[136] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis* 15 (2007): 209.

call the "matching frontier."[137] Their method identifies the treatment and control groups with maximum balance for each potential sample size. Visually, the result is a curved line (the frontier) that demonstrates the improvement in balance as the sample size is reduced (Figure 7).[138] Their method also allows the researcher to quickly examine how improved balance and reduced sample size impacts the estimated effects and model dependence.

In order to identify treatment and control groups that can more accurately account for the impact of the Flintlock exercise, I used the six previously identified covariates to construct a Mahalanobis matching frontier for the original sample. This method matches pairs of treatment and control units based upon a distance computation of the difference between the covariates. Balance is improved as the mean of these distances is reduced and respondents without a good match are eliminated from the sample. This reduced sample is then used in the difference-in-differences model to estimate the feasible sample average treatment effect on the treated (FSATT). This designation identifies the fact that any estimated causal effect only applies to those treated observations that have a good match.[139] For simplicity, the FSATT will be referred to simply as the treatment effect.

I used the R statistical program and the associated Matching Frontier statistical package to conduct the data processing, matching, and DD regression.[140] The matching frontier for confidence in Nigerien institutions is shown in Figure 7. The shape of the two frontiers are nearly identical so the second is not shown; the primary difference between them is the size of the original sample—a larger number of non-responses for the second dependent variable (confidence in U.S. institutions) reduced its sample size. For both frontiers, balance improves quickly at first and then continues at a reduced rate as respondents are pruned. Figure 8 shows how this pruning affects the feasible sample.[141]

---

[137] Gary King, Christopher Lucas, and Richard Nielsen, "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," *American Journal of Political Science* 61, no. 2 (2017): 473-489.

[138] King, et al., "Balance-Sample Size Frontier," 24.

[139] Gary King, Richard Nielsen, Carter Coberley, James E. Pope, and Aaron Wells, "Comparative Effectiveness of Matching Methods for Causal Inference," *Unpublished Manuscript* 15 (2011): 3.

[140] R Core Team, *R*; Gary King, Christopher Lucas, and Richard Nielsen. "MatchingFrontier: R Package for Computing the Matching Frontier," R package version 1.0.0, 2015.

[141] The method for production and interpretation of these visuals is derived from King, et al., "Balance-Sample Size Frontier."

Better balance is achieved in a sample that includes a higher percentage of ethnic Hausa. Additionally, the balanced sample has a higher level of exposure to traditional media (radio and television). Conversely, the balancing results in a sample with a lower average age and poverty level, so the balanced sample is younger and better-off than the whole. This process creates a smaller, more balanced sample for assessing impact, creating greater confidence in the resulting analysis but also limiting its applicability. The reduction did not significantly impact the strength of the findings, and, as I will show, only the strength of the impact is changed, not its direction. Despite this rigorous matching process, the presence of unobserved confounding factors is still possible and remains as a limiting factor for non-random experimental design.

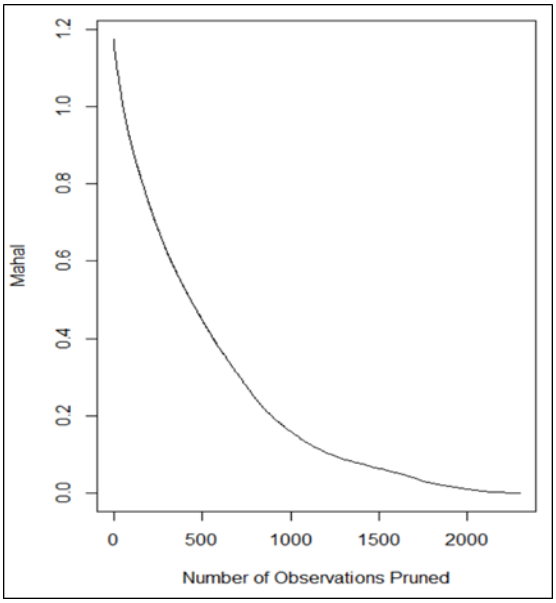


Figure 7. Matching frontier for Nigerien institutional confidence

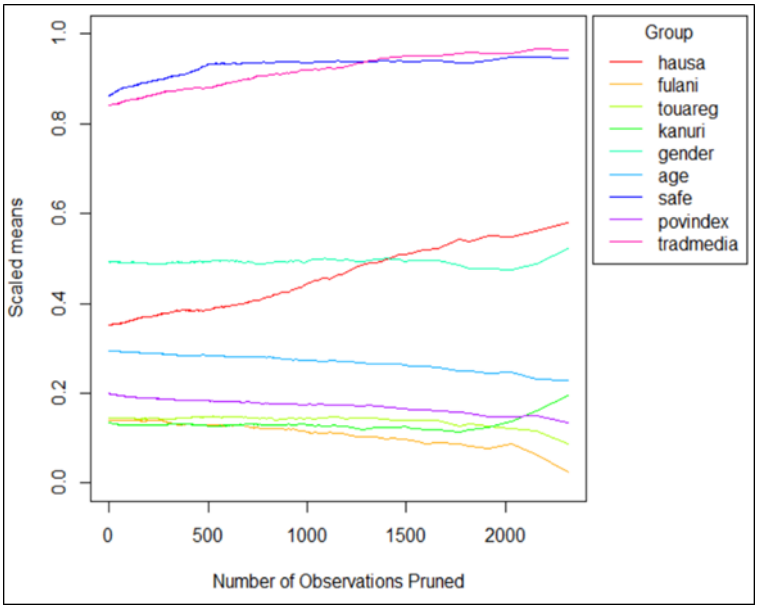


Figure 8. Change in covariate means as balance improves

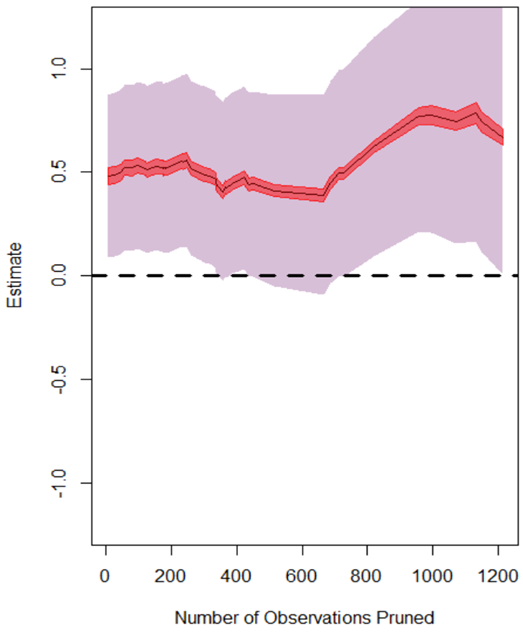
D. RESULTS: ASSESSMENT OF TREATMENT EFFECT

With the Matching Frontier identified, I then applied the difference-in-differences model to the matched dataset to estimate causal effects at each potential sample size. Figure 9a and 9b show the change in the treatment effect coefficient (black line), model dependence (red band), and 95% confidence intervals (purple band) for both dependent variables of interest as the sample becomes more balanced and more closely approximates a randomized experiment. For both dependent variables, model dependence remains fairly static, but the strength of the treatment effect, as indicated by the vertical distance from the dotted 0.0 line (null hypothesis), increases. For the impact on confidence in Nigerien institutions, the statistical power of the sample was not affected by the reduction in size and improved balance. However, for confidence in U.S. institutions, statistical power begins to fray after roughly 1000 observations have been pruned. Based on these characteristics, a balanced sample of roughly 800 observations was selected to identify the FSATT for confidence in Nigerien institutions and a balanced sample of roughly 900 was selected for assessing the impact on confidence in U.S. institutions. The substantive results of the difference-in-differences regression results for

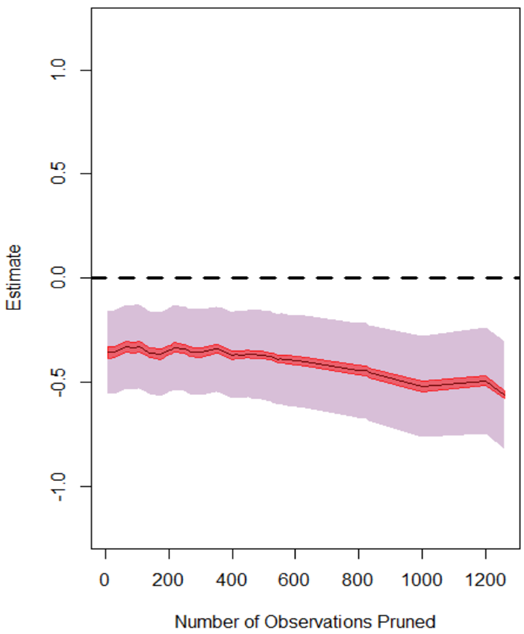
these samples are shown in Table 2. For both of the dependent variables of interest, balancing resulted in the illumination of a stronger treatment effect.

Two interesting and significant findings were identified through the difference-in-differences analysis. The first strengthens an assertion made in the original assessment: the observation of a positive and statistically significant regression coefficient ($p < 0.01$) for treatment effect on confidence in U.S. institutions, indicating that these training events and engagements increased overall confidence in the U.S. government and military (Figure 9c). This identifies, with greater confidence, a positive impact of the exercise: improving the relationship between the United States and a critical regional partner in the fight against violent extremist groups in the Sahel. This improvement in goodwill serves to further U.S. objectives in the region and helps ensure the health of the long-term partnership between these nations.

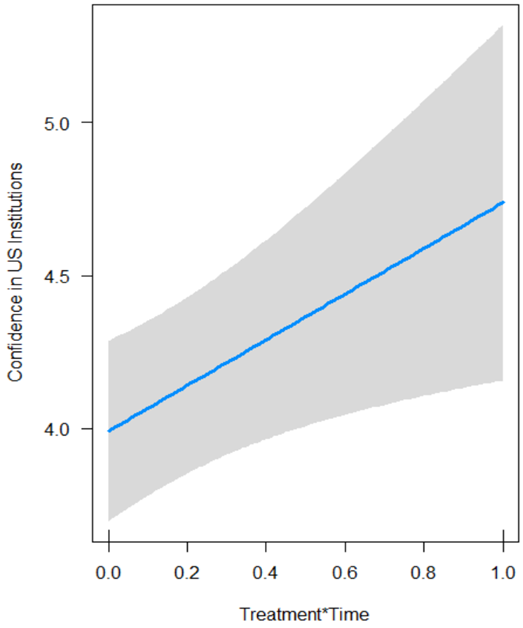
However, this finding is colored by an additional outcome that was not indicated by the original assessment: the observation of a negative and statistically significant regression coefficient ($p < 0.01$) for treatment effect on confidence in Nigerien institutions, indicating that these training events and engagements had a dampening effect on overall confidence in the military and government of the host nation (Figure 9d). This is a disappointing result and indicates that there is a counter-productive impact that needs to be addressed in future planning. Qualitative reporting from the exercise indicated that a combined patrol of exercise participants, both advisors and partner nation military forces, was interpreted incorrectly as an insurgent convoy, inciting a rumor that spread quickly in the town. This incident may have contributed to the result identified in the assessment and it only became visible once the population was separated into treatment and control—as those in closer physical proximity to training events would be more likely to be affected by the rumor.



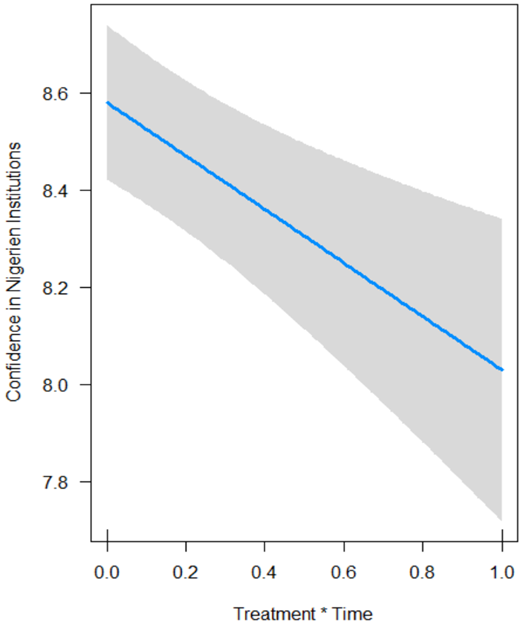
(a) Change in estimated treatment effect as sample becomes more balanced: Impact on confidence in U.S. institutions



(b) Change in estimated treatment effect as sample becomes more balanced: Impact on confidence in Nigerien institutions



(c) Treatment effect coefficient: Impact on confidence in US government and military



(d) Treatment effect coefficient: Impact on confidence in Nigerien government institutions

Figure 9. Treatment effect on institutional confidence

Table 2. Difference-in-differences regression

	Confidence in U.S. Institutions		Confidence in Nigerien Institutions	
	Raw	Matched	Raw	Matched
Treatment Effect	0.463**	**0.771***	-0.353***	**-0.551***
Time	0.662***	0.402**	0.424***	0.577***
Treatment	-0.192	-0.477**	0.282***	0.334**
Constant	2.263***	2.245**	7.024***	6.841***
Controls Included?	Yes	Yes	Yes	Yes
Observations	1,853	897	2,474	817
Log Likelihood	-3,800.150	-1,847.943	-3,821.495	-1,171.442
Akaike Inf. Crit.	7,626.300	3,721.886	7,668.991	2,368.883

$^{**}p < 0.05$ $^{***}p < 0.01$

This result highlights the fact that specific programmatic impact can sometimes be hidden within otherwise promising measures of effectiveness. In this instance, the directional change in mean is actually positive for both variables of interest, seemingly signaling a positive impact. However, the DD method identifies that for confidence in Nigerien institutions, the observed positive impact was actually lower than it *should have been* in the absence of treatment and thus the treatment effect is negative. At the time of the exercise there was a surge in violence perpetrated by Boko Haram in the south. The military responded effectively to those attacks, leading to a general surge in confidence in the government. The assessment appears to show that the exercise had a dampening effect on that surge in confidence. This effect is shown in Figure 10, which inputs the assessment data into the DD concept graph presented earlier. Without the use of quasi-experimental design, this dampening effect is invisible, and may even be interpreted as a positive result.

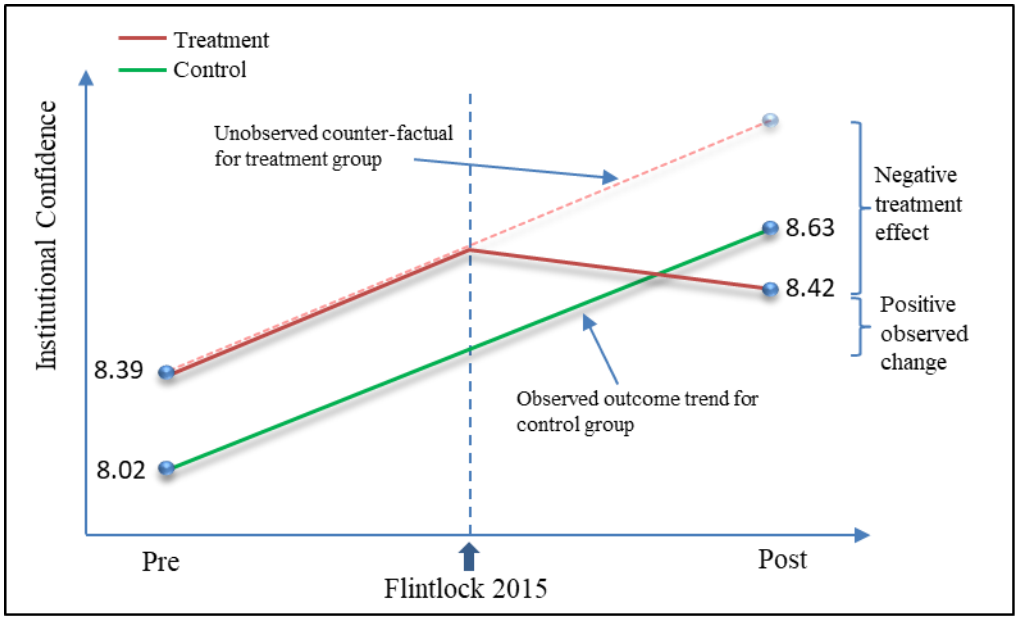


Figure 10. Difference-in-differences for Nigerien institutional confidence

The asymmetry of impact on the two dependent variables in this assessment demonstrates that populations are selective in their responses to the various entities involved in combatting violent extremist threats. The result is consistent with recent studies that examined the role that identity plays in responses to violence in irregular war. Lyall (2010) demonstrated that the identity of soldiers conducting cordon and search operations in Chechnya had a significant effect on patterns of violence following those operations.[142] He found that operations conducted by pro-Russian Chechen forces were more effective than Russian-only sweeps, and that the productivity of Chechen-only forces did not transfer to a mixed force of Chechen and Russian counterinsurgents. Similarly, Lyall, Blair, and Imai (2013) demonstrated that civilian responses to insurgent and counterinsurgent violence in Afghanistan were contingent on the identity of the perpetrator and subject to intergroup bias.[143] They showed that responses to post-violence mitigation and assistance strategies were also dependent on the identity of the

[142] Lyall, "Are Coethnics More Effective Counterinsurgents?" 18.

[143] Lyall, Blair, and Imai, "Explaining Support," 696.

combatant, and that these post-harm aid strategies were especially effective in reducing support for the opponent.[144] Similar to these studies, the results of the Flintlock impact assessment indicate that civilian populations will respond selectively to the component elements of a combatant coalition that includes both incumbent and foreign forces, and that the visible presence of foreign forces may have a normative influencing effect on attitudes toward the incumbent.

I theorize that the large build-up of foreign and local military forces that occurs during an exercise like Flintlock, when compared to the small-footprint, low visibility approach that is typical of the special warfare approach, creates a descriptive norm with counter-productive effect. Essentially, a large presence of foreign forces creates the perception of an insecure environment or a state where domestic institutions are incapable of addressing security concerns on their own. This descriptive norm has the power to change perceptions and behavior and reduces the competitive advantage of the state. This effect is likely made even stronger by the inclusion of events where foreign forces are directly involved in the provision of essential services such as medical care. The fact that the assessment indicated a significant improvement in confidence in U.S. institutions means that the population understood who was responsible for the provision of care, even if attempts were made to conceal that responsibility. The increase in positive perceptions of the U.S. is an important result and should be appreciated in its own right. However, that positive impact does not automatically transfer to the partner government and military forces. Further study could identify mechanisms or engagement strategies that would mitigate this asymmetry of response.

Certainly, this result leads to further questions that can only be answered through a program of designed and deliberate assessment of training results and social impact: Is the relative impact on the units being trained sufficient to overcome a possible setback in institutional confidence? Is the impact on confidence temporary? How can it be reversed? Was it incidental to the rumor and its impact or to the general surge in confidence? These questions should motivate the assessment process for future exercises and steady-state

[144] Lyall, Blair, and Imai, "Explaining Support," 693.

capacity-building operations. For special warfare practitioners, the assessment mandate should be particularly strong in permissive and semi-permissive areas where this type of evaluation is possible, that way the lessons-learned can be applied in non-permissive environments where it is not.

In the following chapter, I will describe how these results affect the proposed theory of change, thus completing the cycle of theory, action, and assessment. I will also make recommendations for the application of these findings to current operational practice and future research and offer some concluding remarks.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY OF FINDINGS

This thesis explored a deficit within current assessment practice for U.S. special warfare efforts and identified the concept of impact assessment as a means to improve causal inference for those activities and campaigns. A survey of other disciplines showed that impact assessment is a productive method to test the theoretical links between program and desired outcome. Those disciplines use randomized controlled trials or quasi-experimental assessments to evaluate the real impact of their interventions and then make iterative changes to program design or implementation. The application of impact assessment represents a philosophical choice to exert the effort necessary to understand how the changes observed in the environment are actually related to unit actions (or not). This thesis demonstrates that it is, indeed, possible to apply the rigorous methods necessary to make those connections, even in austere environments where foreign internal defense and other special warfare activities take place. The impact assessment demonstration was productive in identifying impact that was otherwise hidden in observational data, and it greatly improved understanding of cause and effect by more closely associating observed changes in institutional confidence with the capacity-building activities of special operations forces in the country. This outcome was possible in spite of the weaknesses of the data, encouraging confidence that the results of a designed study would be even more convincing.

B. COMPLETING THE CYCLE: IMPLICATIONS FOR THEORY

Substantively, the demonstration presented in Chapter IV identified a nearly equal and opposing impact of Flintlock events on confidence in U.S. and Nigerien institutions. This result is both encouraging and frustrating and should provide motivation for further study of the social impact of these activities. Returning to the cycle of theory, action, and assessment presented earlier, the impact assessment demonstration shows how assessment can be productive in testing and improving the theoretical understanding of cause and effect in a given environment. The results of that assessment indicate that an

additional caveat (or an additional "AND" statement) should be added to improve the theory. In order to reduce the normative influencing effect, capacity building operations must seek to minimize the presence or visibility of foreign advisors, trainers, and support personnel. Fortunately for proponents of the indirect approach, this minimalist approach represents the standard practice. A large exercise like Flintlock represents a departure from the discreet operational mode of special operations forces advising and assisting foreign forces. The addition of this component to the theory of change, shown in Figure 11, does not dramatically change the logical pathway, but demonstrates how the cycle of theory, action, and assessment functions to improve understanding of cause and effect and limit the impact of spoilers and bias.

The results of the assessment also highlight a critical flaw in the legal foundation of humanitarian civic action like the medical outreach event that occurred in Agadez. The legal authority for civic action, found in U.S. Code, Title 10, Section 401, mandates that U.S. forces participate directly in the provision of services in order to "promote the specific operational readiness skills of the members of the [U.S.] armed forces who participate."[145] That means that U.S. medics and doctors need to treat patients, or U.S. engineers need to be operating machinery. The results of the impact assessment and the asymmetric response to exercise events demonstrates that this policy is counterproductive. In the context of an indirect special warfare campaign, where U.S. forces are assisting a partner nation in the competition for legitimacy, the mandate to participate directly is an obstacle to more effective engagement. Instead, U.S. efforts should focus on supporting and funding civic action planned and executed by the partner military in order to create competitive advantage.[146]

[145] "Humanitarian and Civic Assistance Provided in Conjunction with Military Operations," 10 U.S.C § 401 (2010), https://www.gpo.gov/fdsys/granule/USCODE-2010-title10/USCODE-2010-title10-subtitleA-partI-chap20-sec401.

[146] See Lansdale, *In the Midst of Wars,* 70-71.

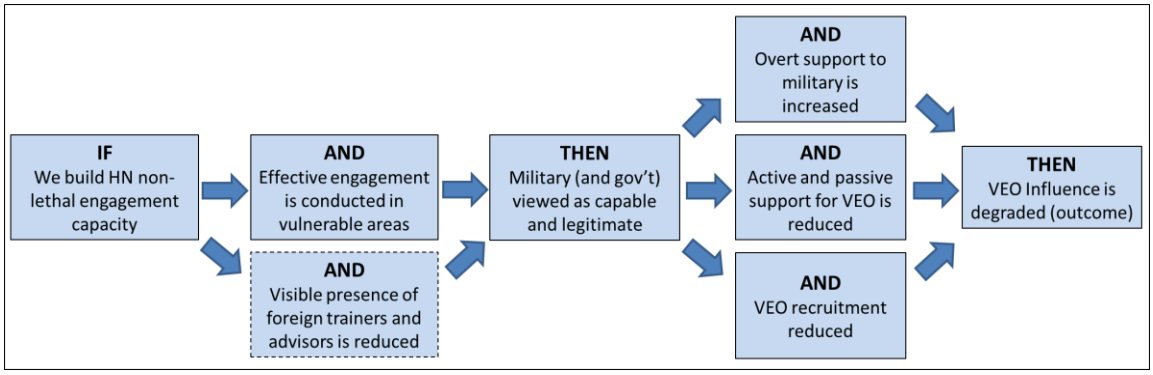


Figure 11. Updated theory of change

C. RECOMMENDATIONS AND FUTURE RESEARCH

1. Formulate a Complete Theory of Change

As mentioned earlier, the logic model for civil engagement capacity building formulated here is incomplete, and future research could expand on this work to develop it further. The complexity of the modern conflict environment leads to divergence within the special warfare community about the correct approach and differing opinions about cause and effect. For the most part, these differences of opinion are built upon the real experiences of practitioners, but also on organizational culture and bureaucratic politics. A research process that leveraged these differences, and used design methodologies to construct a theory of change in reverse, from desired end state backward through intermediate outcomes to necessary actions, would contribute greatly to an improved understanding of cause and effect in special warfare.

Part of that process would be the identification of links to other components and organizations that may be working in the same area. The causal pathways might run in parallel and overlap in certain areas. Awareness of those overlaps provides the ability to leverage one-another to bring about shared intermediate outcomes or end states. For example, consider a scenario where an interagency partner is working in the same area to improve links between vulnerable populations and legitimate governance, using local radio programming to do so. The second step of the theory of change I presented deals with the application of engagement capacity within areas vulnerable to extremism. If a SOF organization is working to build capacity and advising their partner force in the

application of that capacity, then the interagency radio effort could be leveraged to replicate and broadcast positive and productive interactions between civilian leaders and military forces. This would amplify the effect of those engagements, providing mutually beneficial outcomes.[147] A cooperative design process would identify many such opportunities for collaboration and mutual support. At a minimum, special warfare campaigns should be designed using the theory of change method, with each component describing their assumptions about cause and effect and how their actions will contribute to the over-all objective. Then the theory can be used to design an impact assessment program to measure progress toward that goal.

2. Design Assessment Structure Prior to Execution

The primary weakness of the impact assessment presented in this thesis is its reliance on existing data that was not sufficiently aligned with exercise events and not intended to create distinct treatment and control groups for comparison of effect—this structure was created by the author after the fact. If that structure was designed into the assessment from the beginning, it would facilitate even stronger confidence in the resulting analysis by removing some of the geographic bias that is still present in this assessment. The design of such a study could identify matched zones of treatment and control prior to the exercise or training event, similar to the study of law enforcement strategy conducted Braga and Bond.[148] Although that study's randomized assignment of treatment is difficult in the context of capacity building and civil-military engagement, the statistical methods used in this thesis would greatly reduce threats to validity from confounding and treatment bias.

Exercise Flintlock, along with other capacity building events like it, presents a unique opportunity to conduct assessment in a more controlled environment. Exercise planners and trainers exert a tremendous amount of influence over where exercise events

[147] This strategy was adopted during the 2015 Flintlock Exercise in Chad. See U.S. Agency for International Development, *2015 Retrospective: Office of Civilian-Military Cooperation* (Washington, DC: U.S. Agency for International Development, 2016), 4, https://www.usaid.gov/sites/default/files/documents/1866/CMC-2015-Retrospective.pdf (accessed 18 August 2017).

[148] Braga and Bond, "Policing Crime and Disorder," 582.

take place, what the participant forces will be doing, and with whom they will interact from among the civilian populations of the host nation. That level of control means that the possibility exists to conduct rigorous baseline assessment in the precise areas where exercise events will occur, and to identify very well-matched controls in areas where events will not take place. These advantages make Flintlock and other multi-national exercises an ideal venue for testing the components of special warfare engagement, so that those components can be applied with greater confidence in areas where that kind of assessment is difficult or impossible.

3. Strengthen Assessment Requirements and Identify Funding

Policy and authority documents play a critical role in governing the actions taken by deployed forces and those training to deploy for a specific mission. Executive orders and deployment orders for programs of record, like civil-military engagement (CME), should include more substantial instructions on when and where assessment is required. These documents should dictate that any shift in operational approach or area of responsibility requires a baseline assessment, that includes areas where no engagement is planned, to support the structure necessary for impact assessment with before and after measurement and a defined counterfactual. They should also elevate responsibility for continuity of assessment to the staff level at sub-unified commands like a theater special operations command (TSOC) or service component command. Rotational forces would obviously retain a great deal of responsibility in the process, but the staff would ensure that their actions fit into the broader structure that is supported by an assessment design. In the case of special warfare operations in Africa, this level of assessment management is limited to assessments of partner force capability only, and should also include an assessment structure for identifying the impacts of those forces on the environment and human domain. However, that additional responsibility would requirement additional manpower.

In general, policy and authority documents should both mandate more rigorous assessment and provide the funding mechanism necessary to conduct it. These processes require funding to support paying for survey mechanisms and contracted personnel with

the necessary methodological expertise. One method to ensure that assessment is being conducted is to designate a certain percentage of program budget for use on assessment—a technique used by the civilian foreign service agencies. Another method is the use of a centralized assessment contracting mechanism, like the Global Research and Assessment Program, a U.S. Special Operations Command contract designated for assessment of military information support operations. This program could be expanded to approach assessments more holistically, to include the political/social impacts of capacity building, CMO, and support to governance, in addition to the information-related programs that are currently assessed. Also, GRAP could be improved by incorporating quasi-experimental designs in its assessments. The GRAP contract is generally held by a large polling and research firm, like ORB International, that is capable of bringing on the necessary expertise to conduct the type of analysis required by the employer. The methods used in this paper would be within the scope of their expertise, and their implementation simply requires inclusion in the scope of work for a given assessment.

4. Include Impact Assessment in Advanced Training for Special Warfare Practitioners

Within the U.S. special operations community, Civil Affairs forces are given special responsibility for the impacts on human and social components of modern conflict. As a result, CA has a particular responsibility to incorporate the kind of impact assessment methods I have described, but they are not adequately trained to do so. Arguably, based on requirements laid out in their operational authority documents, the Psychological Operations community has done a much better job of incorporating training on assessment practice. For CA, there has been a great deal of discussion regarding the dearth of advanced training available at the Special Operations Center of Excellence at Ft. Bragg, including a number of important recommendations in the 2016–2017 Civil Affairs Issue Papers.[149] I agree with Daniels and Keay, who recommended the incorporation of data analytics for the measurement of structural fragility and its

[149] Melton and Holshek, "Symposium Workshop Report," xlvi.

correlation with conflict prevention activities.[150] Additionally, I would add impact assessment and social science-based evaluation methods to draw stronger linkages between those activities and the observed effect. This advanced training and education is particularly important for officers and NCOs destined for positions in TSOCs and other sub-unified or joint commands. Last year's issue paper summary discusses the draft Army Concept for Civil Affairs and its designation of the regiment as "the lead DOD 'human geography' capability to engage civil societies and agencies by applying unique knowledge, skills, and abilities."[151] Some of that uniqueness should come in the form of distinctive competence in assessment of the social impact of military operations and plans. A more thorough understanding of the design and implementation of strong assessment programs would allow CA staff officers to effectively and accurately communicate the impacts of CA forces operating within their command, and describe how those tactical impacts are linked to operational effects.

5. Narrow Focus for Greater Effect

The assessment demonstration conducted in this thesis showed that concentrated capacity building effort in a relatively small area can produce observable impact. Without an assessment structure in place, it might be tempting for the small number of SOF elements in a particular area of operations to cast a wide net—to maximize the number of productive engagements and generate information from a wide area. But the philosophy of experimental design—thinking in terms of treatment and control—would motivate a more concentrated effort, in order to maximize the observable and attributable effect in a smaller area. This creates the need for more analytical thinking about prioritization of effort and a focus on more specific areas of interest where populations are most vulnerable to extremist influence. Intense focus on indigenous capacity in those specific areas will likely help to create resilient networks that can be expanded into outlying areas when the threat inevitably migrates. This is the ideal operational approach for SOF teams

[150] Clay Daniels and Morgan G. Keay, "Supporting the Trickiest Task: How Civil Affairs Can Bring Essential and Missing Capabilities to Geographic Combatant Command's Mandate to Prevent Conflict," in *2016-2017 Civil Affairs Issue Papers: Leveraging Civil Affairs*, ed. Christopher Holshek, (Carlisle, PA: Peacekeeping and Stability Operations Institute), 22.

[151] Melton and Holshek, "Symposium Workshop Report," xix.

and the requirement to assess the social impact of SOF missions can only encourage its employment.

D. CONCLUSION

The role of leaders in the assessment process cannot be overstated, because leaders have the ability to enable and incentivize a more dedicated pursuit of knowledge about cause and effect and create a culture of learning. If success is the only option, then teams will report successes and ignore things that look like failure, resulting in the loss of countless opportunities to learn from those failures. Astro Teller, the leader of X, Google's so-called moonshot division, has cultivated a culture there that celebrates failure by rewarding teams that identify specific and convincing evidence that their project *will not work.* He describes a balance between "unchecked optimism," and "enthusiastic skepticism," that keeps the organization moving forward productively.[152] For special warfare practitioners, who are rightly optimistic about their methods because they have seen them work in the real world, I think a similar balance is required. They need to pair that optimism with a healthy amount of "enthusiastic skepticism" to identify what is actually working and what is not, and then adapt their theory, actions, and assessments in response.

Impact assessment can provide valuable information for making adjustments to theory and actions so that the intended effect is achieved. If the goal of a partnership is to strengthen the competitive advantage of legitimate governance in the face of violent extremist groups that threaten stability, then SOF must focus the partnership on those activities that improve popular confidence in that governance. The two attacks that were described at the beginning of this thesis clearly demonstrate how productive relationships between security forces and vulnerable populations can tip the balance against insurgent and extremist forces. Rigorous impact assessment methods give practitioners an effective tool to understand the shifting dynamics of those relationships and focus on efforts that bring about a positive result. Without them, it is difficult to effectively communicate

[152] Astro Teller, "The Secret to Moonshots? Killing Our Projects," *WIRED*, February 16, 2016, https://www.wired.com/2016/02/the-secret-to-moonshots-killing-our-projects/

success, and failures and unexpected results will surprise and confound us. Or worse, we will carry on without even knowing the failure occurred.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX. INDEX VARIABLES

A. INDEPENDENT VARIABLES

(povindex) – Poverty Index: combines responses from five questions regarding access to basic commodities (water, fuel, electricity, food, and healthcare). High score = higher level of poverty. Response range is on a scale of 4–16.

Component Questions: D6_A – D6_E. D6 base question: "For each of the following items, please tell me if your household has access to it always, often, rarely, or never. To begin, what about access to water for drinking and cooking—do you always, often, rarely, or never have access? And what about access to?" D6_A: "Water for drinking and cooking." D6_B: "Fuel for heating and cooking." D6_C: "Electricity in the home." D6_D: "Enough food for the whole family." D6_E: "Medical care when needed."

B. DEPENDENT VARIABLES

(instconf) – Institutional Confidence Index: combines responses from three questions regarding confidence in Nigerien government institutions (central government, military, and police). High score = higher confidence. Response range is on a scale of 0 - 9.

Component Questions: L1_A, L1_B, L1_C. L1 base question: "And now please tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of the following institutions." L1_A: "Your national government." L1_B: "Your military." L1_C: "Your police."

(USconf) – U.S. Institutional Confidence Index: combines responses from two questions regarding confidence in United States government institutions (government, and military). High score = higher confidence. Response range is on a scale of 0 - 6.

Component Questions: L1_F, L1_G. L1 base question: "And now please tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of the following institutions." L1_F: "The U.S. Government." L1_G: "The U.S. military."

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Aldrich, Daniel P. "Mightier than the Sword: Social Science and Development in Countering Violent Extremism." Washington, DC: U.S. Agency for International Development, 2012. http://pdf.usaid.gov/pdf_docs/PNADY453.pdf.

Al Jazeera. "Boko Haram Attack in Niger 'Kills 32 Soldiers.'" June 4, 2016. http://www.aljazeera.com/news/2016/06/boko-haram-attack-niger-kills-32-soldiers-nigeria-160604125912477.html.

Angrist, Joshua D., and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press, 2009.

Arquilla, John, and David Ronfeldt. "The Advent of Netwar (Revisited)." In *Networks and Netwars: The Future of Terror, Crime, and Militancy,* edited by John Arquilla and David Ronfeldt, 1–28. Santa Monica, CA: RAND, 2001.

Azari, Bardha. "Flintlock '15 Wraps up in N'Djamena, Chad." *United States Africa Command Media Room*. March 09, 2015. https://www.africom.mil/media-room/article/25269/flintlock-15-wraps-up-in-ndjamena-chad.

Baldwin, Kate, and Shylock Muyengwa. *Impact Evaluation of Supporting Traditional Leaders and Local Structures to Mitigate Community-level Conflict in Zimbabwe.* Arlington, VA: Social Impact, 2014.

Baginski, Theresa, Brian J. Clark, Francis Donovan, Karma Job, John S. Kolasheski, Richard A. Lacquement, Simon D. Roach, Sean P. Swindell, Curt A. Van De Walle, and Michael J. McMahon. *A Comprehensive Approach to Improving U.S. Security Force Assistance Efforts.* Carlisle, PA: Strategic Studies Institute, U.S. Army War College, 2009.

Blanken, Leo J., and Jason J. Lepore. "Principals, Agents, and Assessment." In *Assessing War: The Challenge of Measuring Success and Failure*, edited by Leo J. Blanken, Hy Rothstein, and Jason J. Lepore, 3–15. Washington: Georgetown University Press, 2015.

Blattman, Christopher, and Jeannie Annan. "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State." *American Political Science Review* 110, no. 1, (2016): 1–17.

Braga, Anthony A., and Brenda J. Bond. "Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial." *Criminology* 46, no. 3 (2008): 577–607.

Braga, Anthony A., Brandon C. Welsh, and Cory Schnell. “Can Policing Disorder Reduce Crime? A Systematic Review and Meta-analysis.” *Journal of Research in Crime and Delinquency* 52, no. 4 (2015): 567–588.

Bratton, William, and George Kelling. “There Are No Cracks in the Broken Windows.” *National Review*. February 28, 2006. http://www.nationalreview.com/article/216913/there-are-no-cracks-broken-windows-william-bratton-george-kelling.

Bolduc, Donald C. “A View from the CT Foxhole: Brigadier General Donald C. Bolduc, Commander, Special Operations Command Africa.” Interview by Brian Dodwell, *CTC Sentinel* 9, no. 5 (May 2016): 7–10.

———. *Commander’s Narrative: Ongoing Discussion on Special Operations in Africa*. Stuttgart, DE: Special Operations Command Africa, 2016.

Campbell, Jason, and Michael O’Hanlon, “Measuring Progress in Iraq,” *Wall Street Journal*. July 13, 2007. https://www.wsj.com/articles/SB118429013509165496.

Campbell, Jason, Michael O’Hanlon, and Jeremy Shapiro, “How to Measure the War,” *Policy Review* 157, (2009): 15–30.

Childress, Sarah. “The Problem with ‘Broken Windows’ Policing.” *Frontline*. June 28, 2016. https://www.pbs.org/wgbh/frontline/article/the-problem-with-broken-windows-policing/.

Columbia University Mailman School of Public Health. “Difference-in-Differences Estimation.” Accessed June 5, 2017. https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation.

Connable, Ben. *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*. Santa Monica: RAND, 2012. http://www.rand.org/pubs/monographs/MG1086.html.

Cohen, Craig. “Measuring Progress in Stabilization and Reconstruction,” *United States Institute for Peace Stabilization and Reconstruction Series* 1, (2006). https://www.usip.org/publications/2006/03/measuring-progress-stabilization-and-reconstruction.

Courtney, Morgan L. *In the Balance: Measuring Progress in Afghanistan.* Washington, DC: Center for Strategic and International Studies, 2005.

Cragin, Kim, and Scott Gerwehr. *Dissuading Terror: Strategic Influence and the Struggle Against Terrorism.* Santa Monica: RAND, 2005.

Currey, Cecil B. *Edward Lansdale: The Unquiet American*. Washington, DC: Brassey's, 1998.

Daniels, Clay, and Morgan G. Keay. "Supporting the Trickiest Task: How Civil Affairs Can Bring Essential and Missing Capabilities to Geographic Combatant Command's Mandate to Prevent Conflict." In *2016–2017 Civil Affairs Issue Papers: Leveraging Civil Affairs*. Edited by Christopher Holshek, 15–28. Carlisle, PA: Peacekeeping and Stability Operations Institute, 2017.

Department of the Army. *The Operations Process.* FM 5-0. Washington, DC: Department of the Army, 2010.

DigitalGlobe. *Exercise Flintlock 2015 Population Study: Analysis of public awareness and support in Niger regarding Exercise Flintlock 2015*. Tampa, FL: DigitalGlobe, 2015.

Downes-Martin, Stephen. "Operations Assessment in Afghanistan Is Broken," *Naval War College Review* 64, no. 4 (2011): 103–124.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. *Using Randomization in Development Economics Research: A Toolkit.* London: Centre for Economic Policy Research, 2007. http://www.cepr.org/active/publications/discussion_papers/dp.php?dpno=6059.

Eck, John E. and Edward R. Maguire. "Have Changes in Policing Reduced Violent Crime? An Assessment of Evidence." In *The Crime Drop in America*, edited by Alfred Blumstein and Joel Wallman. Cambridge: Cambridge University Press, 2000.

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

Finkel, Steven E., Reynaldo T. Rojo-Mendoza, Cassilde L. Schwartz, Chris A. Balasco, and Anne Kreft. *Impact Evaluation of Peace through Development II (P-DEV II) Radio Programming in Chad and Niger*. Washington: USAID, 2015.

Funnell, Sue C., and Patricia J. Rogers. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco: Jossey-Bass, 2011.

Fridovich, David P., and Fred T. Krawchuk. "The Special Operations Forces Indirect Approach: Winning in the Pacific." *Joint Forces Quarterly* 44, (2007): 24–27.

Gaarder, Marie, and Jeannie Annan. *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*. Policy Research Working Paper No. 6469. New York: World Bank Independent Evaluation Group, 2013.

Ghani, Ashraf, Clare Lockhart, and Michael Carnahan. “Closing the Sovereignty Gap: An Approach to State-Building.” *Overseas Development Institute*, Working Paper 253, (2005).

Gladwell, Malcolm. *The Tipping Point: How Little Things Can Make a Big Difference*. Boston: Little, Brown and Company, 2000.

Goldstein, Noah J., and Robert B. Cialdini. “Using Social Norms as a Lever of Social Influence.” In *The Science of Social Influence: Advances and Future Progress*, edited by Anthony R. Pratkanis, 167–191. New York: Psychology Press, 2007.

Gregor, William. “Military Planning Systems and Stability Operations,” *Prism* 1, (June 2010): 99–114.

Griswold, Eliza. “Can General Linder’s Special Operations Forces Stop the Next Terrorist Threat?” *New York Times Magazine*, June 13, 2014. https://www.nytimes.com/2014/06/15/magazine/can-general-linders-special-operations-forces-stop-the-next-terrorist-threat.html?_r=0

Grynkewich, Alexus G. “Welfare as Warfare: How Violent Non-State Groups Use Social Services to Attack the State.” *Studies in Conflict & Terrorism* 31 (2008): 350–370.

Harcourt, Bernard E., and Jens Judwig. “Reefer Madness: Broken Windows Policing and Misdemeanor Marijuana Arrests in New York City, 1989-2000.” University of Chicago Public Law and Legal Theory Working Paper, no. 142 (2006).

Hatry, Harry P. *Performance Measurement: Getting Results*. Washington, DC: The Urban Institute Press, 1999.

Herbst, Jeffrey. *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton: Princeton University Press, 2000.

Hernandez, Alejandro S., Julian Ouellet, and Christopher J. Nannini. “Circular Logic and Constant Progress: IW Assessments in Afghanistan.” In *Assessing War: The Challenge of Measuring Success and Failure*, edited by Leo J. Blanken, Hy Rothstein, and Jason J. Lepore, 214–233. Washington, DC: Georgetown University Press, 2015.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.” *Political Analysis* 15 (2007): 199–236.

Holland, Paul W. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81, no 396, (1986): 945–960.

Johnson, Charles Michael, Jr. *Counterterrorism: DOD Should Enhance Management of and Reporting on Its Global Train and Equip Program*. GAO-16-368. Washington, DC: Government Accountability Office, 2016.

Joint Chiefs of Staff. *Commander's Handbook for Assessment Planning and Execution*. Joint Chiefs of Staff, 2011.

———. *Joint Operations*. JP 3-0. Joint Chiefs of Staff, 2017.

———. *Foreign Internal Defense*. JP 3-22. Joint Chiefs of Staff, 2010.

———. *Joint Planning*. JP 5-0. Joint Chiefs of Staff, 2017.

Jones, Howard. *Death of a Generation: How the Assassinations of Diem and JFK Prolonged the Vietnam War.* New York: Oxford University Press, 2003.

Kahle, D., and H. Wickham. "ggmap: Spatial Visualization with ggplot2." *The R Journal* 5, no. 1 (2013): 144–161. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Kalyvas, Stathis N. *The Logic of Violence in Civil War*. New York: Cambridge University Press, 2006.

Keele, Luke. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23, no. 3, (2015): 313–335.

Kelling, George L., and James Q. Wilson. "Broken Windows: The Police and Neighborhood Safety." *Atlantic Monthly.* March, 1982. https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/.

Kelling, George L., and William H. Sousa, Jr. *Do Police Matter? An Analysis of the Impact of New York City's Police Reforms*. New York: The Manhattan Institute, 2001.

King, Gary, Christopher Lucas, and Richard Nielsen. "MatchingFrontier: R Package for Computing the Matching Frontier," R package version 1.0.0, 2015.

———. "The Balance-Sample Size Frontier in Matching Methods for Causal Inference." *American Journal of Political Science* 61, no. 2 (2017): 473–489.

King, Gary, Richard Nielsen, Carter Coberley, James E. Pope, and Aaron Wells. "Comparative Effectiveness of Matching Methods for Causal Inference." Unpublished Manuscript 15 (2011).

Lansdale, Edward Geary. *In the Midst of Wars: An American's Mission to Southeast Asia*. New York: Fordham University Press, 1991.

Lechner, Michael. “The Estimation of Causal Effects by Difference-in-Difference Methods.” *Foundations and Trends in Econometrics* 4, no. 3 (2010): 165–224.

Leuthner, Scott D., and Emmanuel Cabahug. “Joint Combined Exchange Training Evaluation Framework: A Crucial Tool in Security Cooperation Assessment.” Master’s thesis, Naval Postgraduate School, 2015. http://hdl.handle.net/10945/47809.

Lyall, Jason. “Are Coethnics More Effective Counterinsurgents? Evidence from the Second Chechen War.” *American Political Science Review* 104, no. 1 (2010): 1-20.

———. “Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya.” *Journal of Conflict Resolution* 53, no. 3 (2009): 331–362.

Lyall, Jason, Graeme Blair, and Kosuke Imai. “Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan.” *American Political Science Review* 107, no. 4 (2013): 679-705.

Mampilly, Zachariah. *Rebel Rulers: Insurgent Governance and Civilian Life during War*. Ithaca: Cornell University Press, 2011.

McChrystal, Stanley. “COMISAF Initial Assessment (Unclassified).” August 30, 2009. *Washington Post*, September 21, 2009. http://www.washingtonpost.com/wp-dyn/content/article/2009/09/21/AR2009092100110.html.

Melton, Kevin, and Christopher Holshek. “Symposium Workshop Report.” In *2016–2017 Civil Affairs Issue Papers: Leveraging Civil Affairs*, edited by Christopher Holshek. Carlisle, PA: Peacekeeping and Stability Operations Institute, 2017.

Meierhenrich, Jens. “Forming States after Failure.” In *When States Fail: Causes and Consequences*, edited by Robert I. Rotberg. Princeton: Princeton University Press, 2004.

Meyer, Ross. “SOF Regional Engagement: An Analysis of the Effectiveness of Current Attempts to Shape Future Battlefields.” Master’s thesis, Naval Postgraduate School, 2003. http://hdl.handle.net/10945/950.

Moyar, Mark, Hector Pagan, and Wil R. Griego. *Persistent Engagement in Columbia*. Tampa: The JSOU Press, 2014.

ORB International. *2014 NIGER GAP Analytic Report*. London: ORB International, 2014.

———. *Flintlock 2015 Survey – Niger Analytic Report*. London: ORB International, 2015.

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

Paul, Christopher. "Foundations for Assessment: The Hierarchy of Evaluation and the Importance of Articulating a Theory of Change." *Small Wars Journal* 9, no. 7 (2013): 1–7.

Paul, Christopher, Brian Gordon, Jennifer D. P. Moroney, Lisa Saum-Manning, Beth Grill, Colin P. Clarke, Heather Peterson. *A Building Partner Capacity Assessment Framework: Tracking Inputs, Outputs, Outcomes, Disrupters, and Workarounds*. Santa Monica: RAND Corporation, 2015.

Paul, Christopher, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala. "Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference." Santa Monica: RAND Corporation, 2015. www.rand.org/t/RR809z1.

———. "Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners." Santa Monica: RAND Corporation, 2015.

Panzarella, R. "Bratton Reinvents 'Harassment Model' of Policing." *Law Enforcement News*, June 15–30 (1998): 13–15.

Peters, Mark G., and Philip K. Eure. *An Analysis of Quality-of-Life Summonses, Quality-of-Life Misdemeanor Arrests, and Felony Crime in New York City, 2010-2015.* New York: Office of the Inspector General for the New York Police Department, 2016. http://www1.nyc.gov/assets/oignypd/downloads/pdf/Quality-of-Life-Report-2010-2015.pdf.

Organizational Research Services. "Theory of Change: A Practical Tool for Action, Results and Learning." 2004. Accessed March 19, 2014: http://www.aecf.org/upload/publicationfiles/cc2977k440.pdf.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2017. https://www.R-project.org/.

Rate, Christopher R., and Dennis M. Murphy. *Can't Count It, Can't Change It: Assessing Influence Operations Effectiveness*. Carlisle Barracks, PA: U.S. Army War College, 2011.

Reisman, Jane, and Anne Gienapp. "Theory of Change: A Practical Tool for Action, Results and Learning." Prepared for the Annie E. Casey Foundation by Organizational Research Services, 2004.

Reuters. “Niger Forces Kill 57 Members of Boko Haram: Defense Ministry.” April 10, 2017. http://www.reuters.com/article/us-nigeria-security-idUSKBN17C29B.

Robinson, Linda. “The SOF Experience in the Philippines and the Implications for Future Defense Strategy.” *PRISM* 6, no. 3 (2016): 151–167.

Schroden, Jonathan. “Why Operations Assessments Fail: It’s Not Just the Metrics.” *Naval War College Review* 64, no. 4 (2011): 89–102.

Shelden, Randall G. “Assessing ‘Broken Windows’: A Brief Critique.” *Center on Juvenile and Criminal Justice.* Accessed June 6, 2017. www.cjcj.org/uploads/cjcj/documents/broken.pdf.

Searle, Garrett. “Partner Nation Civil-Military Capacity Building: Exploring the Counter-Violent Extremism Effect.” Unpublished manuscript, 2015.

Stein, Daniel, and Craig Valters. *Understanding Theory of Change in International Development.* London: Justice and Security Research Programme, 2012.

Stuart, Elizabeth A., Eva DuGoff, Michael Abrams, and David Salkever. “Estimating Causal Effects in Observational Studies Using Electronic Health Data: Challenges and (some) Solutions.” *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 1, no. 3 (2013). http://repository.edm-forum.org/egems/vol1/iss3/4.

Taplin, Dana H., and Helene Clark. *Theory of Change Basics: A Primer on Theory of Change.* New York: ActKnowledge, 2012.

Teller, Astro. “The Secret to Moonshots? Killing Our Projects.” *WIRED*, February 16, 2016. https://www.wired.com/2016/02/the-secret-to-moonshots-killing-our-projects/.

U.S. Agency for International Development. *2015 Retrospective: Office of Civilian-Military Cooperation.* Washington, DC: U.S. Agency for International Development, 2016. https://www.usaid.gov/sites/default/files/documents/1866/CMC-2015-Retrospective.pdf.

———. “Center of Excellence on Democracy, Human Rights and Governance.” Accessed September 12, 2017. https://www.usaid.gov/who-we-are/organization/bureaus/bureau-democracy-conflict-and-humanitarian-assistance/center.

———. “Fact Sheet: Office of Transition Initiatives, Niger.” August 2015. https://www.usaid.gov/political-transition-initiatives/niger.

———. *Mid-Term Evaluation of USAID's Counter-Extremism Programming in Africa*. Washington, DC: U.S. Agency for International Development, 2011.

———. *MISTI Stabilization Trends and Impact Evaluation Survey Analytical Report, Wave 5: Sep 28 – Nov 3, 2014*. Arlington, VA: Management Systems International. http://pdf.usaid.gov/pdf_docs/pa00ks3x.pdf

———. "USAID Democracy, Human Rights and Governance Center Learning Agenda." Washington, DC: U.S. Agency for International Development, 2016. https://usaidlearninglab.org/library/usaid-democracy,-human-rights-and-governance-center-learning-agenda.

———. *USAID Evaluation Policy*. Washington, DC: U.S. Agency for International Development, 2011 (Updated 2016). www.usaid.gov/evaluation/policy.

von Einsiedel, Sebastian, Louise Bosetti, James Cockayne, Cale Smith, and Wilfred Wan. *Civil War Trends and the Changing Nature of Armed Conflict.* Occasional paper 10, United Nations University Centre for Policy Research, 2017. https://cpr.unu.edu/civil-war-trends-and-the-changing-nature-of-armed-conflict.html.

War is Boring. "U.S. Military Setting up Shop in Niger." August 29, 2015. https://warisboring.com/u-s-military-setting-up-shop-in-niger-8824462bb365#.f73o746e9.

Weiss, C.H. "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families." In *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*, edited by J. Connell, A. Kubisch, L. Schorr, and C. Weiss, 65–92. New York: Aspen Institute, 1995.

White, Howard, and Shagun Sabarwal. "Quasi-Experimental Design and Methods." *Methodological Briefs: Impact Evaluation* 8 (2014).

Wuestner, Scott G. *Building Partner Capacity/Security Force Assistance: A New Structural Paradigm.* Carlisle, PA: Strategic Studies Institute, 2009.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California