T H E   EES   N E W S L E T T E R

EVALUATION
CONNECTIONS

Jörg Faust Guest editor. **Ole Winckler Andersen** Associate Editor. **Tereza Zvolská** Editorial Assistant.

EES
EUROPEAN
EVALUATION
SOCIETY

A U G U S T   2 0 1 6

A R T I C L E S   –   N E W S   –   E V E N T S

CONTENT

PRESIDENTIAL MESSAGE

Riitta Oksanen

Dear EES members and friends,

In order to maintain its policy relevance evaluation needs to evolve in response to changes in the operating environment. Accordingly, this special issue of your newsletter focuses on peace and security – an increasingly salient evaluation challenge.

The articles that follow demonstrate that evaluation practice in this domain needs continuous development. Global peace and prosperity are now tightly connected and the dynamics of conflict and development are not the same as they were 15 years ago. Increased uncertainty, volatility and complexity are felt everywhere, including in Europe. This means that new evaluation approaches and practices are needed especially since many of the evaluation challenges triggered by insecurity and conflict characterize other sectors as well.

What role and focus should a voluntary organization for professional evaluation, such as the EES, strive towards in this turmoil? This topic will be included in the EES agenda in 2016 as the EES board engages in consultations about the future role and priorities of your Society. We plan to involve the full membership in strategy sessions at

the forthcoming Maastricht Conference, and beyond.

These deliberations will be informed by an extraordinarily rich and diverse set of Conference presentations. I look forward to a fulsome debate and to further on-line interactions following the Conference. This will give us a sound foundation to design an EES Action Plan focused on the following issues:

- What it the appropriate role of the EES: should we focus on providing fora for discussions as conveners; knowledge brokers and/or standard setting?
- Should we adopt an activist advocacy role and should we proactively promote innovation to tackle emerging issues?
- What topics and themes should we focus on? How should they be selected?
- What can we practically and realistically achieve as a voluntary organization?
- How can we more actively involve in initiatives most relevant to the European evaluation community, with and by the membership?

I look forward to welcoming you in Maastricht.

Riitta Oksanen
President of the European Evaluation Society

EDITORIAL: CONFLICT PREVENTION AND PEACE-BUILDING: WHAT ROLE FOR EVALUATION?

Ole Winckler Andersen and Jörg Faust

In spite of guidelines for evaluation of peace-building and conflict prevention (see e.g. OECD, 2012) many challenges remain. First, the assessment of relevance of intervention goals and designs hinges on an understanding of the conflict contexts, which is hard to come by. The choice of methods to assess effectiveness and efficiency is equally demanding. Finally, the assessment of sustainability and impact calls for a combination of evaluations at the levels of the intervention: the local context, the country and region.

Underlying these challenges is a particular disconnect between the research community dealing with peace and conflict on the one hand and the evaluation community on the other. In principle, similar questions are also important for other areas of evaluation, but they seem to be more prominent in the field of conflict prevention and peace-building. The current European refugee crisis has given the relevance of these issues a new dimension with questions around the effectiveness of development policy and humanitarian action in fragile contexts having entered the center of domestic political debate about refugee policy in Europe.

Against this background, the papers assembled in this special issue of Connections provide various perspectives on these questions. They point to rich debates about how to go about evaluation in conflict affected environments. The papers provide insights about the feasibility of micro-level approaches to impact evaluation, and they amplify conceptual and theoretical challenges of evaluation practice at micro and macro-level.

Primarily from a methodological perspective, Trichler, BenYishay and Parks show how geo-spatial impact evaluation methods provide a highly relevant tool to overcome some of the traditional challenges of getting physical access to conflict areas and of analyzing relevant data for outcome and impact measurement. The authors argue that this approach can lead to a faster completion of evaluations at lower financial costs without negative quality implications. Also at the micro-level of impact evaluations, Banholzer illustrates achievements and challenges regarding the evaluation of disarmament, demobilization and re-integration programs. She strongly argues for a combination of quantitative and qualitative methods and calls for more thoroughly working on theory driven approaches from which context-specific intervention-logics for interventions can be deduced. She also discusses some key design issues, including various potential biases as well as using cross-section comparisons and longitudinal analyses.

The paper by Zürcher presents the results of a systematic review, which tackles the particular relevant question of whether foreign aid contributes to violence in conflict-affected environments. The paper concludes that it depends on the context whether aid dampens or increase violence, and that aid is best allocated to areas with some stability.

In the second batch of papers Paffenholz criticizes that theories of change are often not made explicit in peacebuilding programs or remain comparatively vague. Based on her own experience, she outlines an approach on how the theory of change can be reconstructed and assessed. She illustrates the approach with an example from civil society peace building initiatives. Chigas and Goddard in their paper provide arguments for not only analyzing the outcome of interventions but also for monitoring and evaluating conflict sensitivity of participating organizations and of the intervention. They also discuss how the OECD DAC evaluation criteria for peacebuilding can be adapted to address conflict sensitivity.

The third group of papers deals with macro-level topics. Picciotto proposes the introduction of a multi-dimensional Conflict Prevention Index in order to promote policy coherence for peace and security of OECD-countries. Kennedy-Chouane compares the challenges of peacebuilding evaluations and climate change evaluations, identifying crucial similarities but also differences. From her comparative perspective, she concludes that evaluations in the field of aid and climate change are on average at a stage, where peacebuilding evaluations were some years back. The final paper by Faust and Roxin deals with the potential roles of evaluation of humanitarian aid and broader development cooperation in the current refugee crisis. While acknowledging the value of micro-perspectives, they emphasize the importance of macro and meso level perspectives on evaluation of the European refugee crisis. They also encourage evaluators to take a more proactive position in the current debate of the refugee crisis and warn against a 're-bilateralization' of development assistance.

Reference

OECD (2012). *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility*. Paris: OECD Publisihng.

THE GEOSPATIAL REVOLUTION IN IMPACT EVALUATION

Rachel Trichler, Ariel BenYishay and Bradley Parks

We introduce a new impact evaluation tool for spatially-distributed development interventions and discuss its advantages and limitations in conflict and fragile state settings. When geocoded program data are merged with spatially-referenced survey, census, and satellite data on outcomes like poverty, disease, conflict, governance, and environmental degradation, evaluators can employ quasi-experimental methods of causal identification that control for potential confounds and omitted variables at fine geographic levels, thereby potentially addressing the longstanding critiques of evaluations that do not employ randomization methods. These geospatial impact evaluation (GIE) methods create new opportunities to rigorously measure programmatic impact at a substantially lower time and financial cost. Whereas randomized controlled trials often require expensive primary data collection over customized samples and extensive coordination and collaboration between evaluators and implementers over the entire course of a development program, GIEs can be conducted retrospectively and remotely with more easily accessible program, outcome, and covariate data. These features make GIEs particularly useful to evaluators working in conflict and fragile state settings.

Availability of Geocoded Data

Over the last several years, the international development community has witnessed an explosion in the availability of sub-nationally georeferenced programmatic intervention and outcome data. The World Bank was a first mover in publishing spatial program data, launching an ambitious effort in 2010 to geocode the universe of active IDA and IBRD projects. The World Bank now publishes precise latitude and longitude coordinates of all investment projects, and a growing num-

ber of bilateral and multilateral aid agencies have followed suit.

A similar pattern of diffusion has been observed in the developing world. Malawi's Ministry of Finance was the first owner and operator of a country-level aid information management system (AIMS) to embrace subnational geocoding of development projects (Weaver et al., 2014). There are now more than two dozen finance and planning ministries in Africa, Asia, and Africa with subnationally geocoded AIMS data, including Somalia, Colombia, Iraq, Honduras, Nigeria and Afghanistan.[1]

At the same time, subnational data on intended and unintended outcomes have rapidly expanded in number, scope, and accessibility. Census and household survey data, including the Demographic and Health Survey (DHS) and Afrobarometer, are increasingly georeferenced to the level of enumeration areas (Nunn and Wantchekon, 2011). Satellite data on luminosity, a proxy for economic development, is now available at high-level spatial resolution and fine time scales (Bundervoet et al., 2015). Data on land cover, population density, climate, distance to roads and cities, and other physical attributes are also widely available.

For evaluations that seek to determine whether a given intervention or bundle of interventions affect conflict outcomes, several sources of sub-nationally geocoded data also exist. The Uppsala Conflict Data Program's Georeferenced Events Dataset (UCDP GED) documents organized violence as far back as 1975 and the Armed Conflict Location and Event Database (ACLED) catalogs political violence and protest in developing countries since 1997. The Integrated Crisis Early Warning System (ICEWS)

and the Social Conflict Analysis Database (SCAD) are also rich sources of geocoded, time-varying conflict data.

Geospatial Impact Evaluation

In order to illustrate how these spatial program, outcome, and covariate data open up new avenues for evaluation in conflict and fragile state settings, consider an evaluation that is currently underway on the intended economic effects and unintended conflict effects of natural resource concessions in Liberia. This GIE, which is being led by researchers from AidData, the University of Texas at Dallas, and the Concessions Working Group in Monrovia, aims to identify the effects of different types of foreign direct investment (FDI). For the intervention data, the evaluators drew upon existing but fragmented data sources to assemble a georeferenced dataset of all available concession contracts the Liberian government has awarded to investors since the end of the civil war. The evaluators are merging these data with spatiotemporal outcome and covariate data on poverty, electrification, deforestation, social conflict, and violence from satellites, ICEWS, and multiple DHS waves. Recognizing that FDI projects are not randomly located across Liberia, the evaluators will use propensity score matching methods to compare conflict and economic development outcomes in areas with and without certain types of investment projects. By controlling for potential confounds and omitted variables at fine geographical levels that otherwise might explain changes in the outcomes of interest, this evaluation design will strengthen the counterfactual and improve estimates of programmatic impact. Another important methodological component of this work will be to ensure that the spatial characteristics of the data

---

[1] AidData has subnationally geo-referenced nearly 140,000 development project locations worth approximately $870 billion in aid. These data can be accessed at http://aiddata.org/subnational-geospatial-research-datasets.

are appropriately accounted for, such that spatial autocorrelation and spillover effects do not bias the results.

GIE methods can be particularly useful to evaluators working in active conflict and fragile state settings, where outcome data are often scarce and data collection capabilities are limited. When evaluators rely on satellite-based outcome data, the challenges of geographical access and repeated data collection can be easily overcome. Survey, census, and administrative data can also be collected by national statistical offices and line ministries and used in GIEs for multiple donors and programs, thus replacing donor-specific and program-specific data collection efforts. By way of illustration, consider a recent evaluation of UN Peacebuilding Fund (PBF) activities aimed at improving social cohesion within communities in Burundi that hosted a wave of returning ex-combatants and internally displaced persons (Campbell et al., 2014). In this case, a statistical algorithm was used to match similar locations with and without PBF activities using data collected at various spatial scales. Many of the spatial covariate data – at the colline, commune, and provincial levels – had already been collected by Burundi's national statistical office (ISTEEBU) and various line ministries.

Strengths and Limitations

GIE is a useful but underutilized tool for evaluating development interventions that are geographically distributed. It can produce rigorous estimates of programmatic impact in a timely and cost-effective manner. Such evaluations can also be undertaken after project completion and with existing data. For programs with well-documented intervention locations, a desk-based evaluation can often be completed in six months. The potential for replication in various settings also offers significant external validity benefits, and with time-series outcome data evaluators can sometimes look at treatment effects 5 or 10 years after project completion (e.g. BenYishay et al., 2016).

At the same time, evaluators who wish to use GIE methods face a unique set of challenges in settings of active conflict or fragility. The existence of geocoded program data may be limited in areas with highly sensitive programs or high risks of data misplacement. Some organizations opt not to collect detailed locational information about their programs in conflict settings, while others place strict limitations on the use of such data. Additionally, when GIE and qualitative methods are used in conjunction, they can shed light on both causal impacts *and* mechanisms (Campbell et al., 2014); however, evaluators are often not able to visit treatment and control sites and communities in active conflict situations. GIE thus does not offer a panacea for evaluation challenges, but in the right circumstances, it can complement theory-driven and contextually-relevant analysis with rigorous statistical evidence.

References

BenYishay, A., Runfola, D., Trichler, R., and Heuser, S. (2016). Indigenous Land Rights and Deforestation: Evidence from the Brazilian Amazon. *AidData Working Paper #22*. Williamsburg, VA: AidData.

Bundervoet, T., Maiyo, L, Sanghi, A. (2015). Bright Lights, Big Cities: Measuring National and Subnational Economic Growth in Africa from Outer Space, with an Application to Kenya and Rwanda. *World Bank Policy Research Working Paper no. 7461*, World Bank, Washington, DC.

Campbell, S., Dexter, T., Findley, M., Hoftmann, S., Marineau, J., and Walker, D. (2014). *Independent External Evaluation, UN Peacebuilding Fund Project Portfolio in Burundi*. The Graduate Institute of Geneva Centre on Conflict, Development, and Peacebuilding.

Nunn, N., and Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review* 101: 3221–3252.

Weaver, C., Davenport, S., Baker, J., Findley, M., Peratsakis, C., and Powell, J. (2014). *Malawi's Open Aid Map*. Washington, DC: The World Bank.

RIGOROUS IMPACT EVALUATION OF DDR PROGRAMMES

Lilli Banholzer

Disarmament, Demobilisation and Reintegration (DDR) programmes seek to manage the return of combatants through the collection of weapons, the provision of a civilian status and re-integration into civilian life, all of which are meant to prevent re-recruitment for combat. Little is known about the effectiveness of DDR programmes because systematic evaluations are lacking and so is the answer to the question of under which conditions DDR is successful (Schulhofer-Wohl & Sambanis, 2010).

The "success" of DDR programmes can be examined from different angles. Academic studies tend to define "success" as a country remaining peaceful in the aftermath of a programme's implementation. Practitioners prefer to focus on technical aspects and programme design. What is often missing is the focus on individual combatants. While theory acknowledges the need for micro-level approaches, they are rarely implemented. Fieldwork in conflict-affected areas is time-consuming and challenging, but it can be done. The following sections outline possibilities and challenges for systematically evaluating DDR programmes through field research. Two points will be addressed: (1) defining, operationalising and measuring DDR success; and (2) choosing an appropriate and feasible research design.

Defining, operationalising and measuring DDR success

DDR programmes commonly unfold in three phases, either sequentially or simultaneously. Disarmament refers to the gathering together of combatants, the collection and destruction of weapons, and the assessment of eligibility for further assistance. Demobilisation refers to the formal release of combatants from their military organisations. Reintegration refers to the process of building a new life as a civilian (United Nations, 2005).

Disarmament and demobilisation are considered a success if all or at least a large number of targeted combatants hand in their arms and acquire civilian status. Impact can be measured by comparing pre-estimations of to-be-disarmed combatants with the number of actually disarmed combatants. There are however two primary challenges: i) obtaining reliable estimates of the number of active combatants who need to be disarmed (quantity); ii) determining whether the core members of an armed organisation have been demobilised (quality). Information problems are a key challenge in civil war research. While some rebel groups are built on hierarchical structures with clear memberships, others are fragmented and membership is far less evident, which makes it difficult to determine the real number of combatants to be demobilised. Quality refers to the question of whether core fighters i.e. highly capable or highly ranked personnel disarm.

The success of reintegration is more difficult to assess because it is a lengthy and complex process. There is neither a single measure for reintegration, nor is there a common understanding of which proxies to use, but two dimensions are commonly found: economic reintegration and social reintegration. Some studies cover psychological factors, too.

Economic reintegration is arguably easier to judge. The question is whether former combatants find civilian forms of employment. If large parts of the population are unemployed, the combatants might however be "reintegrated into poverty" (Knight and Özerdem, 2004: 516). Rather than simply asking whether an ex-combatant has a civilian job, it is more meaningful to carry out longitudinal comparisons to assess whether combatants are more likely to find and retain a job if they participated in a DDR programme.

Social reintegration is even harder to capture. Researchers often create their own "mix" of measures. One fairly widespread concept is "acceptance" – do individuals feel accepted by their communities? Reintegration is also often measured by determining whether individuals continue to have ties to their former armed groups or whether they show willingness to become politically engaged. As reintegration is a multidimensional concept, it makes sense to measure different dimensions. However, comparability between studies would benefit from a common understanding among scholars of the core concepts in use.

Choosing an appropriate research design

Unstable security situations and a lack of infrastructure render survey research difficult. The available sample is often a compromise between satisfying standards for random samples and meeting practicability limits, especially if the evaluation takes place ex-post. To ensure that the evaluation is an integrated part of the DDR programme from the beginning on is therefore highly recommended.

Two research designs are recommended to determine the impact of DDR programmes on individual combatants: a) A cross-sectional comparison between participating and non-participating ex-combatants; b) A longitudinal/panel comparison of participating combatants. Preferably, both designs should be combined as they provide complementary information about DDR effectiveness.

Cross-sectional comparison

One method to determine DDR success is to compare DDR programme participants with non-participants. In the disarmament and demobilisation phase, this approach provides information on why some combatants choose to disarm while others do not. In the reintegration phase, comparing these two groups shows whether participants are better off in terms of their social and economic wellbeing.

DDR participants are normally registered and concentrated in cantonment sites or reception centres run by international organisations or NGOs. Selecting a random population is therefore often fairly feasible. Obtaining a representative and random

sample of the population of non-participating combatants is more difficult and depends on case-specific factors. Combatants can have different motivations not to participate, and this heterogeneity of motivations can lead to biases in the sample, including:

• Combatants refusing to disarm: Researchers primarily talk to those who have left their units voluntarily or who were discharged. Unless you are a very reckless researcher, it might not possible to reach those combatants who remain with their armed units.

• Self-selection into programmes: Some combatants do not participate because they do not face any challenges in returning to a civilian life. Reversely, combatants deciding to participate in DDR programmes might be those facing great difficulties in adjusting back into civilian life. Non-participants might therefore actually do better than participants in terms of reintegration.

• Spillover effects: In case of a high level of programme participation, even non-participants can benefit indirectly. A comparison between participants and non-participants might therefore show little difference.

Longitudinal /Panel design

A second approach to evaluate the effectiveness of DDR programmes is to interview former combatants before they participate in the programme (to establish a baseline), while they are engaged in the programme (at several different points in time) and after they have completed it (again at several different points in time). This allows for a longitudinal comparison and greater insights into the effectiveness of programme components.

There are a few practical challenges though: Funding needs to be secured over a long period and the commitment of the researcher needs to be strong. Tracking former combatants is probably the most challenging task. But it is not impossible. Some researchers send evaluation surveys via mobile phones; others work closely with local organisations and community leaders. Another approach is to organise former DDR participants into cooperatives or groups and have the group maintain contact with the former combatant.

Conclusion

DDR programmes would benefit from more systematic impact evaluations. These can be

conducted though a cross-sectional and/or a longitudinal approach. The collected data will likely include some biases, which need to be recognised to avoid causal fallacies. To ensure a research design that best meets scientific standards, the evaluation should be an integral part of the DDR process from the beginning onwards.

References

Knight, M. and Özerdem, A. (2004). Guns, camps and cash: disarmament, demobilization and reinsertion of former combatants in transitions from war to peace, *Journal of Peace Research* 41 (4): 499-516.

Schulhofer-Wohl, J. and Sambanis, N. (2010). Disarmament, Demobilization, and Reintegration Programs An Assessment. Research Report. Folke Bernadotte Academy Publications.

United Nations Disarmament, Demobilization and Reintegration Resource Centre. (2005). What is DDR? Secretary-General, note to the General Assembly, A/C.5/59/31, May 2005.

WHAT DO WE KNOW ABOUT AID AND VIOLENCE

Christoph Zürcher

Can development aid reduce violence in countries in or after war? Research and evaluation on this crucially important topic has grown in importance in the wake of the costly wars in Iraq and Afghanistan. But what do we really know?

One way to answer this question is to conduct a systematic review. A systematic review is designed to identify all available evidence on a given topic, in this case, the impact of aid on violence. A systematic review follows a transparent search protocol and has clearly defined inclusion criteria (Waddington et al., 2012). For example, a recent systematic re-

view conducted by this author included only studies with a clear and transparent identification strategy that allows assessing the counterfactual (what would have happened without the intervention?). Furthermore, only studies with a focus on the sub-national level were included because higher levels of aggregation make it too difficult to understand causal processes. All types of aid with the exclusion of humanitarian aid were included. Based on these inclusion criteria, approximately 20 studies were identified. While this may appear to some as a surprisingly small number, it should be noted that studies with a robust identification strategy

depend on very granular, sub-national data on aid and on the security situation. Such data is not yet widely available.

It is only very recently that donors began to provide micro-level data on their aid disbursements. Given these constraints, it is fair to say that we have now a decent sample of high-quality studies on the impact of development aid on violence They include various types of aid, such as community driven development programs (CDD) and commander emergency response program (CERP) widely employed by the US military in Afghanistan and Iraq, conditional cash transfer programs

(CTT), employment programs and large-scale infrastructure programs in Afghanistan, Iraq, Philippines, Colombia, India and others. Clearly, gaps remain with regard to regional coverage and other types of aid, but it is not too early to describe the emerging evidence.

The first and most important finding is that aid can have a violence dampening or a violence increasing effect. While a number of studies find that aid dampens violence, an approximately equal number of studies find that aid actually increases violence. Secondly, neither a violence-dampening nor a violence-increasing effect is clearly associated with a particular type of aid: all of the above mentioned aid types can lead to either outcome. This suggests that the impacts of aid are in important ways modified by the environment in which an aid program is implemented. In other words, scope conditions – that is, the characteristics of the environment in which aid takes place - are of paramount importance. Where violence is rife it is not the type of aid which makes a difference. Rather, the impact depends primarily on the environment in which aid projects are implemented.

Aid may have a negative impact when it is given to regions where insecurity is already high, e.g. when armed rebels are relatively strong and predominately ideologically motivated (as opposed to rebel-entrepreneurs, who are primarily interested in making profits out of war). They have good reasons to be wary of aid projects since such interventions may foster better relations between local communities, the government and its international backers. In response they may be inclined to disrupt these relations, if needed by violent means (see for example Weintraub, 2016). Hence in the aggregate, it is not surprising that aid projects that are implemented in regions with high insecurity and strong insurgent presence tend to increase violence rather than dampen it.

This is an important finding, given that massive amounts of aid have been spent in highly insecure zones in Iraq and Afghanistan, mostly by the US military. This contradicts the seminal study by Berman et al. (2011),

which tests the effects of CERP (Commanders Emergency Response Program) in Iraq and finds that "improved service provision reduces insurgent violence" (Berman, Shapiro and Felter, 2011: 767).

But we also find that aid, when implemented in relatively secure regions, can dampen violence. How does this violence-dampening mechanism work? Recent research points to three important mechanisms: aid can have a violence-dampening effect by capturing hearts and minds. When aid is perceived as valuable enough the local population may be less likely to side with (or join) the insurgence. Beath, Christia and Enikolopov (2014) provide some evidence for this. Aid can also help by addressing group-level grievances. When one particular group, typically an ethnic minority, holds economic and social grievances, well-targeted aid may enable redistribution, lessen inequalities, create solidarity links between population groups and remedy grievances (Arcand, Bah and Labonne, 2010).

Finally, a violence-suppressing effect of aid is often explained by an opportunity cost model. Economic opportunities, it is argued, can provide employment for young men, which make recruiting them as fighters more expensive. In the long run, violence decreases because rebels can no longer recruit fighters. Dasgupta, Gawande and Kapur (2016) provide evidence citing the case of a large employment guarantee scheme in rural India.

It is important to bear in mind, however, that these violence –dampening mechanisms only work when there is already a reasonable level of security so that insurgents lack the capacity to penalize those local communities who work with government agencies and donors, and when the local political economy makes it unlikely that violent actors misappropriate aid resources.

One clear and important implication for policy makers emerges from this review: donors should allocate aid mostly to environments where conditions are stable enough for aid to generate a benign effect. In war torn countries such targeting is not easy since the

operating environment is typically fluid and localized. One district may provide benign conditions, and the neighboring district may provide adverse conditions. Donors often lack the specific local knowledge and the analytical tools needed to make informed decisions. Here is where future evaluations of the impacts of aid on violence can and should help: by developing, first, a better understanding of contextual conditions which modify the effects of aid, and, second, by developing assessment tools that can help to identify such conditions "on the ground"

References

Arcand, J.-L., Bah, A. and Labonne, J. (2010). Conflict, Ideology and Foreign Aid. *CERDI, Etudes et Documents, E 2010.21.*

Beath, A., Christia, F. and Enikolopov, R. (2012). Winning Hearts and Minds through Development: Evidence from a Field Experiment in Afghanistan. *Policy Research Working Paper 6129.* The World Bank.

Berman, E., Shapiro, J. N. and Felter, J. H. (2011). Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq. *Journal of Political Economy* 119, no. 4: 766–819.

Dasgupta, A., Gawande, K. and Kapur, D. (2016). (When) Do Anti-Poverty Programs Reduce Violence? India's Rural Employment Guarantee and Maoist Conflict (April 22, 2016). Available at SSRN: http://ssrn.com/abstract=2495803 or http://dx.doi.org/10.2139/ssrn.2495803 (forthcoming in International Organization).

Waddington, H. et al. 2012. "How to Do a Good Systematic Review of Effects in International Development: A Tool Kit." *Journal of Development Effectiveness* 4 (3): 359–87.

Weintraub, M. (2016). Do all good things go together? Development Assistance and Violence in Insurgency. (Forthcoming in *Journal of Politics*).

METHODOLOGIES IN PEACEBUILDING EVALUATIONS: CHALLENGES AND TESTING NEW APPROACHES

Thania Paffenholz

One of the biggest methodological challenges in peacebuilding evaluations is the evaluation of effectiveness. Peacebuilding evaluation guidelines put a lot of emphasis on the importance of assessing theories of change (Church and Rogers, 2006; Paffenholz and Reychler, 2007; OECD/DAC, 2012). Yet, it is rare to come across projects, programmes and policies with well-developed baselines and theories of change embedded in a results-based framework. Hence, evaluating peacebuilding effectiveness remains a challenge.

In this short article, I discuss an innovative methodological approach to the evaluation of peacebuilding effectiveness. This approach entails drawing upon evidence-based comparative research on what worked and what did not work in similar peacebuilding interventions in order to analyse outcome plausibility. It has been tested in a global evaluation of civil society peacebuilding projects in eight countries, and proved extremely useful, especially in cases where there are no clear baselines or where theories of change are confused.

Step 1: Reconstructing theories of change

In many peacebuilding evaluations, baselines and theories of change that provide the rationale for *how* and *why* peacebuilding projects are assumed to have an impact, are implicit rather than explicit. It is also quite common that the theories listed in project documents no longer reflect the intervention logic due to changes in project design or context. In such cases, the theory of change has to be reconstructed as part of an evaluation.

Data constraints pose a tricky problem in this regard. Bamberger, Rugh and Mabry (2004) suggest reconstructing baselines and theories of change with the help of secondary data from programmes/projects, national statistics like national household surveys and by interviewing the main evaluation stakeholders, including beneficiaries. Applying methods to compensate for missing data

involves a number of constraints, including the issue of stakeholder bias when it comes to recalling the past (ibid.). A further issue in complex conflict contexts is that national authorities are often dysfunctional, and reliable national statistics or surveys, and public perception studies are rarely available. As a consequence, more testing is required to acquire reliable approaches and methods.

In cases where an explicit theory of change is elaborated in project documents, the evaluator nevertheless needs to assess whether the theory remains valid.

Step 2: Assessing theories of change

Evaluating the logical plausibility of a theory of change involves relating a project's activities and outcomes to its desired impacts or goals in order to establish whether such impacts might reasonably be achieved (Rossi, Lipsey and Freeman, 2004: 158–159). The evaluator can also assess whether a theory of change is relevant in a given context. Firstly, this can be done by analysing the causes and dynamics of conflicts over time, and secondly, by evaluating if the theory of change addresses these factors in a logical sequence in order to facilitate peacebuilding.

Step 3: Assessing outcome plausibility

In the absence of detailed baseline data, an innovative way of assessing peacebuilding outcomes is analysing outcome plausibility by comparing the theory of change and the main activities conducted with existing comparative data on what has worked and what has not worked in similar peacebuilding interventions. However, such assessments are only possible if sound evidence from research or other evaluations exists. In my experience, even when such research results exist, they require adaptation to the purpose of evaluation, i.e., they need to be "translated". A practical example of this approach is presented below.

I used such a "translation" approach as part of a global evaluation of support to civil society peacebuilding initiatives in eight countries (Paffenholz et al., 2011: 3–10), drawing on the results of a multi-year international research project on the role of civil society in peacebuilding (Paffenholz, 2010; IPTI, 2016). For the purposes of this research project, I co-developed a framework for the analysis of civil society in peacebuilding (Paffenholz and Spurk, 2010) which elaborates seven functions civil society can fulfil: protection, monitoring, advocacy, socialisation, social cohesion, facilitation, service delivery (see Figure 1). The effectiveness of the seven functions has then been assessed in 13 country case studies on the level of cumulative impact by function.

Figure 1: Seven civil society peacebuilding functions (Paffenholz and Spurk, 2006).

Seven civil society peacebuilding functions

1. **Protection** of citizens from violence from all parties;
2. **Monitoring** of human rights violations, the implementation of peace agreements, etc.;
3. **Advocacy** for peace and human rights;
4. **Socialisation** for democratic and peace values as well as for in-group identity of marginalised groups;
5. **Inter-group social cohesion**, bringing people together from adversary groups;
6. **Facilitation** on the local and national level between all types of actors;
7. **Service delivery** for creating entry points for peacebuilding, i.e. for the six above functions.

In order to make use of the research results for the evaluation in question, they were translated into effectiveness conditions against which the projects could be assessed. For example, with regard to function 5, "inter-group social

cohesion", the research found that the relevance of activities which bring people from adversarial groups together depends largely on the context in which these activities take place (Paffenholz, 2010: 405–424).

More specifically, the research results showed that the following reasons limited the effectiveness of most dialogue projects (ibid: 427):

- Radicalisation within society hinders this type of peace work;
- The main focus of most initiatives is on the main conflict lines only;
- Most initiatives are of a scattered, short-term and fragmented nature;
- Most participants are English-speaking, elite-based representatives who are often already "converted" to the idea of positive images of the other group;
- People-to-people programmes do not reach society at large as they only focus on the individual level;
- The apolitical nature of most initiatives frame a deeply political problem as a relationship problem, something that can often be misleading, and result in limited acceptance and ownership within society;
- Many initiatives aim at changing attitudes, yet even over the long-term, this seems ineffective. Existing evidence from Bosnia, Cyprus and Israel/Palestine demonstrate that attitude change might not be necessary for behavioural change. Instead, work-related activities, which brought people from different groups together, proved to be more successful than peace-related work. Here people expressed positive experiences from working with

the other group, often producing concrete outcomes or common work initiatives.

These findings can then be translated into a checklist in order to establish whether the peacebuilding projects under evaluation have built these conditions into their theory of change and subsequent project implementation.

In a nutshell, the outcome plausibility evaluation approach is a viable alternative in the absence of baseline data. Nevertheless, in order to ensure more effective peacebuilding evaluations in the future, a lot more emphasis should be put both on improving the quality of project planning in peacebuilding and on creating a culture of monitoring and evaluation as an integrated part of project implementation.

References

Bamberger, M., Rugh, J. and Mabry, L.S. (2006). *RealWorld Evaluation. Working under Budget, Time, Data and Political Constraints*. London and New Delhi, Sage Thousand Oaks.

Church, C. and Rogers, M. M. (2007). *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*. Search for a Common Ground. Washington D.C.

Inclusive Peace and Transition Initiative (IPTI) (2016). *Civil Society at a Glance*. http://www.inclusivepeace.org/content/civil-societys-inclusion-peace-processes-glance.

OECD/DAC. (2012). Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Paris, OECD. http://www.oecd-ilibrary.org/development/evaluating-donor-engagement-in-situations-of-conflict-and-fragility_9789264106802-en [accessed 22.03.2016].

Paffenholz, T. (2010) (ed.). *Civil Society and Peacebuilding. A Critical Assessment*. Boulder: Lynne Rienner.

Paffenholz, T. and Reychler, L. (2007). *Aid for Peace. A Guide to Planning and Evaluation for Conflict Zones*, Nomos Baden Baden.

Paffenholz, T. and Spurk, C. (2006). "Civil Society, Civic Engagement and Peacebuilding". Social Development Papers, Conflict Prevention and Reconstruction Paper No. 36. Washington, D.C.: World Bank.

Paffenholz, T. and Spurk, C. (2010). A Comprehensive Analytical Framework. In: T. Paffenholz (ed.) *Civil Society and Peacebuilding. A Critical Assessment*. Boulder: Lynne Rienner, 65–78.

Paffenholz, T. et al. (2011). "The German Civil Peace Service: Synthesis Report. Volume I: Main Report." Unpublished evalution report. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.

Rossi, P., Lipsey, M. W. and Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. London and New Dehli, Thousand Oaks.

■

MONITORING & EVALUATION OF CONFLICT SENSITIVITY: METHODOLOGICAL CONSIDERATIONS

Diana Chigas and Nicole Goddard

Conflict sensitivity (CS) is the ability of an organisation to:

1. Understand the context in which it is working, especially the dynamics of relationships between and among groups in that context.
2. Understand how the details of its interventions interact with that context. This includes not only the outcomes of the interventions, but also:
   a. *Details of its programs* (beneficiaries/participants selection, sites and timings of programs, etc.)
   b. *Details of its operations* (hiring, procurement, security, etc.)
   c. *Specifics of its policies* (criteria-setting for both programs and operations).
3. Act upon this understanding to minimize the negative impacts of its interventions on the context and maximize positive impacts.[1]

A conflict sensitive approach requires the application of conceptual tools and frameworks that help to ensure its appropriateness (steps 1 and 2 above) and adaptability (step 3 above), thereby contributing to its overall effectiveness. The application of CS tools can help to ensure the sustainability of development or peacebuilding outcomes and the successful implementation of humanitarian responses.

Why Monitor and Evaluate Conflict Sensitivity?

Unintended negative impacts on conflict dynamics have the potential to undermine any intervention. Failure to adapt a programme adequately in response to changes in the context can prevent achievement of objectives, weaken trust with recipients of aid and other stakeholders, and put the organization, its staff, its partners, and its reputation at risk. Conflict sensitivity is thus a relevant consideration for all programmes in conflict contexts – much like gender, environment or disability – mitigate conflict risks and, where possible, build on positive dynamics in the context. M&E of conflict sensitivity is key to identifying and addressing possible conflict-escalating side effects of all interventions early on and revising programming to avoid contributing to violence and tensions.

What does M&E of conflict sensitivity entail?

Monitoring CS tracks sources of tension (dividers) and connection in the conflict context, how all aspects (operational and programmatic) of the intervention affect and are affected by them, and how the organisation is adapting to minimize negative and maximize positive effects on them. *Evaluating* an intervention's conflict sensitivity involves reviewing the effects of the conflict on the programme, the effects of both the outcomes (peacebuilding, development, humanitarian, etc.) *and* the organisation's operational decisions on the conflict, and the functionality of the *processes* in place to ensure conflict sensitive action.

Evaluation of conflict sensitivity is different from evaluation of peacebuilding. Evaluation of peacebuilding assesses the effects of program *outcomes* on conflict drivers. Evaluation of conflict sensitivity examines whether the intervention's outcomes were achieved in a conflict sensitive way, i.e. whether and how the conflict interfered with outcomes and/or the outcomes were achieved without significant (and avoidable, though unintentional) side effects that exacerbate conflict or undermine connectors. This includes:

a) Analyzing the *conflict* and collecting data on how conflict dynamics have evolved. For projects and programs, it will usually be sufficient to analyse sources of tension, and sources of cohesion in the context. At strategic level, a broader analysis of conflict drivers and dynamics may be necessary, but may be done more rapidly than would be appropriate for M&E of a peacebuilding initiative.
b) Assessing whether *processes* for conflict sensitivity are in place and functioning well. Has a conflict analysis been conducted and updated? What processes are in place to gather information and reflect on how the intervention could escalate or mitigate tensions? How has the intervention been revised in light of the analysis?
c) Assessing how the intervention has adapted to changes in the conflict context. This includes identifying conflict risks to the intervention, and to beneficiaries, staff and partners; assessing how the conflict affects the appropriateness or feasibility of the intervention and the measures taken to address these challenges.
d) Assessing the interaction between the intervention and the conflict context. How has the intervention affected the conflict context, and what has been done to mitigate any negative effects the intervention may be having, with what effect? At the strategic level, decisions such as the overall choice of instruments, the sectoral and geographical focus and choice of partners might be examined. Program- or project-level monitoring and evaluation of conflict sensitivity focuses more on how

1 AFPO, CECORE, CHA, FEWER, International Alert, & Saferworld. (2004). *Conflict-sensitive approaches to development, humanitarian assistance and peacebuilding: A Resource Pack.* London: Saferworld et al.

the design and implementation of specific activities could be contributing inadvertently to tensions.

An example: Evaluating conflict sensitivity

An evaluation of a food aid program in a conflict-affected province focused on whether the program reduced malnutrition. Evaluating the programme's conflict sensitivity would also examine how the programme mitigated conflict risks and how it affected the behavior of the conflict parties. For example, did it free up resources for conflict parties to pursue violence or conflict or provide incentives to armed actors to behave in ways that increased tensions? In this case, the program had positive results on malnutrition but did so in a conflict-*insensitive* way by inducing farmers to abandon wheat in favor of poppies, also inadvertently enriching and expanding the power and resources of a key armed actor (AFPO, CECORE, CHA, FEWER, International Alert, & Saferworld. (2004).

Methodological Approaches for M&E of conflict sensitivity

The OECD/DAC Criteria for Evaluating Development Assistance, adapted for peacebuilding interventions the DAC's guidance, Evaluating Peacebuilding Activities in Settings of Conflict and Fragility, can be used to evaluate conflict sensitivity, with adaptations to some of the questions (OECD/DAC, 2012).[2] (see table below).

Indicators and results frameworks can help in monitoring conflict sensitivity. *Conflict* indicators provide information on how the conflict issues relevant to the intervention are evolving. *Interaction* indicators track how the intervention is affected by the conflict, and how it affects the conflict trends, considering facts and perceptions of who actually benefits from the intervention (e.g., employment, contracts, aid recipients, even theft and corruption). In evaluation, theory-based approaches, such as contribution analysis, theory-based evaluation, or process tracing, with their focus on context and on mechanisms, can be helpful in identifying the *contribution* of the intervention to observed positive or negative effects, provided they are conducted in a "eyes wide open" manner, with attention to what is happening on the ground (Bamberger, 2012).

At the same time, conflict sensitivity impacts are often unintended, and sometimes also unanticipatable even with robust conflict sensitivity analysis in programme design and monitoring. Therefore, these methods should be supplemented in both monitoring and evaluation with more open-ended approaches to capture these effects, such as Most Significant Change or Outcome Harvesting, as well as beneficiary feedback and open-ended listening processes, especially with stakeholders who are *not* program participants.

Conclusion

The gaps in evidence about conflict sensitivity remain large. Methods for M&E of conflict sensitivity have not been sufficiently docu-

mented or systematically applied, in part because M&E of conflict-sensitivity happens only episodically and because CS often takes place within the "black box" of implementation.

CDA Collaborative Learning Projects (www. cdacollaborative.org) continues to examine M&E of conflict sensitivity as a field of ongoing learning and advancing practice. We encourage organisations to share their lessons learned from application of M&E processes to conflict sensitivity.

References

AFPO, CECORE, CHA, FEWER, International Alert, & Saferworld. (2004). *Conflict-sensitive approaches to development, humanitarian assistance and peacebuilding: A Resource Pack*. London: Saferworld et al.

Bamberger, M. (2012). *Unanticipated consequences of development interventions: a blind spot for evaluation theory and practice*. Lecture at International Program for Development Evaluation Training (IPDET). Retrieved from https://www.youtube.com/watch?v=R31Udi_5x6o.

Goldwyn, R. and Chigas, D. (2013). *Monitoring and Evaluating Conflict Sensitivity: Methodological challenges and practical solutions*. CCVRI Practice Product. London: DFID.

OECD/DAC. (2012). *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Learning for Results*. Paris: OECD.

∎

OECD DAC evaluation criteria application to peacebuilding		OECD DAC evaluation criteria application to conflict sensitivity
Do objectives and activities address key drivers of conflict *and* are responsive to conflict?	**Relevance**	Is the intervention responsive and adjusting to conflict context?
Have intended objectives met with respect to peacebuilding and immediate or secondary outcomes' relation to peacebuilding and conflict dynamics?	**Effectiveness**	Did the intervention effectively manage conflict-related risks?
What are the effects, intended or unintended, medium or long-term, on the wider conflict dynamics, i.e., key drivers of conflict and peace?	**Impact**	What are the effects, intended or unintended, on tensions, vulnerabilities, grievances, connectors and capacities for peace?

2 The adaptations are taken from Goldwyn and Chigas (2013).

TOWARDS A CONFLICT PREVENTION INDEX (CPI)

Robert Picciotto

This brief article outlines an index that would measure and compare the commitment to conflict prevention by OECD countries and major regional powers. Used in conjunction with other public information activities the proposed Conflict Prevention Index (CPI) would encourage the adoption of effective conflict prevention policies.

Rationale

Policy indexes and league tables have mushroomed in recent years. Well-constructed, properly researched and effectively disseminated, they have a variety of uses: they help to educate citizens; promote internal reforms; improve public policy; guide decision making; etc.

Indexes are also an integral part of the evaluator's tool kit. While they do not substitute for policy, program and policy evaluation they bring together in a concise way relevant knowledge about topics of evaluative interest. They embody plausible theories of change. They facilitate identification of major policy drivers. By allowing ranking of individual countries or organizations in terms of their contribution to a specified policy goal they help in the targeting of further evaluative inquiry.

Indexes are often contested but the resulting public controversies help in stakeholders' involvement in policy debate. The proposed index is no exception: readers are encouraged to challenge its rationale with a view to improve its design.

Why a CPI?

The need for a CPI is almost self-evident. All countries, rich and poor, have a stake in the prevention of armed conflict. Most contemporary wars take place within low and middle income countries but they do not respect national borders and they threaten international stability.

Country indexes that measure the proneness to conflict of fragile states are in ample supply. But no index currently tracks the critical role played by OECD countries in the prevention of armed conflict. Yet in an interconnected global security system their contribution to conflict prevention is critical to the prevention of war. More often than not it is the combination of domestic and international actions (or inaction) that fuels conflict and spreads it across borders.

Thus just as OECD countries' commitment to development has been measured and disseminated in an index published by the Centre for Global Development in Washington DC since 2001[1] the proposed CPI would drill down within the security dimension domain to track the adoption of conflict sensitive policies by the rich and powerful nations of the world.

CPI design

Contemporary wars are shaped not only by internal forces but also by the policies pursued by powerful countries and the international institutions that they influence. As detailed below and in line with available policy research findings the CPI would reflect the following dimensions of conflict prevention policies:

- **Security:** the rules of the game of the global security system exert significant influences on the behavior of local actors. They may or may not create incentives where grievances accumulate and conflicts spill over national borders.
- **Development:** external interference through development cooperation can either amplify and sustain civil strife. When applied in a timely and judicious way it can help moderate, mediate and resolve internal strife.

Security	Development	Governance
1 Direct military participation in conflicts	4 Adequacy of development aid to fragile states	7 Support of UN (and regional) peace building and conflict prevention initiatives
2 Direct arms exports to countries affected by (or prone to) conflict and controls over the diversion/re-transfer/brokering of arms exports	5 Conflict sensitivity of aid and non-aid engagement w/ fragile states	8 Adherence to international agreements that address global security concerns including reduction of carbon emissions
3 Diplomatic efforts specifically targeted towards conflict mediation, resolution and management	6 Private sector and voluntary agencies' engagement with fragile states	9 Support for public private partnerships aimed at conflict sensitivity of trade and investment policies

1 http://www.cgdev.org/cdi-2015.

- **Governance:** Policy coherence across the security and development domain implies the design and delivery of joined up government initiatives within and across governments.

Security

The **first cluster** of indicators would address direct military interventions. The risks, uncertainties and costs associated with illegitimate military involvement have proven to be so high that it is arguably more rational for external actors to invest judiciously in conflict prevention than to intervene with deadly force, in the absence of a Security Council resolution. The proposed evaluative frame of reference for assessing this delicate aspect of developed country performance under the CPI would be the Responsibility to Protect (R2P) endorsed by the United Nations in 2005.

The **second cluster** of indicators would focus on small arms and light weapons. Technology has made small arms deadly, cheap, light, portable, and easy to use – even by a child. Nor do assault rifles, grenades, rocket launchers, landmines, and explosives require logistical support or elaborate maintenance. As a result, they have been omnipresent and responsible for most of the casualties in intrastate conflicts. Thus, a key policy priority for OECD countries community is to reduce the proliferation of small arms and light weapons.

The **third cluster** of indicators would reflect the principle that an ounce of prevention is worth a pound of cure. Diplomacy is the instrument of choice in early prevention. It is also embedded in post conflict interventions by external actors keen on avoiding a recurrence of warfare. It can increase the effectiveness of other forms of actions designed to reduce tensions (preventive deployment, preventive disarmament, preventive humanitarian action, etc.).

Development

The **fourth cluster** of indicators would reflect the well-established policy research finding that development aid directed to poor and vulnerable countries (and redress the aid orphan problem) can generate very high returns since it can help prevent, postpone, or reduce the incidence of conflict. In a nutshell aid and capacity building assistance to conflict-prone states is certainly risky but it is less risky than not providing it.

The **fifth cluster** of indicators would measure conflict sensitivity of the aid. Policy research has established that aid should not be allowed to increase horizontal inequalities, provoke social resentment, tolerate natural resource mismanagement, distort the exchange rate, exacerbate youth unemployment or induce corruption. Coherent interventions and well selected aid instruments aligned with domestic processes should be favored and designed with care to avoid worsening social tensions. Aid should promote broad based growth, reduce inequities, protect human rights and recognize the central role of good governance.

The **sixth cluster** of indicators would recognize the role of the private and voluntary sectors. The CPI would incorporate the influences that foreign direct investment (FDI) may have on conflict prevention. It would also reflect the growing peace building role of voluntary organizations and foundations.

Governance

The **last three clusters** would acknowledge the importance of good international citizenship geared to global stability. It would reward countries that (i) support multilateral institutions concerned with conflict prevention and international law; (ii) comply with international conventions that directly or indirectly support conflict prevention and peace building and (iii) leadership of (and participation in) public private partnership networks that set standards in such security sensitive areas as the Extractive Industry Transparency Initiative) or the observance of voluntary controls over the trade of conflict resources (e.g. the Kimberley process).

Conclusion

In line with experience with other policy indexes, the proposed CPI would provide a means of holding targeted countries to account for their contribution to a specified goal – international peace and stability. It would gauge the quality of government conflict prevention policies as opposed to their stated intentions. By raising public awareness about good and bad international policy practice, it would provide incentives for the adoption of coherent policies likely to reduce violent conflict in developing countries.

References

Collier, P. and Hoeffler, A. (2000), *Greed and Grievance in Civil War*, Washington, DC. World Bank.

Muscat, R. J., (2002), *Investing in Peace: How Development Aid Can Prevent or Promote Conflict*, Armonk, NY, M.E. Sharpe, Inc.

Picciotto, R., Olonisakin, F., Clarke, M., (2007), *Global Development and Human Security*, New Brunswick, NJ. Transaction Publishers.

WHAT CLIMATE CHANGE EVALUATORS NEED TO LEARN FROM PEACEBUILDING EVALUATORS

Megan Grace Kennedy-Chouane

The success of the 21st international climate Conference of the Parties last December, including commitments of more than $200 billion in finance, signals a new era in global efforts to stop catastrophic climate change. For the global community to limit warming to 'well below 2°C' as promised – and have the best chance of limiting the worst impacts on people – national and collaborative efforts will have to be implemented rapidly and effectively.

Evaluators working on climate mitigation[1] will have an important role to play in bolstering such activities. But to do so, they must continue to improve climate change evaluation methods and practices. This article gives some sign posts for this journey, drawing lessons from the field of conflict prevention and peacebuilding evaluation.

Various authors have described the operational, ethical and political challenges facing evaluations in conflict settings (OECD, 2008; Andersen, Bull and Kennedy-Chouane (eds.), 2014; Bush & Duggan, 2015). As Bush & Duggan aptly state, the combination of volatile contexts and thick politics "magnify existing research and evaluation challenges, rendering them more extreme" (Bush & Duggan, 2015).

Climate change shares many of these characteristics, including systems complexity, transboundary dynamics and contested political contexts. Hence evaluations of activities in both areas face similar methodological and process-related challenges. These challenges may explain in part why both fields have been relatively under-evaluated, compared to the scale of investment (Kohlin, 2016).

These shared evaluation obstacles include:
- **Wicked (attribution) problems:** Both fields are characterized by multi-dimensional causal pathways making it difficult for evaluators to discern the specific effects of any one activity. Many of the tools deployed for creating a counter-factual are not readily useable, except for evaluating smaller, sub-components of programmes (such as the deployment of household solar lights or attitudinal changes towards conflict groups caused by soap operas (see e.g. Gaarder & Annan, 2013), which are relatively insignificant in terms of the overall climate and conflict systems.

- **System complexity:** Many individual interventions in conflict and climate are aimed at systems transformation, which means programme outcomes may be constrained (or enhanced) by interrelated forces and positive or negative feedback loops. Change is non-linear and largely unpredictable – and in the case of climate change the problem is itself not fully understood.

- **Sense of urgency resulting in low evaluability:** Evaluation is hampered by programmes that are not designed and implemented based on understanding and evidencing causal pathways and results. The overwhelming urgency of the existential threats posed by both climate change and violent conflict mean that many practitioners find themselves working under time pressure to deliver results 'yesterday' – an ambiance that does not facilitate careful planning. Baselines are missed, data are not collected, theories of change are unclear, and monitoring is done on the fly, with little documentation of learning.

- **Politicization and potential to cause harm:** Evaluations may be used or misused in ways that exacerbate conflict, or in the

Evaluation Criteria	Conflict	Climate
Relevance	Conflict drivers often poorly understood, limited analysis of conflict sensitivity	One metric (GHG emissions) can be used to ascertain relevant sectors/geographies
Effectiveness	Much progress made to articulate theories of change	Over-emphasis on abatement potential, less on theories of change
Efficiency	Cost benefit analysis very hard to use	Cost benefit analysis easier to use
Impact	Difficult both to measure and to attribute	Ultimate impact is measureable in GHG, but difficult to attribute
Sustainability	Conflict management capacities and resilience not well understood	Similar to other fields, with clear policy, market or behaviour changes to sustain results

Table 1. Relative ease of assessing evaluation criteria: Conflict vs. Climate.

---

1 Climate interventions are those policies, programmes and activities aimed at reducing greenhouse gas emissions, or increasing carbon stocks (e.g. energy, land use and forests, short-lived climate pollutants). In this article I do not look at climate adaptation.

case of climate change, by detractors to block political action. Well-documented instances, including the so-called 'climategate' or ExxonMobil's obfuscation of scientific data, attest to the highly politicized nature of climate action – and of evaluating activities aimed at undermining incumbent powers.

- **Timeline to impact problem:** Interventions in both areas aim at broad socioeconomic changes that will take generations to realise. Both fields provide ripe examples of "disconnects" between donor funding and reporting cycles and the timeframe of desired changes. The two spheres differ however, in that while the impacts of unsuccessful climate interventions will not be immediately visible for some time, failed conflict projects can have instant (negative) effects.

There are also differences between climate change and conflict with implications for evaluation. Perhaps most notably, climate faces a much steeper challenge to motivate action. As a result, a disproportionate amount of intellectual effort seems to go into making the case for climate action and understanding the problem, with less emphasis on effective programme delivery at scale. In peacebuilding there is much less need to convince people that violent conflict is bad and ought to be addressed. So while evaluators might encourage peacebuilding practitioners to focus more on defining outcomes (the 'what'), climate evaluators can add value by helping practitioners articulate intervention logics (the 'how').

Table 1 captures how these similarities and differences can affect evaluation methods, by criteria.

Advances in the peacebuilding field over the past ten years have shown that real progress can be made to improve evaluation (and use evaluation to improve results) through collaborative scholarship and experimentation.

Climate evaluation is today where peacebuilding was fifteen years ago: with increasing funding and attention, and a deepening understanding that evaluation methods and practice need to adjust to the unique (and not so unique) characteristics of climate mitigation interventions.

Many evaluators are working to improve climate evaluations.[2] They may benefit from lessons learned in the peacebuilding field, including the importance of:

- Cultivating demand among practitioners and focusing on usability, with fit-for-purpose methods that provide actionable, timely insights for decision making.
- Integrating evaluation into programme design and management, starting with clear theories of change.
- Expanding the evaluation tool box, using theory-based approaches, such as process tracing, outcome mapping, and contribution analysis.
- Exploring the application of complexity theory and systems thinking to evaluation.

As an evaluation community, we must quickly rise to the challenge of doing more

and better climate evaluations. Though they may sound like an unlikely source, peacebuilding evaluators can help by sharing their experiences. Let us each do our part to help decision-makers act effectively to stop global warming.

References

Andersen, O. W., Bull, B. and Kennedy-Chouane, M. (eds.) (2014). *Evaluation Methodologies for Aid in Conflict.* Routledge.

Bush, K. and Duggan, C. (2015). *Evaluation in the Extreme: Research, Impact and Politics in Violently Divided Societies.* Sage Publications.

Gaarder, M. and Annan, J. (2013). Impact evaluation of conflict prevention and peacebuilding interventions. *Policy Research working paper 6496;* Series no. IE 93. World Bank.

Jamieson, D. (2014). *Reason in a Dark Time: Why the Struggle Against Climate Change Failed.*

Kohlin *et al.* (2016). *In Search of Double Dividends from Climate Change Interventions Evidence from Forest Conversation and Household Energy Transitions.* EBA.

OECD. (2008). *Evaluating conflict prevention and peacebuilding: Towards DAC Guidance.* OECD: Paris.

2 The UK's International Climate Fund, and GEF Independent Evaluation Office are two notable examples.

FOREIGN AID, THE EUROPEAN REFUGEE CRISIS AND THE ROLE OF EVALUATION – BEYOND MICRO LEVEL PERSPECTIVES

Jörg Faust and Helge Roxin

Since summer 2015, the European Refugee Crisis has become one of the most prominent political topics on the continent, challenging not only individual European countries and governments but also the normative and political foundations of the European integration process. One relative broad consensus in the current debate within and between European countries is that the refugee crisis cannot be solved without improving the situation of those, who have desperately left their homes because of protracted violent conflict; no matter whether such improvement is within civil war torn countries or their neighbors.

Consequently, the role of humanitarian aid and development cooperation as a way to help those fleeing from civil war has become not only a highly relevant foreign policy issue but also a core and sensitive domestic political issue throughout Europe. Given the protracted Middle East crisis – it is likely to remain in the limelight for a very long time. Politicians as well as the broader public are increasingly linking the effectiveness of humanitarian aid and development cooperation with the domestic issue of uncontrolled migration.

Against this background, the assessment of foreign aid interventions in the context of protracted violent crises has become highly relevant to domestic audiences within Europe. But are humanitarian aid and long term development interventions capable of making a lasting contribution to controlled migration? This question poses challenges for evaluators and evaluation units as providers of sound, policy-relevant evidence.

Currently, these challenges are mostly addressed at the micro-level of evaluation where the difficult and highly volatile contexts are identified as challenging barriers for implementing sound evaluation designs and processes. For example, the "first, do-no-harm principle" has gained traction given that the transparency created by evaluations can influence the dynamics of conflict. While

such well-grounded micro-perspectives are highly relevant, evaluation should also move to the macro and meso level of the policy debate triggered by the European refugee crisis. In this regard, the macro-level relates to the grand debate around the role of defense, diplomacy and development aid in the context of violent conflict, while the meso-level consists in analyzing the coordination and harmonization of different aid organizations and instruments in such contexts.

The Macro Level: The role of evaluation in the grand debate

As providers of evidence for sound policy making evaluators and evaluation units should raise their sights and reconsider their role in the current debate. Decision-makers – often driven by domestic pressures – tend to overestimate the role of humanitarian aid and structural development cooperation in reducing the number of refugees from crisis-driven societies. Development evaluation practitioners often consider themselves as technicians, who limit their role to the provision of concrete evidence geared to gradual improvements. This cautious stance is often inadequate. It should be complemented by a more proactive and enlightened involvement in policy debate and a deliberate focus on the coherence between diplomacy, defense and development.

Humanitarian and more lasting aid will not be capable of providing substantial structural effects as long as there is a protracted violent crisis. Thus, from an external perspective, the primary instruments to end violent conflict are in the fields of diplomacy and security policy (e.g. Weiss et al 2010). Moreover, in countries neighboring war torn societies, absorption capacities for refugees are limited given restricted financial resources and institutional capacities. Thus, even if the situation of refugees in neighboring countries might be temporarily improved by means of humanitarian aid, these are mainly focused on short term effects rather than long term

solutions. They may even conceal the risks of de-stabilization inevitably associated with large-scale and semi-permanent refugee settlements.

Therefore, to combat unrealistic notions regarding the potential of foreign aid in contexts of protracted crisis, evaluation should use its formalized links to decision-makers and its knowledge dissemination activities to help manage public expectations and ensure that humanitarian and development aid interventions are assessed against realistic goals. Such a pro-active role offers the potential of increasing policy transparency for a broad public constituency. At the same time, it implies independence without which evaluation might fail to bring sobering but highly relevant messages about the role of foreign aid in the European refugee crisis into the broader policy debate.

The Meso Level – Organizing actors and instruments

Despite the limitations described above, there is an important role for humanitarian and development aid in ameliorating the effects of protracted crisis for refugees and neighboring countries (Mosel and Weingärtner, 2014). Much has been learned through evaluation about improving interventions at program and project level (e.g. Healy and Tiller, 2013; Mowjee et al., 2015). Yet, it is also important to highlight the importance of coordinating and harmonizing instruments and organizations as critical for reaching policy coherence. If foreign aid in the context of refugee crisis is to make a substantial difference, humanitarian and more structural aid have to be coordinated in a highly context-sensitive way and huge challenges remain in coordinating and harmonizing the vast amount of bilateral, multilateral and civil-society actors (e.g. Carpenter and Benett, 2016; Culbertson et al. 2016).

Unfortunately, Europe has witnessed a "re-bilateralization" of its aid policies in recent years. Faced with a wave of nationalist popu-

lism that is highly critical of development cooperation, politicians have often reacted in a defensive way by attempting to highlight the distinctive contribution of bilateral aid and its visibility, thereby putting the value of coordination and harmonization on the backburner. In this context, evaluation has to resist this constriction by rejecting the use of narrow analytic lenses that do not take account of the crucial need for partnership across individual donor organizations and countries.

Evaluation not only has to rigorously analyze any given evaluation questions that address the contribution of individual interventions but also to put coordination and harmonization as key elements of effective aid provision on the evaluation agenda. While undoubtedly evaluation on its own cannot hope to fully overcome collective action challenges, it must address the reductionism of simplified "bilateral value for money" arguments by pointing to the potential efficiency and effectiveness losses, if international coordination and harmonization is not given the due relevance it deserves "on the ground". In this context, there is also need for reengaging in joint evaluations to enhance evaluation efficiency and to foster joint learning processes. Again, this requires to put evaluation issues on the agenda, which are not necessarily demanded by managers and decision-makers of individual organizations but are nevertheless of strategic relevance for the overall achievement of aid effectiveness.

References

Carpenter, S. and Bennett, C. (201). *Managing crises together: towards coherence and complementarity in recurrent and protracted crises.* Humanitarian Policy Group, Overseas Development Institute, London.

Culbertson, S. *et al.* (2016). *Rethinking Coordination of Services to Refugees in Urban Areas.* Report by the RAND Corporation, Santa Monica.

Healy, S. and Tiller, S. (2013). *A Review of the Humanitarian Response to the Syrian Refugee Crisis in Jordan, 2012 – 13.*

Mosel, I. and Levine, S. (2014). *Remaking the case for linking relief, rehabilitation and development How LRRD can become a practically useful concept for assistance in difficult places.* HPG Commissioned Report.

Mowjee, T., Hanley, T., Toft, E., Fleming, D. (2015). *Evaluation of the Strategy for Danish Humanitarian Action 2010 – 2015.* Danish Ministry of Foreign Affairs.

Weiss, S. et al (2010). *Diplomacy, Development and Defense: A Paradigm for Policy Coherence: A Comparative Analysis of International Strategies.* Bertelsmann Foundation.
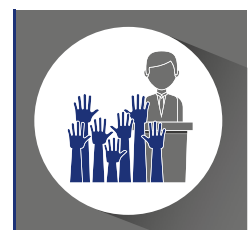
THE AUTHORS

Lilli Banholzer

is Professor for Political science and Conflict research at the University of Mannheim. Her research focuses on peace and conflict studies, political violence, civil war, armed groups, demobilization, disarmament and reintegration of former combatants, civil society and non-state actors, inclusive economic growth and migration. Prior to teaching at the University of Mannheim, she was a post-doctoral researcher at the German Development Institute, a think-tank advising the German federal government on development cooperation matters. She has a strong background in quantitative, survey and multidisciplinary research methods and holds a PhD from the University of Konstanz, Germany.

Ariel BenYishay

is Assistant Professor of Economics at The College of William & Mary and Chief Economist of AidData. He specializes in empirical microeconomics, geospatial impact evaluations and randomized control trials. He heads AidData's Research and Evaluation Unit, and oversees the AidData Research Consortium, some 120 academics at 50 universities worldwide. His current research focuses on the impacts of foreign aid programs on agriculture and deforestation as well as human health and social capital. He leads several large-scale RCTs, including one in the Philippines and several in Malawi. The John D. and Catherine T. MacArthur Foundation, The William and Flora Hewlett Foundation, US-AID, MCC, KfW, The World Bank Independent Evaluation Group, and 3ie, have funded and are currently supporting his research portfolio. He previously served as a Lecturer at the School of Economics with the University of New South Wales and as Associate Director of Economic Analysis & Evaluation at the Millennium Challenge Corporation. Ariel's work has been published in leading journals, including the Journal of Human Resources, the Journal of Comparative Economics, and Economic Development and Cultural Change. Ariel holds a PhD in Economics from the University of Maryland.

Diana Chigas

joined CDA Collaborative Learning Projects in 2003. Her professional areas of expertise are peacebuilding effectiveness, conflict analysis, systems thinking, peacebuilding evaluation, participatory program design and review, theories of change, and use of Reflecting on Peace Practice in evaluative processes. She also co-directs CDA's collaborative learning work. Prior to joining CDA, Diana worked with the Conflict Management Group (CMG). Diana worked as a facilitator, trainer and consultant in negotiation, dialogue and conflict resolution. Her work included development of strategies, training and advice on preventive diplomacy in the OSCE, and facilitation of dialogue in El Salvador, South Africa, Ecuador, Peru, Georgia/South Ossetia, and Cyprus. Diana is also Professor of the Practice of Negotiation and Conflict Resolution at the Fletcher School of Law and Diplomacy.

Megan G. Kennedy-Chouane

is a development professional with expertise in evaluation and social change, Megan Grace Kennedy-Chouane manages Climate Evidence, Measurement and Evaluation at the Children's Investment Fund Foundation. She provides evidence insights to shape investment design and funding decisions and manages a dozen third party evaluations of climate programmes in urbanisation, energy systems transformation, finance and landuse. Prior to coming to CIFF, she spent seven years at the OECD, supporting the DAC Network on Development Evaluation. At the OECD she carried out a major comparative analysis of development evaluation systems in 2010 and led the process of developing guidance on evaluating in situations of conflict and fragility for the OECD. She has worked on evaluation capacity development, formative evaluation, joint evaluation and engagement with emerging donors. She manages production of the OECD's Evaluation Insights series and works to disseminate evaluation findings to inform development policy debates. She holds a Master's degree in Public Administration and International Management, with a focus on public policy evaluation. She holds a post-graduate diploma in Children, Youth and Development Studies and a dual Bachelor's degree in Economics and in Peace

& Global Studies from Earlham College (USA). Ms. Kennedy-Chouane is a U.S. national and currently resides in London.

Jörg Faust

is director of the German Institute for Development Evaluation (DEval), in Bonn. Prior to his position at Deval, he was head of department at the German Development Institute (DIE). He holds a doctoral degree in political science and is also extraordinary professor of political science at the University of Duisburg-Essen. He is a member of the Executive Committee of the European Association of Development Research and Training Institutes (EADI) and of the scientific advisory board for the Bertelsmann Transformation Index (BTI). Beyond evaluation and quantitative as well as qualitative research methods in social sciences, his areas of expertise comprise bilateral and multilateral development cooperation as well as political determinants of economic development and political transformation processes. He has extensive experience in implementing and managing complex research, consultancy and evaluation projects.

Nicole Goddard

has been working at CDA Collaborative Learning Projects since 2007. Her professional focus is the application of conflict sensitivity to technical specialities, integration of conflict sensitivity in organizational policies, procedures, planning and programing, and conflict sensitivity capacity development. She is a Do No Harm trainer and Trainer of Trainers, and has experience training CDA's Reflecting on Peace Practice program planning tools. Nicole most recently co-authored Do No Harm in Land Tenure and Property Rights: Designing and Implementing Conflict Sensitive Land Programs and co-designed the CDA course Practical Tools for Promoting Positive Change in Complex Contexts.

Riitta Oksanen

EES President, is a senior advisor on development evaluation in the Ministry for Foreign Affairs of Fin-

land. She previously held the post of director for development policy in the Ministry and recently chaired the OECD/DAC Evaluation Network task team on evaluation capacity development. She has also acted as advisor on management and effectiveness of development cooperation; served in Finland's permanent EU delegation as counsellor responsible for EU development policy and cooperation, and chaired the Council's working group on development cooperation during the Finnish EU Presidency in 2006. Before joining the Ministry in 1999 she worked as a consultant specialising in development cooperation planning, management and evaluation. She is a University of Helsinki graduate specialized in marketing, business administration and economics with emphasis on the forestry sector.

Thania Paffenholz

is the Director of the Inclusive Peace and Transition Initiative at the Graduate Institute. With over 25 years of experience as both an academic and policy advisor, Dr. Paffenholz is internationally renowned for her work on and in support of peace and political transition processes worldwide, focusing on mediation and peacebuilding, process design, inclusion and participation, as well as the conditions under which these processes produce sustainable outcomes. In recognition of her work, Dr. Paffenholz received the prestigious Wihuri International Prize in 2015. Dr. Paffenholz has roots in Europe, Africa and Asia.

Brad Parks

is the Executive Director of AidData (http://aiddata.org) and Research Faculty at the College of William and Mary's Institute for the Theory and Practice of International Relations. He is responsible for managing the day-to-day operations of the AidData research lab, its portfolio of grants and contracts, and its team of 37 faculty and staff. His research is focused on the cross-national and sub-national distribution and impact of international development finance, and the design and implementation of policy and institutional reforms in low-income and middle-income countries. His work has been published in a variety of disciplinary and inter-disciplinary journals, including Science, World Development, the Journal of Development Studies. Dr. Parks is currently leading an

evaluation of the World Bank's impact on bio-diversity conservation outcomes, an evaluation of the economic and social impacts of foreign direct investment in Liberia, and a large-scale effort to collect experiential survey data from development policymakers and practitioners in 126 developing countries. Brad holds a Ph.D. in International Relations and an M.Sc. in Development Management from the London School of Economics and Political Science.

Robert Picciotto

Robert ('Bob') Picciotto, AcSS, is Visiting Professor, King's College (London) where he teaches evaluation. He holds graduate degrees in public and international affairs (Princeton University, USA) and aeronautical engineering (ENSA, France). He serves on the boards of the UK Evaluation Society and the European Evaluation Society. At the World Bank he was Director General of the Independent Evaluation Group from 1992 to 2002 after holding a variety of operational and senior management positions including Vice President, Corporate Planning and Budgeting and Director, Projects in three of the World Bank's Regions. Since his retirement from the World Bank he has provided senior independent evaluation advice to many international organizations, philanthropic foundations and think tanks.

Helge Roxin

has been working for the German Institute for Development Evaluation (Deval) since it was founded in 2012. Before he worked as a lecturer at the University of Freiburg from 2001 – 2006 (dealing with development politics, African politics and introductions to social sciences) and the Arnold-Bergstraesser-Institute Freiburg from 2006 – 2012 as a senior researcher and evaluator. His main fields of research are crisis and conflict resolution, governance and democratization, gender and micro-finance. His evaluative focus lies on impact evaluations.

Rachel Trichler

is a Senior Research Analyst and Geospatial Impact Evaluation Practice Manager at AidData. Rachel oversees design protocols and analysis for

rigorous evaluations using geospatial program and outcome data as part of the Research and Evaluation Unit's work with bilateral and multilateral organizations and private foundations. Previously, Rachel supported monitoring and evaluation activities for the World Bank's Independent Evaluation Group, Social Impact, and the Millennium Challenge Corporation. Rachel holds a MS in Public Policy from Georgetown University.

Ole Winckler Andersen

is Deputy Permanent Representative at the Danish Delegation to the OECD. He holds a M.Sc. in Economics and a Ph.D. in Public Administration and has worked for the Danish Ministry of Foreign Affairs for the last 20 years, including as Head of Evaluation Department (2007 – 13). Before that he was assistant and associate professor at Roskilde University in Denmark. He has managed a number of evaluations and has served as team leader for missions to Asian, African and Latin American countries. He has also been member of several management committees for international development evaluations. He has published on evaluation in various international journals and recently co-edited Evaluation Methodologies for Aid in Conflict (Routledge 2014).

Christoph Zürcher

is Professor at the Graduate School of Public and International Affairs at the University of Ottawa. His research interests include conflict research, state-building and international development. He has been a consultant for, among others, the German Federal Ministry for Economic Cooperation and Development, the German Federal Ministry of Defense, the Swiss Federal Department of Foreign Affaires, the GIZ and the World Bank. His recent practical work was mainly on evaluating the impact of development aid on violence in Afghanistan.

GUIDANCE
TO CONTRIBUTORS

Brevity and nimbleness are the hallmarks of Connections: published articles are normally 800–1,200 words long. Even shorter contributions (news items; opinion pieces; book reviews and letters to the editor) are accepted with a view to promote debate and connect evaluators within Europe and beyond. While Connections is not a peer reviewed publication only articles that add to knowledge about the theory, methods and practices of evaluation should be submitted.

Contributions that highlight European values and evaluation practices are given priority but Connections also reaches out beyond Europe to the international evaluation community and favours articles reflecting a diversity of perspectives. Within the limits of copyrights agreements articles that summarize in a cogent way the substantive content of published (or to be published) studies, papers, book chapters, etc. are welcome (with suitable attribution).

Individuals or organizations wishing to sponsor special issues about an evaluation theme or topic of contemporary interest should contact the EES Secretariat (secretariat@europeanevaluation. org). Such special issues usually consist of 6–8 articles in addition to a guest editorial. A Presidential letter may be included. The guest editor(s) are responsible for the quality of the material and the timeliness of submissions. The regular editorial team ensures that special issues meet 'Connections' standards and takes care of copy editing.

To facilitate copy editing, authors are encouraged to use end notes rather than footnotes and to use the APA style guide for references. Here are some examples:

For books: Bergmann, I. (1997). Attention deficit disorder. In *The new Encyclopedia Britannica* (Vol. 26, pp. 501–508). Chicago, IL: Encyclopedia Britannica.

For journal articles: Rindermann, H., & Ceci, S. J. (2009). Educational policy and country outcomes in international cognitive competence studies. *Perspectives on Psychological Science*, 4(6), 551–568. doi:10.111/j1745-6924. 2009.01165.x

Website: United States Environmental Protection Agency. (2007, May 4). Climate Change. Retrieved 12 June 2014 from the Environmental Protection Agency website: http://www.epa.gov/climatechange.

In-text reference: (United States Environmental, 2007).

www.ees2016.eu