# Evaluation Capacity Development through Cluster Evaluation

Khalil Bitar
*Co-Founder and Director of the Palestinian Evaluation Association (PEA)*

Rola Hbeichi
*Monitoring and Evaluation Consultant*

Lamia Al-Zou'bi
*Director/ Evaluation and Impact Assessment Unit*
*Ministry of Planning and International Cooperation*
*Co-Founder of EvalJordan*

Craig Russon
*Chair, Evaluation Capacity Development Group (ECDG)*

**Background:** The term "cluster evaluation" was first coined in 1988 by W. K. Kellogg staff in an evaluation of a Foundation-funded initiative. The concept was further developed and practiced by the Kellogg Foundation evaluation consultants and other practitioners in the evaluation community.

**Purpose:** The article explains how cluster evaluation, as originally conceived by the W. K. Kellogg Foundation, was adapted to the realities of the UN system and used to build and enhance capacities of participating evaluators.

**Setting:** Custer evaluation was an approach used to evaluate the programming of a small specialized UN agency in three countries (Lebanon, Jordan and Palestine) of the Arab States.

**Intervention:** Not applicable.

**Research Design:** The authors present a case study of an evaluation of a cluster of programs from the Arab States region (Lebanon, Jordan and Palestine) that was conducted by a small, specialized agency of the UN.

**Data Collection and Analysis:** The evaluation was designed as a series of country program reviews. Information from the reviews was to be aggregated in order to understand their collective contributions to the region.

**Findings:** The case study demonstrates the potential benefits of the model as well as some of the challenges.

*Keywords:* cluster evaluation; international development; case study.

The term "cluster evaluation" was first coined in 1988 by staff of the W. K. Kellogg Foundation (a private philanthropic organization based in the USA) in an evaluation of a large social program that had received funding. In the Foundation's official booklet on cluster evaluation (1995), the definition of the approach is brief and somewhat vague [1]

> Cluster Evaluation is the process of assessing the progress, outcomes, and impact of a programming initiative. It also entails helping WKKF use this information to plan and manage programs, and educate grantees, policymakers, and other relevant communities. (WKKF, 1995)
> Cluster Evaluation is a mechanism for gathering information to use in determining the extent to which WKKF grants are helping people improve their quality of life. (WKKF, 1995)

Over the years, the concept was further developed by the Kellogg Foundation consultants and other practitioners in the evaluation community (e.g. Barley. and Jenness, 1993; Jenness and Barley, 1995). While there is probably not a single, comprehensive definition of cluster evaluation that is agreed upon by all who are familiar with the approach, various authors have pointed out several basic characteristics (Russon & WKKF, 2005):

- It seeks to determine impact through aggregating outcomes from multiple sites or projects, whereas multi-site evaluation seeks to determine outcomes through aggregating indicators from multiple sites.

---

[1] Cluster evaluation, as conducted by foundations of their grantee projects, should not be confused with the evaluation of clusters, an organizational management approach developed by the United Nations to improve humanitarian assistance.

As set out in a Humanitarian Response Review (2014) paper commissioned by the UN, clusters are thematic groups composed of selected sectors or topics, depending on the specific country setting. A cluster is coordinated by a cluster manager and the sectors or topics of a cluster are managed by sector specialists.

The basic premise is that accountability, predictability and reliability can be improved by identifying organisational leaders for areas in which there was an identified gap in humanitarian response. These organizations are then made responsible for specific areas, or clusters.

- It looks across a group of projects to identify common threads and themes that, having cross-project confirmation, take on greater significance.
- It seeks not only to learn what happened with respect to a group of projects, but also why those things happened by discovering factors that contribute to or constrain success within the environment of the projects and the strategies employed.
- It occurs in a collaborative way that allows all players—projects, foundation and cluster evaluator—to contribute to and participate in the process so that what is learned is of value to everyone.
- It helps to strengthen the evaluation efforts of all who are involved, and it supports the work of foundation program officers who have limited time to interact with the projects because of the numerous other demands on their time.
- Cluster evaluators receive most of their data from the project or grantee evaluators, although some primary data are also collected.

In practice, selected projects, usually five or more, are grouped as a cluster because of similarities in strategy or targeted population group. They may have been at different stages in their project cycle but to the extent possible, should have been within the first half of implementation. This allows projects adequate time to adapt and modify their activities as appropriate as they participated in the cluster evaluation and networking with other projects.

The projects were brought together to share evaluation plans for each individual project and to discuss important questions that were common to all projects that formed the basis for the cluster evaluation. The evaluators then prepare plans for seeking answers to important questions through evaluation activities across the cluster. This often involves a variety of methodologies.

Cluster evaluations typically are funded for upwards of four years. During that time, all projects in the cluster convene during a series of networking meetings. These meetings generally included updates on individual projects and on the cluster evaluation, but also revolved around concerns of broad interest to that cluster, e.g., legislation in their subject area, management questions, and the use of computer technology. In some networking meetings, cluster evaluators engage in some form of data collection.

The information collected through cluster evaluation is used to enhance effectiveness, clarify strategies and inform policy. Cluster evaluation is not a substitute for project-level evaluation, nor do cluster evaluators "evaluate" projects. It focuses on progress made toward achieving the broad goals of a programming initiative. Cluster evaluators provide feedback on commonalties in program design, as well as innovative methodologies used by projects during the life of the initiative (Sanders, 1997; 1998).

As one might imagine, the philanthropic context in which cluster evaluation originated is quite different from that of the United Nations. The differences might include some of the following (Worthen and Schmitz, 1997):

*Purpose*— the primary purpose of cluster evaluation in the philanthropic context is learning. Many organizations within the United Nations system have declared their intention to implement results-based management (RBM) as part of the 1997 reform programme (OIOS, 2008). As such, learning is often secondary to accountability.

*Clients*— cluster evaluation in the philanthropic context would have different clients than cluster evaluation in the United Nations context. The difference in clients, who wants information for different purposes, would require differences in design, methodology and reporting.

*Methodology*— in the philanthropic context, a cluster evaluator might find him/herself in the dual role of programming and evaluating. The UN Evaluation Group's Norms and Standards place strong emphasis on independence. In the UN context, a dual role could be seen as a breach of independence.

*Timeframes*— in the philanthropic context, initiatives often have long-term funding. Cluster evaluations of such initiatives can continue over the course of several years. In the UN context, the projects and programs that make up the cluster to be evaluated often have much shorter life cycles.

The authors think that cluster evaluation has many interesting possibilities for use within the UN system. However, the concept requires adaptation to the UN context in order to realize the possibilities.

## Case Study

The case study found below explains how cluster evaluation, as originally conceived by the W.K. Kellogg Foundation, was adapted to the realities of the UN system. In the philanthropic sector, cluster evaluation is a programmatic intervention. That is not the way that the approach was used in the context of this evaluation. Instead of evaluation as intervention, it was evaluation as capacity development.

## Context of the Evaluation

Cluster evaluation was a model used to evaluate the programming of a small specialized UN agency in three countries (Lebanon, Jordan and Palestine) of the Arab States. According to the Terms of Reference, the rationale for clustering the countries in this proposed manner was based on the following commonalities:

- All three are middle income countries that dealt with common thematic issues, including the economic downturn;
- They constituted a critical mass of the specialized agency activities (Technical Cooperation in these countries represented about two-thirds of the agency's total portfolio in the region);
- They were all affected by regional crisis: Palestinian and/or Syrian refugees; and
- The specialized agency was strongly engaged with the UN in these countries.

The purpose of the evaluation was to assess how the three programs jointly achieved intended results linked to regional and national development frameworks as well as to the mandate of the specialized agency.

Because each country was at a different level of strategic program development, the evaluation was also to review how the specialized agency adapted its work methods to different contexts, while supporting the development of a technical cooperation portfolio targeting specific outcomes. Lastly, the cluster evaluation intended to identify common lessons learned and best practices.

In order to accomplish the purposes explained above, the evaluation was designed as a series of program reviews. Information from the reviews was to be aggregated in order to understand collective contributions to the region. Local consultants were hired by the specialized agency to conduct each of the program reviews.

The work of the local consultants was coordinated by a Senior Evaluation Officer from the headquarters of the specialized agency. The manner in which the work was coordinated was to create a learning triad of national consultants. The consultants worked together in five areas: document review, evaluation design, data collection/analysis, lessons learned and reporting.

## Document Review

A Dropbox folder was set up on the onset of the evaluation as a repository for all relevant documents provided by the national, regional and headquarter teams. All documents were saved in the respective folders and, in addition to the team leader, all three local evaluators had full access to the documents and folders uploaded. The local evaluators analysed project documents, progress reports and inception reports to determine their contribution to international, regional and national development frameworks.

Having a shared repository for documents ensured that the consultants had a common base of information to conduct the evaluation. For example, it allowed all members of the team to access other countries' documents in order to not only review documents related to their respective countries but also to be able to read and understand the regional issues as well. During the mission, all three independent national evaluators utilized the same Dropbox folders to upload documents they collected, interview notes they conducted and draft reports they wrote.

Hence, each of the three evaluators had the chance to view other national evaluators' folders and files including notes of interviews conducted in each country. Sharing interview notes, documents collected from the field, and early draft reports for each country allowed the national evaluators in the three countries to be aware of progress made in the other countries.

Because Lebanon hosts the Regional Office, while Jordan has no country office and Palestine's office is supervised by the Lebanon office, the sharing of documents, interview notes and draft reports was essential to the Jordan and Palestine evaluators to learn about issues related to the Regional Office without the need to physically interview the Regional Director or the technical specialists. This also allowed the Lebanese evaluator to be well informed about progress made in the two other countries without the need to physically interview program implementers and other stakeholders there.

## Evaluation Design[2]

---

[2] An evaluation design consists of the evaluation criteria and questions, the approach to be used for answering those questions, a plan for collecting the information needed to formulate an answer, a plan for analysing the information that is collected and a description of the deliverables.

At the outset of the evaluation, Terms of Reference (ToR) were developed that included evaluation criteria and questions based on the literature review and preliminary interviews. The initial ToR was designed on the assumption that each country in the evaluation had some form of a strategy that could address the evaluation criteria. However, once in the field, it was observed that for each country, some of these questions had to be modified to better fit the reality on the ground. For example, contrary to the Jordanian case, both Lebanon and Palestine had no country strategy in place.

Hence, in both Lebanon and Palestine, questions of relevance were modified to answer questions such as "What were the social, political and economic contexts of the problems that Country Programs sought to address?" and "How well did the Country Program results promote the agency's Strategic Policy Framework?" instead of finding answers related to these countries' strategies.

In doing so, and unlike what was stated in the original ToR, the evaluators were able to assess how the work of the country programs contributed to a national strategy or promoted disparate program elements in the absence of a strategy. This could not have happened without having a team leader and ToR flexibility that allowed these modifications to be made by the evaluators to better fit the reality on the ground.

Once the consultants finished with the initial document review, a matrix of interviews with relevant stakeholders was devised. For each country, the national consultant was provided a focal point at the specialized agency in order to set up meetings with stakeholders. Each consultant in turn met/coordinated with the focal point to ensure that at least one representative from all the stakeholders was represented in the interview schedule.

Individuals to be interviewed were generally identified by the Evaluation Matrix Sheet. These included: representatives from the specialized agency; the government and social partners; UN partners; other international agencies; and "outsiders."

The focal point and the national evaluator identified potential interviewees from each category for each question prior to the visit by the team leader. In certain cases, interviewing individuals from some of the categories was not appropriate or helpful so the evaluator and the

specialized agency focal point decided not to ask them about the given issue(s). The specialized agency country office contacted the potential interviewees and scheduled interviews in coordination with the local evaluator.

## Data Collection/Analysis

In order to answer the evaluation questions, the cluster evaluation team employed a combination of techniques and different practices, including desk reviews of project documents, interviews with stakeholders, focus groups, field visits, surveys and informed judgment and possibly scoring, ranking or rating techniques.

A mixed-method approach was used by the national evaluators to conduct the evaluation. According to Stufflebeam (1999), use of both quantitative and qualitative methods, a distinguishing characteristic of the mixed-method approach, assures dependable feedback on a wide range of questions; depth of understanding of particular programs; a holistic perspective; and enhancement of the validity, reliability and usefulness of the full set of findings. The evaluation complied with the UN Evaluation Group Norms and Standards for Evaluation.

The evaluation exercise started with the recruitment of three national consultants. After weeks of work undertaken by the national consultants, a field mission to the Arab States region (20 days) was conducted by the Evaluation Unit representative and team leader.

The team leader was then presented with a list of people to be interviewed upon arriving in the country. Interviews were conducted by both the team leader and the national evaluator. Notes from the interviews were drafted by the two evaluators and saved on the same day of the DropBox country folder, which allowed other national evaluators to review the notes as time allowed.

Field visits in each country were conducted for a week by the specialized agency Senior Evaluation Officer, as the lead evaluator, and the Independent National Evaluator. Field visits began with Lebanon, where the specialized agency regional office is also based, followed by Jordan and Palestine.

Once each day's interviews were completed, they were typed and uploaded to the Dropbox folder. Upon completion of each country field visits, the national evaluator shared with the other consultants, lessons learned during field visits and interviews, what worked and what didn't work.

## Lessons Learned

After the site visit in Beruit, Rola shared lessons that she learned with Lamia and Kahlil (see 13 April email below). Lamia did the same. In this manner, each successive site visit was able to benefit from the lessons derived from the previous ones.

April 13, 2013

Dear Lamia and Kahlil,

I hope you are doing well. Craig requested that I share with you how our Beirut mission went and what are some of the lessons learned that you may benefit from and would help you in your country evaluations.

Overall the mission went very well. We met with representatives from all tripartite constituents, as well as the specialized agency team and some UN agencies. When I was selecting the interviewees for the evaluations conducted in a short period of time, I was mostly concentrating on the main stakeholders because this evaluation is high level and assessing the specialized agency strategy or framework in the country/region, rather than a project evaluation. It was important to talk to these stakeholders (workers, employers and government) in order to assess their understanding of DW and at what stage they perceive the work to be at. It was also a way to assess the level of the specialized agency's engagement with them and how they view this relationship, what is being done well and how the specialized agency could support them better. We got some very interesting views and important information. I will not go into that, I'm sure that Craig will fill you in on the details, but I will list some of the things that I think may be relevant to your evaluation when conducting it:

One of the key elements that I thought was important is the communication among the specialized agency and the tripartite constituents, and between the specialized agency Team and Program team. In Lebanon we have the regional office based here (which plays the role of a country office in the absence of one), so it gave me the opportunity to discuss with them face to face what their views are. In Jordan they have program staff and supported remotely by the team based in Beirut, and in the absence of an office there, it would be worthwhile to get a feel on how that

affects communication for program staff among each other and with the regional office. In Palestine, while I understand the specialized agency does have an office, the technical team is based in Beirut, some of whom are not able to travel to Palestine for direct support. How the specialized agency program is supported in Palestine with such a challenge, how they communicate with the regional office is also something to look into.

Some of the evaluation questions are related to the Country Program. They are very applicable to Jordan as they are the only country being assessed that has a strategy; however, for Lebanon, and also in the case of Palestine, this is not the case. It's hard to assess when there's no framework to benchmark against. From our talks with the team we have come to understand that there is an approach being followed in Lebanon. This means that additional questions were added to get answers and some of the evaluation questions were re-worded in ways to be able to answer what the evaluation is trying to answer. There may be a need, especially in Palestine, to go over the questions and reformulate or add supporting questions in order to better understand the specialized agency's decent work approach and the accomplishments in that area. Reading about the projects they do and what they've accomplished on a program level for Lebanon has helped me to understand more and sometimes use a certain project to understand the overall framework in my interviews.

I think these were the main lessons learned here in Lebanon. If I were to do it again, perhaps I would have tried to get more documents from the specialized agency before the mission, instead of having them given to us during the mission. The good thing is I'm located here and if I have any questions I can always go back to the specialized agency and talk to them, as I'm preparing the report.

Also, as you are conducting your evaluations, if you feel the need to ask the regional office questions regarding your countries, I would be happy to go pay them a visit and ask the relevant people on your behalf, if you like.

Good luck and I hope you have a smooth and productive mission. It was great working with Craig. He is easy going and fun to work with. He is also very helpful and knowledgeable; if you have any questions regarding the evaluation that you would like to ask him.

If I can be of any assistance, please do not hesitate to contact me.

Best,

Rola

## Reporting

After drafting the three national reports, the final regional draft report was compiled by the Team Leader and shared with the national evaluators for comments and verification. Each national evaluator provided comments on the compiled draft version and shared her/his comments and suggestions not only with the lead evaluator, but also with the rest of the team.

Despite the fact that each national evaluator conducted research of the national country program that she/he was evaluating, in addition to reading the regional documents, each of the national evaluators was well informed about the other two countries' findings. This allowed each national evaluator to take these findings into consideration when revising her/his own findings in the compiled report and ensuring that the individual reports had the same spirit.

Although some of the recommendations in the final compiled report did not apply to all country programs evaluated, national evaluators were aware of these differences and hence the recommendations were customized to the relevant country program.

Although the three independent national evaluators never physically met during the field mission, during the report writing phase, draft reports were shared and comments were welcomed.

## Post-Evaluation

Two of the three participating national evaluators later met thousands of miles away from the region while at the International Program for Development Evaluation Training (IPDET) held in Ottawa, Canada during June-July 2013.

During the one-month extensive international M&E training program, the two evaluators participated in the same workshops and other activities and reported that they often discussed different evaluation issues related to the cluster evaluation in which they had just participated.

In addition, they held more general discussions about ways in which to cooperate in the future. When the evaluator from Palestine helped establish the Palestinian Evaluation Association (PEA), he also shared lessons learned from establishing a national Voluntary Organization for Professional Evaluators (VOPE) with his Jordanian and Lebanese colleagues.

When the Jordan Development Evaluation Association (EvalJordan) was established during May 2014, the Jordan evaluator was one of the founding members. The Lebanese Evaluation Association (LebEval) is currently being established. Issues related to establishing the new VOPEs were discussed among the four evaluators thanks to the relationship previously established during the cluster evaluation.

When the EvalPartners' and the *International Organization for Cooperation in Evaluation's* (*IOCE*) Call for Proposals for the Peer-to-Peer Support Program was announced, the first Middle East and North Africa (MENA)-based project proposal came from the Palestinian Evaluation Association (PEA) and the Jordan Development Evaluation Association (EvalJordan). They submitted an application together to EvalParners and IOCE to work jointly on a project that aims to foster and enhance the first global, open, researchable and non-organization specific evaluation database [website: www.evaluationdatabase.com], which was initiated by the Palestinian Evaluation Association and launched in Amman, Jordan during the Third General Assembly and Conference of the MENA Evaluators Network (EvalMENA) in April 2014.

Through the project, members of both evaluation associations, including the cluster evaluation national evaluators from Jordan and Palestine, will manage all aspects of the project together as part of the Joint Working Group especially formed to deliver this task.

The cluster evaluation, and the professional relationships that were developed in the process, led to the creation of a community of practice that had implications, not only for the participating evaluators, but also for the entire MENA region. This article is one of the results of the community of practice. The three national evaluators, with assistance from the Senior Evaluation Officer, agreed to jointly write a professional journal article where they would share their experiences from successfully conducting one of the first cluster evaluations in the MENA region.

Furthermore, the participating national evaluators have, in fact, benefited from knowing each other by linking and introducing their colleagues to other evaluators working in their respective countries and organizations. For example, while in Canada, the evaluator from Jordan introduced her colleagues from Jordan's Ministry of Planning, where she manages the M&E department, to the evaluator from Palestine who reported that he still maintains these professional relationships. This community of practice and the professional network built and sustained resulting from the cluster evaluation surely continues as the four participating evaluators often share with each other any evaluation-related materials or opportunities on a regular basis.

## Conclusions

Analysis of the data found above leads to a number of interesting conclusions. First, in the philanthropic sector, cluster evaluation is a programmatic intervention. That is not the way that the approach was used in the context of this evaluation. Instead of evaluation as intervention, it was evaluation as capacity development.

This evaluation took place in the Arab States, a region in which evaluation capacity is perhaps not as institutionalized as other parts of the world. The purpose of using a cluster evaluation in this context was to strengthen the capacities of the national evaluators.

Each of the national evaluators who were contracted had different areas of strength and weakness. By establishing a strong network amongst themselves, they were able to leverage each other's strength and to compensate for any weaknesses.

The national consultants developed each other's capacity and that greatly benefited the evaluation. The individual evaluation capacity went on to strengthen the collective evaluation capacity of the region. The local evaluators have all gone on to be leaders in the formation of their respective regional and national evaluation organizations.

## Recommendations

It is important to communicate expectations for evaluation capacity development in the Terms of Reference, for example, networking among national evaluators.

Cluster evaluation as conceived by the W.K. Kellogg Foundation could take up to four years. Cluster evaluation in the UN system does not have that luxury. However, sufficient time needs to be provided to ensure all relevant documents are available before the field mission (if it can be helped).

It would have been useful to the evaluators to meet, physically or over Skype, sometime before, during or after the mission. Communication among the evaluators is a leverage point that has minimal cost that could produce significant results.

The benefits derived from this type of cluster evaluation may warrant its increased use in the

UN system. However, this would require strong differentiation between cluster evaluation and the evaluation of clusters.

## References

Barley, Z., & Jenness, M. (1993). Cluster Evaluation: A Method to Strengthen Evaluation in Smaller Programs with Similar Purposes. *Evaluation Practice, 14*(2), 141-47.

Jenness, M., & Barley, Z. (1995). Using cluster evaluation in the context of science education reform. *New Directions for Program Evaluation*, *65,* 53-69.

Office of Internal Oversight Services. (2008). Review of results-based management at the United Nations. A/63/268. UN General Assembly. NY.

Russon, C. (2005). Cluster Evaluation. In S. Mathison (Ed.), *Encyclopedia of Evaluation*, (66-67). Thousand Oaks, CA: Sage Publications.

Sanders, J.R. (1997). Cluster Evaluation. In E. Chelimsky & W. R. Shadish (Eds.) *Evaluation for the 21st Century*. Thousand Oaks, CA: Sage Publications.

Sanders, J. (1998). Promising Practices: Cluster Evaluation. *The Evaluation Exchange*,4(2). Boston, MA: Harvard Family Research Project.

Stufflebeam, D. (1999). Foundational Models for 21st Century Program Evaluation. . Kalamazoo, MI. The Evaluation Center's Occasional Papers Series.

UNEG. Evaluation Norms and Standards. United Nations Evaluation Group.

WKKF. (1995). W.K. Kellogg Foundation Cluster Evaluation Model Model of Evolving Practices. Battle Creek, MI: W.K. Kellogg Foundation.

Worthen, B.R. & Schmitz, C.C. (1997 July). Conceptual Challenges Confronting Cluster Evaluation. *Evaluation, 3* (3), 300-319.