

eVALUation Matters

A Quarterly Knowledge Publication on Development Evaluation

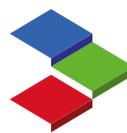


Best Practices and Innovation in Evaluation

Second Quarter 2019

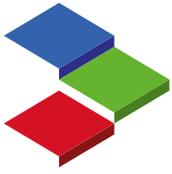


AFRICAN DEVELOPMENT BANK GROUP



IDEV

Independent Development Evaluation
African Development Bank



IDEV

Independent Development Evaluation
African Development Bank

From experience to knowledge...
From knowledge to action...
From action to impact

eVALUation Matters

Is a quarterly publication from Independent Development Evaluation at the African Development Bank Group. It provides different perspectives and insights on evaluation and development issues.

Editor-in-Chief:

Grace Gabala, communications consultant

Acknowledgments:

IDEV is grateful to all contributors, reviewers, editors, and proofreaders who worked on this issue, in particular: Karen Rot-Münstermann; Eustace Uzor; Daniel Alonso Valckx; Debazou Yantio; Boubacar Ly; Kobena Hanson; Aminata Kouma; Marc Guislain Bappa Se.

Editing and translation:

Melora Palmer, Agnès Derelle

Design & Layout:

Créon (www.creondesign.net)

Photo credits:

- ! The African Development Bank Group
- ! The Global Environment Facility (GEF)
- ! The World Bank Group
- ! Nairobi Garage
- ! Shutterstock
- ! United Nations Environment Program (UNEP)
- ! United Nations Office for Project Services (UNOPS)
- ! United Nations Development Program Congo DRC

About Independent Development Evaluation

The mission of Independent Development Evaluation at the AfDB is to enhance the development effectiveness of the institution in its regional member countries through independent and instrumental evaluations and partnerships for sharing knowledge.

Disclaimer:

The views expressed in this publication belong solely to the authors, and not necessarily to the authors' employer, organization or other group or individual they might be affiliated with.

Evaluator General:

Karen Rot-Münstermann (acting) k.rot@afdb.org

Managers:

Foday Turay (OIC) f.turay@afdb.org
Madhusoodhanan
Mampuzhasseril (OIC) m.mampuzhasseril@afdb.org
Karen Rot-Münstermann k.rot@afdb.org

Talk to us:

Phone (IDEV) +225 2026 2841
Phone (AfDB standard) +225 2026 4444

Write to us:

01 BP 1387
Avenue Joseph Anoma,
Abidjan 01, Côte d'Ivoire

E-mail us:

idevhelpdesk@afdb.org

Find us online:

idev.afdb.org
afdb.org

Connect with us on:

 [@evaluationafdb](https://twitter.com/evaluationafdb)
 [IDEV AfDB](https://www.youtube.com/IDEVAfDB)

Second Quarter 2019

In the context of Agenda 2030, all countries are expected to put in place rigorous and effective monitoring and evaluation processes to track progress. Yet countries are at different levels of evaluation capabilities. At the same time, development interventions are becoming more complex, with similar consequences for the evaluations designed to assess them. Thirdly, new technologies and sources of data are increasingly becoming available, and evaluators are continually developing new and more sophisticated methods.

Sharing good practices and innovations in evaluation can help evaluators to learn from each other, to tackle challenges and continually strengthen the profession. This edition of Evaluation Matters aims to showcase selected good, new or innovative evaluation methods that have contributed to better evaluations of development.



News in pictures, page 58

4 **From the Acting Evaluator's General Desk**

Karen Rot Munstermann, Independent Development Evaluation at the African Development Bank

This edition of Evaluation Matters highlights good practices and innovations that have helped make evaluations more useful and impactful. The goal of the edition is to share knowledge of techniques and tools that have proven effective, in order to stimulate further discussion and encourage exploration, so that the field of evaluation can keep moving forward.

8 **Innovations in evaluation of rural development projects: the experience of IFAD's Independent Office of Evaluation**

Oscar Garcia and Hansdeep Khaira, Independent Office of Evaluation, International Fund for Agricultural Development

This article presents three examples of innovations used by the Independent Office of Evaluation at IFAD. These include geo-spatial analysis as a means for data triangulation, the genetic matching method in quasi-experimental ex-post impact evaluations and SenseMaker, a qualitative participatory technique that involves the programme beneficiaries.

18 **Process Tracing as a Methodology for Evaluating Small Sample Size Interventions**

Andrew Anguko, Independent Development Evaluation at the African Development Bank

The author argues that when done 'right', the Process Tracing methodology can tackle questions of attribution and provide evidence to back up causal claims in interventions with small sample sizes.

28 **Innovations in Evaluating Impacts of Environmental Interventions: Approaches and Findings from Independent Evaluation in the Global Environment Facility**

Geeta Batra and Juha I. Uitto, Independent Evaluation Office of the Global Environment Facility

This article presents new tools and methodologies developed by the Independent Evaluation Office of the Global Environment Facility to better evaluate the longer term outcomes, transformative impacts, value for money and sustainability of the GEF's interventions, with a focus on the nexus between the environment and human and social factors.

- 40 **Good Practice: user satisfaction as a criterion in the evaluation of public reforms across Africa : The case of the use of direct deposit payments for employees and civil servants in the Democratic Republic of Congo**

Bobo B. Kabungu, Central Bank, Democratic Republic of Congo

This article underlines the importance of taking into account the satisfaction of end beneficiaries or users in evaluations of public interventions. It illustrates how a recent evaluation of the payroll reform for state agents and officials in the DRC integrated a user survey, and analyzed the results using the double-difference method.

- 50 **The Most Significant Change Approach in Evaluation: A quick guide and its application at the African Development Bank**

Madhu Mampuzhasseril, Independent Development Evaluation at the African Development Bank

The MSC approach in evaluation is a systematic, transparent and participatory method of collecting and analyzing qualitative information about change. In this contribution, the author succinctly describes this technique and highlights the experience of the African Development Bank using this tool.

- 58 **News in pictures**

- 63 **Hot off the Press**

“Since evaluators are not expected to be experts in new methods, evaluation offices need to forge partnerships with the private sector, research institutions and universities who may be the originators of, or experts in, innovative methods.”

Hansdeep Khaira and Oscar Garcia, International Fund for Agricultural Development (IFAD)

Editorial Calendar

Q3 2019
Made in Africa
Evaluations
(Volume 1)

Q4 2019
Made in Africa
Evaluations
(Volume 2)



Find eVALUation Matters at <http://idev.afdb.org/en/page/evaluation-matters-magazine>



From the Acting Evaluator General's Desk

► The “bicycle theory” (which I am very familiar with, coming from a country that has nearly 23 million of them, 1.3 for every inhabitant) states that when riding a bicycle, you must keep moving forward, otherwise you fall over. This theory applies to international development too. Agenda 2030 took a much broader view of development than the previous paradigm, with universal, all-inclusive Sustainable Development Goals that in addition to issues of poverty, water, health, etc. also cover the environment, human rights, partnership, equity and gender equality. At the same time, the magnitude of challenges facing the world, such as climate change, fragility and vulnerability, is growing. These two trends have led to questions of development, and the interventions designed to address them, growing increasingly complex. Developing countries and the bilateral and multilateral development agencies who support them in pursuing economic and social progress must explore new avenues and innovations in order to keep up. Standing still, by continuing to apply the same old methods, is not an option.

The same applies to development evaluation. As development interventions grow more complex, the field of evaluation must also adapt its methods and approaches to remain relevant and useful. New technologies and an increasing array of sources of data offer opportunities but also challenges. How can evaluators harness the opportunities while overcoming the challenges, to keep moving forward?

One way in which the international evaluation community is moving forward is through the ongoing review of the OECD/DAC evaluation criteria, which seeks to modernize the criteria and keep them fit-for-purpose in the era of the SDGs. The strong and steady engagement of large numbers of evaluation stakeholders in this process is testament to the willingness of the field to self-reflect and adapt, and its commitment to continuing to deliver high quality in order to serve development. Moreover, academics and evaluators are regularly developing new and more sophisticated methods to assess the performance of development interventions.

At IDEV, we also continuously seek to improve the quality of our evaluation and knowledge products, paying particular attention to the needs and demands of our end users. One of our three main objectives is to promote learning: learning from evaluations, to improve the design and implementation of future policies, strategies, processes and operations, but also learning from each other. We invest a lot in knowledge and experience sharing, through various channels: our website, social media, creating and disseminating knowledge products, organizing trainings and events, and so on.

This edition of Evaluation Matters is also intended to be an opportunity for sharing and learning. We invited evaluators from different backgrounds to share good or best practices and innovative approaches in evaluation that have led to greater impact in their organizations, helping policy- and decision-makers to learn more about what works well, what does not and why, and to apply that knowledge in decision-making.

The edition aims to contribute to the global body of knowledge on best practices and innovations for a stronger evaluation culture and to enable knowledge sharing and learning from each other.

In this edition, colleagues from the Independent Office of Evaluation at IFAD share their experience in applying three innovative techniques in the evaluation of rural development projects: geospatial analysis as a means for data triangulation, genetic matching in quasi-experimental ex-post impact evaluations, and SenseMaker, a qualitative participatory technique that involves program beneficiaries. Similarly, you will be introduced to some innovative approaches for evaluating the impacts of environmental interventions applied by the Independent Evaluation Office of the Global Environment Facility, including remote sensing and geospatial analysis, a

"The edition aims to contribute to the global body of knowledge on best practices and innovations for a stronger evaluation culture and to enable knowledge sharing and learning from each other".



new framework for evaluating transformational change, and improved methods for examining sustainability.

Two IDEV colleagues also share their knowledge on evaluation methods that they have found useful in some of the evaluations they have worked on: process tracing as a methodology for evaluating small sample size interventions, and the Most Significant Change technique for collecting and analyzing qualitative information about change. Finally, a good practice from the DR Congo is presented: including user satisfaction as a determining criterion for the evaluation of large-scale reforms in Africa. Like with the Most Significant Change technique, the evaluators here applied a participatory method to give greater voice to the beneficiaries of the intervention. Because after all, the beneficiaries are the ones for whom development interventions are undertaken, the rationale for the work that we do.

Our primary objective in publishing this edition of Evaluation Matters is to highlight some best practices and innovations in order to stimulate further discussion and encourage exploration, so that the field of evaluation can keep moving forward. Of course evaluation methods and approaches need to be selected and adapted to the objectives and purpose of the evaluation at hand, and the methods and techniques presented in this edition will not apply in all situations. But we hope that you will find the selected examples informative and useful, and we would love to hear about your own experiences, questions, suggestions and ideas on the topic.

I am signing off as Acting Evaluator General, so this was my last foreword. The new Evaluator General, Roland Michelitsch, arrives early September.

Happy reading!

About the Acting Evaluator General

Mrs. Karen ROT-MÜNSTERMANN is the Acting Evaluator General of the African Development Bank. She joined IDEV in 2014 as manager of its Knowledge Management, Outreach and Capacity Development Division, after spending five years at the Bank's Resource Mobilization and Partnerships Department, where she was among others responsible for coordinating the three-yearly ADF replenishment processes. Before joining AfDB in 2009, Karen was a Senior Policy Advisor at the Dutch Ministry of Finance (Treasury, Foreign Financial Relations Department). She holds degrees in Political Science, European Studies, and Economics.

Innovations in the evaluation of rural development projects: the experience of the International Fund for Agricultural Development's Independent Office of Evaluation

Innovation in the way evaluations are conducted is gaining importance. A host of factors are responsible for this, including: an emphasis on collecting stronger evidence for accountability purposes, advancements in technology and computational power, and a change in perspective concerning the role of beneficiaries in programme evaluations. This article presents three examples of innovations used by the Independent Office of Evaluation (IOE) of the International Fund for Agricultural Development (IFAD). These include geo-spatial analysis as a means for data triangulation, genetic matching methods in quasi-experimental ex-post impact evaluations, and SenseMaker - a qualitative participatory technique that involves programme beneficiaries. As the results obtained from these three methods have been satisfactory, this article suggests that expanding the use of these successful innovations, in addition to forging partnerships with experts and sharing experiences, are all key ingredients to mainstreaming innovations in evaluations.

Oscar Garcia and Hansdeep Khaira, Independent Office of Evaluation, International Fund for Agricultural Development

The need for innovation

While the evaluation community has relied on tried and tested methods of evaluation, an increasing need is being felt to go beyond the "comfort zone" and push the frontiers of evaluation method - to aim for innovation in the way evaluations are conducted. This need has its genesis in several factors.

First, the measurement and assessment of development outcomes is gaining ever-increasing traction, not in the least because of the need to measure and report progress on the Sustainable Development Goals as well as the donor community's emphasis on understanding how their funds are being put to use, i.e. the accountability proposition. This means that better and stronger evidence is needed to attribute programme effects. Second, there is a felt need for cost-effectiveness in evaluations; this emanates from shrinking or static budgets within evaluation offices who, at the same time, are being asked to undertake evaluations of ever-increasingly complex programmes. Third, a paradigm shift in the way beneficiaries of development programmes are viewed – not merely as participants of programmes but also as stakeholders – is giving rise to the use of participatory methods in the design and implementation of programmes, but also in the measurement of their outcomes.

In recognition of the above factors, the IFAD's IOE has made innovation an important part of its evaluations, especially in the past five or so years. It has attempted to take advantage of the leaps in technological advancements and the increasing computational power of

computers to incorporate innovations in methodology. This article shares the learning of IOE from the perspective of three innovative methods that it has used.

Innovations in IOE evaluations

Geo-spatial analysis

Introduction. The wide-scale availability of satellite data has made geo-spatial analysis² an exciting proposition to consider for use in evaluations of rural development initiatives. IOE, in collaboration with the Environment and Climate Division of IFAD, used geo-spatial analysis in an evaluation concerning an agriculture support project in Georgia. The project aimed to rehabilitate aging and dilapidated irrigation water canals so that water could reach the farmers for irrigation of their crops. The development goal of this intervention was to increase farmers' incomes and food security through increases in farm production which they could use for increasing their own consumption while selling the excess in markets. The geo-spatial analysis was used on a pilot basis as a complement to a household survey that was carried out to measure project effects, thus strengthening the triangulation of information sources.

Methodology. The methodology consisted of using time-series satellite imagery to compute temporal variations (before and after the intervention) of the Normalised Difference Vegetation Index (NDVI)³ for comparing the change in vegetation cover between beneficiary farm plots and comparison/control group farm plots, before and after the rehabilitation of irrigation schemes. The rationale being that the intervention would cause ►

► a change in production area (either expansion of the production area or increase in production within the same area) from before to after the intervention, compared with similar areas not affected by project interventions. The output would be an estimate of the magnitude and significance of the difference in "greenness" change in land cover between the intervention area and control areas.

The method used for analysing the data was the before/after control/impact (BACI) index.⁴ The concept is similar to the difference-in-difference method.⁵ The analysis was performed on freely available satellite images using Google Earth, a cloud-based open platform: 250-m NASA MODIS NDVI product (8 days) from 2003 (before-intervention year) to 2016 (after-intervention year). The methodology was completely automatized by developing an algorithm in open source statistical software R.

The first step consisted of analysing the time-series dataset (2003 to 2016) and calculating a multi-annual vegetation development profile, allowing a determination of the vegetation growth period and then classifying the area according to different vegetation development patterns. Only the cluster classes present in the area of intervention (similar land cover and vegetation development patterns) were considered eligible for the analysis.

The second step consisted of assessing the similarity between pixels in the project and control areas. Similarity was defined as the complement of the Root Mean Squared Error (RMSE) between the fractional compositions and one, i.e. similarity $s=1-RMSE$. Values close to one indicated nearly identical overall composition of a control and a treatment site. Twenty control areas with higher RMSE were considered for the calculation of the BACI contrast. In the last step, the impact of the intervention was calculated

as the BACI effect that represented the differential change between project and control areas, when compared before and after the intervention.

The BACI analysis provides two important statistics: the significance level (P-value) of the BACI effect test and the BACI contrast. The BACI contrast is calculated as the difference between project and control subjects, and between the periods of comparison.

$$\text{BACI contrast} = (\mu_{CAa} - \mu_{CAb}) - (\mu_{PRJa} - \mu_{PRJb})^6$$

"Since evaluators are not expected to be experts in new methods, evaluation offices need to forge partnerships with the private sector, research institutions and universities who may be the originators of, or experts in, innovative methods."

By convention, a negative BACI contrast indicates that the variable has increased more (or decreased less) in the intervention site with respect to controls in the time period ranging from before to after the project implementation. The BACI contrast is expressed in the same units of the variable of interest, here NDVI. In order to highlight the magnitude of the contrast with respect to the initial conditions, it was normalised by the mean of the impact area NDVI before the intervention took place and expressed it as a percentage. This derived variable is referred to as "relative contrast".

Results. The results showed a statistically significant negative BACI contrast (i.e. improvement in NDVI of project areas with respect to control areas after the intervention) in 7 out ►



► the covariates are balanced, it becomes much easier to match participants with multiple characteristics.

Although commonly used, PSM is not without its criticisms. For instance, there is no consensus on how exactly matching should be done and how to measure the success of the matching procedure. Further, it is argued that the true propensity score can never be known in observational studies, thereby casting doubt that the propensity score estimates are accurate. Although Rosenbaum & Rubin recommended iteratively checking the propensity score for balance, this can be quite challenging. In order to overcome this limitation, and taking advantage of the growth in computational power, IOE used a different method for its impact evaluation in Georgia, called the Genetic Matching method (Genmatch)⁷. This uses an algorithm to maximize the balance of observed covariates across matched treatment and control units and eliminates the need to manually and iteratively check the propensity score. Such a method is possible due to the increasing popularity of computationally intensive simulation and machine learning methods.

Method. The Genmatch method uses a combination of PSM and Mahalanobis⁸ distance methods, the two methods used to match treatment and control groups on a set of characteristics. It matches samples on their weighted Mahalanobis distances calculated from the distance matrix that includes propensity scores and other functions of the original covariates. Genmatch adopts an iterative approach of automatically checking and improving covariate balance measured by univariate paired t-tests and/or univariate Kolmogorov-Smirnov (KS) tests. In every iteration, weights used in the distance calculation are adjusted to eliminate significant results from the univariate balance tests from the end of the last iteration. The iterative process ends when all univariate balance tests no longer yield progress in increasing p-values. The aim is

to maximise the p-value associated with the covariate which represents the greatest difference between the two samples.

In the case of IOE's impact evaluation, the Genmatch was used for sampling (matching treatment and control clusters, or villages, on covariates). A Genmatch algorithm was used to calculate weights for each covariate and a matching algorithm was then used to identify the most similar communities to the treated communities at the time of the project baseline, prior to treatment.

Results. Although genetic matching generally outperforms PSM, to test whether it did in the case of the impact evaluation, match balance was tested for the samples matched on propensity scores. The results demonstrated that Genmatch led to greater balance on covariates as compared to PSM.

Opportunities and caveats. The main advantage of Genmatch is that it directly optimizes covariate balance. This avoids the manual process of checking covariate balance in the matched samples and then re-specifying the propensity score accordingly. By using an automated process to search data for the best matches, Genmatch is able to obtain better levels of balance without requiring the analyst to correctly specify the propensity score. It makes use of the current advances in computational power. Open source software that implements GenMatch and a variety of other matching algorithms is available for the R programming environment.

The advantage of any new matching method is limited because of the selection on observable assumptions. The plausibility of the assumptions must be carefully scrutinized in each application using evidence beyond the statistical method. In observational studies, key identifying assumptions cannot be tested by simulations nor proven ►

► mathematically. Therefore, more validation studies based on real data are needed to improve observational methods in practice and to clarify the conditions in which these methods are appropriate.⁹

SenseMaker

Introduction. As part of the Country Strategy and Programme Evaluation (CSPE) conducted in the Republic of Cameroon, the IOE evaluation team¹⁰ piloted an innovative approach to confirm the linkages between the support provided by projects and changes in living conditions as perceived by beneficiary households. The study targeted two projects financed by IFAD focused on cassava, rice and onion value chains. It sought to fill the evidence gap regarding the contribution made by project outputs to the changes measured in IFAD impact domains, such as agricultural productivity, incomes and food security.

The approach was based on a participatory methodology called SenseMaker that involves the collection of a large number of brief stories from beneficiaries, recounting one or more notable changes perceived as a result of their participation in producer organizations supported by an

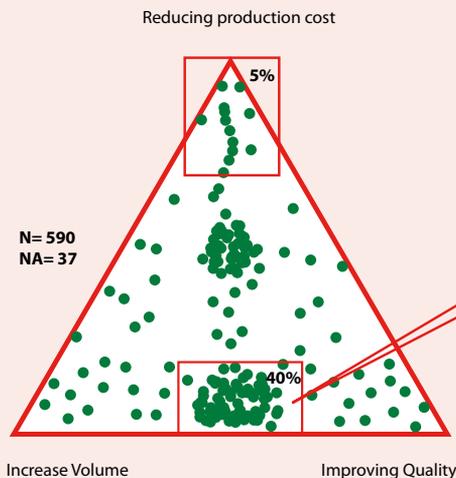
IFAD project. The short stories were then analysed by the respondents themselves through a separate interpretation questionnaire. This lent the analysis greater legitimacy by reducing the bias associated with an external expert’s interpretation of the data.

Method. SenseMaker is based on the collection and indexing of micro-narratives. These anecdotes, experiences or stories are self-signified by the storytellers. This means that respondents assign meaning to their own stories (self-interpretation) immediately after they have shared their anecdotes, experiences or stories through a set of questions (signifiers) rather than an external intermediary interpreting the narratives (common in qualitative approaches). SenseMaker implies the collection of a large number of stories (300+ to thousands) to gain multiple perspectives on the domain of interest. The signification (indexing) of the fragments allows for quantitative pattern analysis backed with explanatory narratives.

Results. In total, 590 stories were collected and self-interpreted from twenty Producer Organizations (PO). The ►►

Figure 2: Respondents’ perception in relation to the production of their crop

Reflecting on the context of your story, where did you see most progress on production issues in the last period ?



I joined the CIG looking for assistance, because i needed to produce more. As it is the CIG that benefits from the projects, fertilizer and seeds, it was important for me to join. When i was not a member of the CIG, i produced between four and six sacks on a quarter of a hectare. In the CIG, i’ve received seeds and fertilizer, and then training on how to sow, prepare the land, and use the fertilizer at the right time(...).
 Young woman, Garoua, North, Rice, 11 April 2017



► opening question to beneficiaries was: Since you have become a member of the common interest group (CIG), can you tell us about an important positive or negative change related to the production, processing or selling/marketing of your crop (onion, rice or cassava) and how this has affected you and your family? Then, an analysis of the responses using dedicated SenseMaker software made it possible to uncover trends embedded in the stories, by positioning the large number of stories on specially designed charts (using visual combinations of multiple choice questions, triads, dyads and stones). The software analysis brought up additional questions which were further explored during four participatory workshops with beneficiaries who took part in the survey.

The beneficiaries interviewed were generally satisfied with services provided by their PO, which were mainly focused on training, processing and storage of produce. At the same time, sustainability issues emerged from the stories, mainly linked to internal tensions in the POs, inadequacy of processing equipment and storage

facilities, and limited availability of seeds and fertilizers after project completion.

Opportunities and limitations. The experience of IOE using a participatory tool involving respondents such as the SenseMaker was positive. The method allowed for quick collection and analysis of qualitative data (cost and time efficient), it provided evidence-based “hard” and “soft” data, and it fit in the evaluation process and adapted to a mixed-methods approach. However, the use of this method also brought forth certain limitations, such as the unavoidability of researcher biases (framework design, theory of change, sampling), the need for technical support at first use and a requirement for the close supervision of data collection.

Conclusions

This article has provided some successful examples of innovations used by IOE in its evaluations. From this exercise, we can conclude that once an innovation ►

► is successful, it should be made to "stick", and its use should be expanded to other evaluations when relevant. The geo-spatial analysis and the SenseMaker have been successfully used by IOE in evaluations of other country programmes as well. A second takeaway is that, since evaluators are not expected to be experts in new methods, evaluation offices need to

forge partnerships with the private sector, research institutions and universities who may be the originators of, or experts in, innovative methods. Finally, is it important for the evaluation community to share its experiences derived from the use of innovative methods to foster collective learning, as this article set out to do.



References

Carbon, M. and Ahmedou, H. (2017). Contribution study of two "value chain" projects. Unpublished manuscript, Country Strategy and Programme Evaluation, Republic of Cameroon.

Diamond, A. and Sekhon, J. S. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies, *The Review of Economics and Statistics* 2013 Vol. 95, 932-945.

Ho, Daniel E., Kosuke I., Gary K., and Elizabeth A. Stuart (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 15, no. 03 (2007): 199-236.

IFAD Independent Office of Evaluation (2014). Impact Evaluation of the Agricultural Support Project in Georgia. International Fund for Agricultural Development, Rome.

IFAD Independent Office of Evaluation (2018). Country Strategy and Programme Evaluation in Cameroon. International Fund for Agricultural Development, Rome.

Meroni et al. (2017). Remote sensing monitoring of land restoration interventions in semi-arid environments with a before–after control-impact statistical design, *International Journal of Applied Earth Observation and Geoinformation*.

Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*. Apr. 1,

Endnotes

1. Geospatial analysis is the gathering, display, and analysis of imagery, GPS, metadata, satellite photography, remote sensing, and historical data of particular areas to identify impact. For example, orbiting satellites can probe the build-up of infrastructure (e.g. number of tents in a refugee camp, roads, irrigation canals), or assess the different wavelengths emitted from the Earth's surface to indicate changes in vegetation, drought (for instance to appraise the risk of famine), or flooding.

2. Normalized Difference Vegetation Index (NDVI) is an index of plant "greenness" or photosynthetic activity.

3. Presented in the research paper: Remote sensing monitoring of land restoration interventions in semi-arid environments with a before–after control-impact statistical design, Meroni et al. 2017.

4. The project effect or impact is calculated as the difference in the value of the variable of interest before intervention and after intervention for both treatment and control or non-treatment groups. The resulting values for both the groups are then subtracted from each other to give the final outcome or effect size.

5. where μ is the site-specific spatial mean of the variable selected to represent the impact (here NDVI); CAa, PRJa stand for control area and project area after project intervention, respectively; CAb and

PRJb stand for control area and project area before intervention, respectively.

6. Presented by Alexis Diamond and Jasjeet S. Sekhon in *The Review of Economics and Statistics*, 2013. Mr Diamond provided methodological guidance to this IOE impact evaluation.

7. The Mahalanobis distance is a measure of the distance between a point P and a distribution D. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. The Mahalanobis distance between two objects is defined as:

$$d(\text{Mahalanobis}) = [(x_B - x_A)^T C^{-1} (x_B - x_A)]^{0.5}$$

Where: x_A and x_B is a pair of objects, and C is the sample covariance matrix.

8. Ibid.

9. The CSPE team comprised Mr. Michael Carbon, IOE Senior Evaluation Officer, and lead evaluator, and Hamdi Ahmedou, IOE consultant.

Authors' profile

Oscar Garcia is the Director of the Independent Office of Evaluation at the International Fund for Agricultural Development (IFAD) in Rome. He is responsible for evaluating the performance of IFAD, with an active portfolio of 230 rural development projects amounting to US\$ 6.9 billion with operations in 99 countries. Before joining IFAD, he served as head of the advisory services for Green Economy at the United Nations Environment Programme –UNEP. Prior to that, Mr. Garcia was senior evaluation advisor at UNDP Evaluation Office in New York, overseeing programmatic and thematic evaluations in Africa, Asia and Latin America. He was also Director General for Trade Policies at the Bolivian Ministry of Economic Development. He holds a Master degree in organizational change management, New School University, New York, an MBA from the Bolivian Catholic University in association with the Harvard Institute for International Development and a BSc in Economics, University of Santa Catarina, Brazil.



Hansdeep Khaira is an Evaluation Officer in the Independent Office of Evaluation (IOE) of the International Fund for Agricultural Development (IFAD). Since 2016 he has worked on a variety of evaluations including country, project and impact evaluations. Previously, he worked in IFAD's Programme Management Department on monitoring, evaluation and performance analysis of development projects. Before joining IFAD, Mr. Khaira worked with the Food and Agriculture Organization in policy research and analysis, statistics and training related to agriculture and food security. His work experience covers over 35 developing countries in sub-Saharan Africa, Near East and North Africa, Asia, Latin America and the Caribbean. He holds a Master degree in Agricultural Economics from the University of London's School of Oriental and African Studies (SOAS) and a Master degree in Business Administration from Mumbai University, India.



Process Tracing as a Methodology for Evaluating Small Sample Size Interventions

Process tracing in evaluation is an approach used to assess the impacts of selected interventions based on qualitative data. It focuses on demonstrating causal inference using in-depth analysis of an intervention. In process tracing, the evaluator's main objectives are to showcase evidence of the extent to which an intervention's key targeted outcomes have materialized and to investigate the causal mechanisms responsible for the outcomes. The main difference between process tracing and other theory-based evaluations is that in process tracing, the theory of change is much more explicit and detailed, and each hypothesized causal relationship is tested using empirical evidence. In this article, the author discusses the relevance and applicability of process tracing in impact evaluations, especially when evaluating interventions with small sample sizes.

Andrew Anguko, *Independent Development Evaluation, African Development Bank*

Introduction

A number of Multilateral Development Banks (MDBs) use theory-based approaches to assess the effects of their interventions in member countries. Unfortunately, this strategy may not capture effects that can solely be attributed to those interventions, given that there is typically a myriad of factors that affect the observed outcomes. In order to claim attribution, a number of MDBs conduct impact evaluations to quantitatively estimate, using statistical techniques, the effects of their interventions. In this case, the interventions are evaluated by comparing data collected from both intervention and comparison populations, coupled with the application of statistical methods to control for observable differences between them. These so-called **counterfactual methods** may be applicable in interventions where there are quantitative indicators that can be reliably measured to demonstrate effect, i.e. where the sample size is large enough to achieve statistical significance. In some interventions, such as those that focus on governance, policy, budget support or advocacy, however, you may not have the **necessary sample size** that is needed to use statistical methods.

A qualitative causal inference method known as **process tracing**, on the other hand, may be used to assess the impact of interventions with small sample sizes. Process tracing is a qualitative research method that attempts to identify the causal processes – the causal chain and causal mechanism – between a potential cause or causes and the outcome. It is a robust technique that tests different theories of causality-in-action by examining the intervening steps. It is used to “unwrap” the causal links that connect independent variables and outcomes, by identifying

the intervening causal processes. It does this by testing alternative hypotheses against the available evidence and attempting to narrow down the number of alternative explanations.

While process tracing may not be able to exclude all but one theory in a given case, it can reduce the range of possible explanations and can disprove claims that a single variable is necessary or sufficient to produce an outcome. Process tracing offers a rigorous approach to assessing causal change, and the potential for examining causality in programmes where attribution is difficult, by providing evidence on how and why an intervention led to change.

Measuring Impact of Development Interventions

Counterfactual methods are well-known in the development community for measuring the impact of interventions (Ravallion 2009). While more thorough and technical descriptions can be found elsewhere (Shadish, Cook et al. 2002; Duflo, Glennerster et al. 2008), the basics are as follows: We want to know the extent to which a particular intervention has affected a particular outcome, e.g. household income. If it were possible to know what the status of this outcome would have been in the absence of the intervention, we could compare it (known as the counterfactual outcome) with the observed outcome. The difference between the two would be the intervention's effect.

We can, of course, never really know for certain what would have happened to a particular individual, household, community, etc... had we never intervened. However, the situation is different if the number of units we are targeting is large. Specifically, if we were to randomly assign a significant number of units ➔

► to both intervention and control groups, the statistical distribution of their characteristics – particularly those that affect outcome – will be very much the same. As such, we can use the observed outcome of the control group to estimate the counterfactual outcome of the intervention group. In the language of the impact evaluation literature, both groups have the same potential outcomes (Morgan and Winship 2007).

However other impact assessment designs that do not randomize intervention exposure, such as process tracing, have also emerged and are being applied in evaluation of development interventions to generate useful evaluative information for policy makers.

Process Tracing as a Methodology for Evaluating Small Sample Size Interventions

As discussed above, while counterfactual analysis may be a popular evaluation method for large n interventions, it may not be possible or appropriate for MDBs to apply it in all their evaluations. Even if we can overcome the ethical and political hurdles associated with randomization, such designs are expensive and often very challenging to successfully implement. However, over the last several decades, significant developments have taken place in drawing causal inferences from non-experimental or observational data (Imbens and Wooldridge 2009).

It is clear that the RCT design is inappropriate when the number of units being targeted is small, e.g. policy, budget support or advocacy interventions in country X. Large numbers of units need to be randomly assigned to intervention and control groups, so that both groups are statistically equivalent. In fact, the more heterogeneous the population, the greater the number required. If we were only targeting a few units, randomly assigning

their exposure to a given intervention would be futile from a causal inference point of view; the two groups would, more than likely, simply be too dissimilar to be comparable.

Fortunately, the counterfactual outcomes framework is not the only approach to credible causal inference (Brady 2004; Hedström 2008). There are other approaches that are more appropriate for small n interventions, one of which is **Process Tracing**.

The best-case scenario is when counterfactual and mechanism-based approaches are used together, i.e. where there is both a rigorous estimation of what would have happened in the absence of the intervention, and strong evidence of what mechanisms were at work to bring about the change (Reynolds 1998). Unfortunately, as mentioned above, the former approach is not suitable for small n interventions. Such interventions, then, must rely primarily on the latter, and this is the impact assessment approach that MDBs such as the African Development Bank may pursue for their governance, budget support, policy influencing, advocacy and advisory/technical assistance interventions.

Evaluation Questions in Process Tracing

It is important that we ask “What evidence would we expect to find if change happened in the ways we predicted, and did we see it?” Process Tracing tries to understand how change happened rather than simply validating the Theory of Change. What outcomes/impacts have actually materialized? Is there evidence that we contributed? And what can we learn about the significance of our contribution? Process Tracing focuses on the following:

- Were the activities carried out?. What evidence is there that the activities were conducted? ►

- ▶ ■ Were the relevant outputs produced? What evidence is there that the relevant outputs were produced?
- What evidence is there for the achievement or otherwise of the intended outcomes?
- What evidence is there for the intervention's contribution to these outcomes?
- How significant is this contribution, compared with other possible contributing factors?

The end goal is to see whether results are consistent with the program theory (theory of change) and/or to see whether alternative explanations can be ruled out.

Methodological Considerations

The Methodology of Process Tracing

Process tracing involves evidencing the specific ways in which a particular cause produced (or contributed to producing) a particular effect. An important component of process tracing is the consideration of alternative, competing explanations for the observed outcome in question, until the explanation(s) most supported by the data remains (Patton 2008a). If these alternative explanations have already been identified, "process verification" is directly undertaken. This involves considering, specifying, and documenting what kinds of evidence, if found, would either validate or exclude each of these alternative explanations.

However, in many cases, some or all of the possible and plausible explanations for the observed outcome will not have been identified in advance. Then "process induction" is undertaken first. This involves undertaking exploratory, inductive research to identify plausible alternative explanations, which are then

developed into explanations that are more thorough, i.e. into hypotheses that can be tested via "process verification," as explained above.

Process tracing is a qualitative method that seeks to evaluate impact through establishing confidence in how and why an intervention worked and for whom. A distinctive feature of process tracing is that it draws on a generative framework to provide a detailed description of a causal mechanism that led to a specific effect, and by doing so demonstrate the causal relation.

"Process tracing is a qualitative research method that attempts to identify the causal processes – the causal chain and causal mechanism – between a potential cause or causes and the outcome"

In process tracing, the purpose of the evaluation is not simply to focus on only one explanation for an observed outcome-level change. Rather, the approach is more nuanced and should accomplish three things: 1) shortlist one or more evidenced explanations for the outcome in question; 2) rule out alternative, competing explanations incompatible with the evidence; and 3) if more than one explanation is supported by the evidence, estimate the level of effect each has had on bringing about the change in question.

The evaluator seeks evidence of the extent to which the intervention's key targeted outcomes have materialized; investigates the causal mechanisms responsible, i.e. how the observed outcome change came about; and, in light of an evidenced understanding of competing explanations, draws conclusions about the significance, if any, of the intervention's ▶



► contribution. This evaluation method uses secondary sources, key informant interviews, and community focus group discussions as sources of evidence. Furthermore, the evaluator develops a data collection matrix, which specifies the outcome, method of data collection and data sources.

Process tracing therefore works through affirming explanations that are consistent with the facts and rejecting those that are not. This is much like a detective who pursues possible suspects and clues, “constructing possible chronologies and causal paths both backward from the crime scene and forward from the last known whereabouts of the suspects” (Bennett 2008).

The Process Tracing Protocol

While not intended to be a mechanical sequence of linear steps of how the research exercise should proceed, the following eight steps form the core of the process tracing protocol.

1. Undertake a process of (re)constructing the intervention’s theory of change, in order to clearly define the intervention being evaluated – what is it trying to change (outcomes), how is it working to effect these changes (strategies/streams of activities) and what assumptions is it making about how it will contribute to these changes (key assumptions).
2. Work with relevant stakeholders to identify up to three intermediate and/or final outcomes considered by stakeholders to be the most significant for the evaluation to focus on (central to the intervention’s theory of change, and useful for learning/forward planning).
3. Systematically assess and document what was done under the intervention to achieve the selected targeted outcomes.

4. Identify and evidence the extent to which the selected outcomes have actually materialized, as well as any relevant unintended outcomes.
5. Undertake ‘process induction’ to identify salient plausible causal explanations for the evidenced outcomes.
6. Gather required data and use ‘process verification’ to assess the extent to which each of the explanations identified in Step 5 are supported or not supported by the available evidence. Looking at these sources in terms of the sequence and structure of events can serve as evidence that a given stimulus caused a certain response in a case.
7. Write a narrative analytical report to document the above research processes and findings.
8. Summarize aspects of the above narrative analysis by allocating project/campaign ‘contribution scores’ for each of the targeted and/or associated outcomes.

The advantages of using the process tracing approach are that: 1) it offers a rigorous approach to assessing causal change and 2) the potential for examining causality in programmes where attribution is difficult, by providing evidence on how and why an intervention led to change. In terms of limitations of this approach, we note that the evaluator has less control, resulting in a process which is more unpredictable and context-dependent. As a result, in spite of the evaluation team’s best efforts, results might still be inconclusive if the evidence collected cannot fully support a causal sequence. To thoroughly test alternative hypotheses, the evaluator needs to have access to a range of stakeholders, data sources and to published and unpublished material. ►►

► **Application of the Process Tracing Methodology in the evaluation of the ACCRA project**

The application of the above protocol is illustrated in the evaluation of the African Climate Change and Resilience Alliance project in Ethiopia (the ACCRA project) by Oxfam GB. The main goal of this project was to promote local adaptive capacity development by advising governance changes at a system level. The following strategies were used in implementing the intervention:

- Policy advice by being accepted as trusted advisers and long-term partners.
- Systemic intermediation – seeking to strengthen and/or realign vertical and horizontal connections within the disaster risk reduction governance system.
- A responsive and flexible approach to capacity building.
- Action research and learning.

Overall, the impact evaluation focused on three key questions:

- What evidence is there for the intended transformation (of governance systems in Ethiopia in order for them to support climate adaptive capacity development, and also become more gender-sensitive and people-centred)?
- What evidence is there for a contribution to this transformation, if any, by ACCRA?
- How significant is this contribution, compared with other possible contributing factors?

It identified two concrete outcomes selected and agreed between ACCRA Ethiopia and the ACCRA International Programme. For purposes of illustration, we shall use one outcome (outcome 1).

Outcome 1: Adaptive capacity building and frameworks mainstreamed into Disaster Risk Reduction governance, supporting a more decentralized and participatory approach. ►►



► During a participatory workshop with Oxfam GB Ethiopia team members, the following steps were undertaken.

1. Defining the outcome the campaign was seeking to bring about (outcome 1);
2. Assessing whether there is evidence to suggest that the desired outcome actually materialized and to what degree;
3. Identifying salient causal stories that explain how the desired outcomes may have been realized;
4. Assessing the ACCRA's contribution to the achievement of the observed outcome, considering other plausible, alternative factors.

Based on coding of key informant interviews and secondary sources, the following causal stories or alternative hypotheses were identified as potential causal explanations for the realized outcome 1.

- **Causal story 1:** The ACCRA interventions played a leading role in the realization of outcome 1
- **Causal story 2:** The World Food Programme played a leading role in the realization of outcome 1
- **Causal story 3:** Other members of the ACCRA consortium played a leading role independently of ACCRA in the realization of outcome 1

5. **Causal story 4:** Systemic contribution by a combination of actors played a leading role in the realization of outcome 1

Empirical evidence was collected on each of these four causal stories or hypotheses to determine their relative contributions to outcome 1. Based on the quality and strength of the evidence, the four stories were assigned contribution scores based on the key below:

Evidence based on interviews with key informants as well as secondary sources suggested that the first causal story or hypothesis offered an important explanation for how Outcome 1 materialized, and it was given a contribution score key of 5. The evidence further found that the other hypotheses contributed little to the realization of the outcome, and so they were assigned lower contribution scores.

Even though several actors including the World Food Programme and Save the Children (independently from ACCRA) played some role, their role was not sufficient for the realization of outcome 1 based on the empirical evidence collected. ACCRA appears to have played a key role in brokering a joint understanding between the Ministry of Environment, Forest and Climate Change and the Disaster Risk Management and Food Security Sector of the value of collaborating, focusing this on the value of mainstreaming Climate Resilient Green Economy and Disaster Risk Reduction together into woreda Annual Development Plans. ►►

Score key	Specific Contribution of intervention
5	Outcome realized in full Evidence that intervention made a crucial contribution
4	Outcome realized in part and evidence that intervention made a crucial contribution Outcome realized in full and evidence that intervention made an important contribution
3	Outcome realized in part and evidence that intervention made an important contribution
2	Outcome realized in part and evidence that intervention made some contribution Outcome realized to a small degree and evidence that intervention made an important contribution
1	Outcome realized, to any degree, but no evidence that intervention made any contribution

► Validity of Findings in Process Tracing

In process tracing, the validity of evaluation findings is ensured by triangulating evidence from a range of data generation methods and sources: document analysis, in-depth interviews, focus group discussions, sense-making meetings with implementing partners, staff and beneficiaries, and workshop meetings with stakeholders, as in the above example.

Data interpretation is also theory-informed, and draws on the experiences of the evaluators using retroductive analysis. Potential biases in analysis are managed and curtailed through the use of feedback processes that involve stakeholders, staff and beneficiaries. Finally, the validity of findings is enhanced through the evaluator holding an inception meeting with the client to develop a common understanding of the assignment and establishing and utilizing a client-evaluators reflection and feedback platform through which they share progress, methodological reflections and changes that grow out of field-based experiences.

Conclusion

Process tracing involves tracing causal mechanisms using in-depth case studies that provide within-case, mechanistic evidence of causal processes. It involves theory testing, theory building and explaining the outcome.

Theory-based impact evaluation cannot rival the rigour with which well-designed counterfactual impact evaluation addresses issues of attribution. However, done 'right', process tracing can tackle the issue of attribution and provide evidence to back up causal claims in interventions with small sample sizes. By emphasizing that the causal process leads to certain outcomes, process tracing lends itself to validating theoretical predictions and hypotheses.

However, there is the possibility that the evidence available will not be sufficient to verify or eliminate all investigated explanations. It is possible, then, for the findings of such studies to be inconclusive. Hence the importance of using various lines of evidence and involving stakeholders throughout the process.



References

- Brady, H., 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Rowman & Littlefield Publishers, Inc.
- Dr. John Colvin et al, 2018. *Governance in Ethiopia, Impact evaluation of the African Climate Change and Resilience Alliance (ACCRA) project Evaluation Report*, Oxfam International Oxford, UK accessed 20 August 2019, <https://oxfamilibrary.openrepository.com/bitstream/handle/10546/620412/er-governance-ethiopia-effectiveness-review-090218-en.pdf?sequence=7&isAllowed=y>
- Duflo, E., Glennerster, R., et al., 2008. Using Randomization in Development Economics Research: A Toolkit. In: T. P. Schultz and J. Strauss: *Handbook of development economics*. Amsterdam: Elsevier, 4, 3895-3962.
- Hedström, P., 2008. Studying Mechanisms to Strengthen Causal Inferences In Quantitative Research. In: J. M. Box-Steffensmeier, H. E. Brady and D. Collier, eds. *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.
- Imbens, G. W. and Wooldridge, J. M., 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47 (1), 5-86.
- Khandker, S., Koolwal, G., et al., 2010. *Handbook on Impact Evaluation: The World Bank*.
- Morgan, S. and Winship, C., 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.
- Patton, M., 2008a. Advocacy Impact Evaluation. *Journal of Multi-Disciplinary Evaluation*, 5 (9), 1-10.
- Ravallion, M., 2009. Should the Randomistas Rule? *The Economists' Voice*, 6 (2).
- Reynolds, A. J., 1998. Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation*, 19 (2), 203-221.
- Shadish, W. R., Cook, T. D., et al., 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.

Author's profile

Andrew Anguko is Chief Quality and Methods Advisor with Independent Development Evaluation at the African Development Bank, where he works with colleagues on methodological approaches and quality assurance of tools and evaluation products for evaluations in the Independent Development Evaluation of the Bank. Andrew previously worked as M & E Manager for the USAID funded Kenya water, sanitation and hygiene project in Nairobi. He also worked as a Global Impact Evaluation Adviser at Oxfam GB where he provided specialist advice on tools, methods and processes for undertaking rigorous impact evaluation on Oxfam's projects in order to capture and communicate Oxfam's effectiveness as an organization and promote effective learning. Andrew previously worked as a Senior Monitoring and Evaluation Adviser with Danya International in Kenya, as Monitoring and Evaluation Specialist with the Malaria Consortium in Kampala, Uganda, and as Senior Data Analyst with the American Centre for Disease Control and Prevention (CDC), where he was involved in randomized control trials on health. Andrew graduated with an M.Sc. in Biostatistics and Epidemiology from the University of the Witwatersrand in Johannesburg, South Africa and an M.Sc. in Medical Statistics from the University of Nairobi, Kenya. He also holds a B.Sc. in Forestry, a Postgraduate Diploma in Education from Moi University in Kenya and a Certificate in Longitudinal Data Analysis from the University of Colorado, Boulder.





Innovations in Evaluating the Impacts of Environmental Interventions: Approaches and Findings from Independent Evaluation at the Global Environment Facility

This article presents new tools and methodologies developed by the Independent Evaluation Office of the Global Environment Facility to better evaluate the longer term outcomes, transformative impacts, value for money and sustainability of the GEF's interventions, with a focus on the nexus between the environment and human and social factors.

Geeta Batra and Juha I. Uitto, Independent Evaluation Office of the Global Environment Facility

Introduction

The Global Environment Facility (GEF) was set up in 1991 as an international partnership to address pressing global environmental problems. Since then, the GEF has provided over \$17.9 billion in grants and mobilized an additional \$93.2 billion in co-financing for more than 4,500 projects in 170 countries. Projects focus on biodiversity, chemicals and waste, climate change, land degradation and international waters. The sixth replenishment of the GEF was recently completed in 2018 under the auspices of a dramatically changing environmental finance landscape (IEO, 2018a).

In the last decade, new funds, such as the Climate Investment Funds and the Green Climate Fund, and new agencies, such as the Asian Infrastructure Investment Bank and the New Development Bank, have been established; while existing multilateral institutions have invested significant resources to address climate change. In addition, targets for the Sustainable Development Goals have been established which reflect the interrelationships between environmental, social and economic goals. GEF programming has also shifted to create greater integration in the capture of synergies across various focal areas, while maintaining the obligations to the Conventions it serves.

Against this backdrop, independent evaluation by the GEF has also evolved, developing and applying a broad spectrum of tools and methodologies to better evaluate the GEF's longer term outcomes, transformative impacts, value for money, and sustainability. This article demonstrates the application of qualitative evaluation approaches combined with remote sensing, to address important evaluation questions

on the relevance, effectiveness and efficiency of GEF interventions drawing on recently completed evaluations by the Independent Evaluation Office (IEO) of the Global Environment Facility. The IEO is increasingly applying remote sensing and geospatial analysis methods to evaluate environmental outcomes, thus utilizing the opportunities made possible by satellite data, effectively combining this with qualitative information and field verification. Geospatial approaches have helped us address several limitations in our evaluations: they provide basic location data of the intervention, fill gaps in baseline data availability, and provide outcome information on important environmental variables over long term trends at a relatively low cost. Drawing on the existing literature, we also present a framework we developed for evaluating transformational change, which is increasingly being applied ex-ante to determine projects' potential for transformative change. Finally, we present the "value for money" analysis that we developed to examine the returns to GEF investments, and improved methods for examining the sustainability of GEF interventions.

"Geolocating projects can provide valuable insights into whether interventions are being implemented in areas which need them the most".

Addressing Relevance with Geospatial data

Geolocating projects can provide valuable insights into whether interventions ➤

Figure 1: Geolocations of GEF Biodiversity projects and Key Biodiversity areas



Study the impact of GEF support to 1292 global protected areas across 147 countries.

► are being implemented in areas which need them the most. Figure 1 shows the location of GEF protected areas overlaid with important biodiversity areas, carried out as part of an IEO study of GEF support to 1292 global protected areas across 147 countries. The results depict the number of GEF protected areas which fall in areas with significant biodiversity, and in this case suggest that GEF biodiversity projects are relevant and are mainly implemented in key biodiversity areas.

environmental impact (Lech et al, 2018). They provide reliable and cost-effective baseline information, help detect changes over time, and track progress toward the achievement of convention targets. We present the findings from GEF interventions in Lake Victoria as well as biodiversity activities in Jordan. Given scarce resources and time constraints, remote sensing and geospatial data and tools are valuable in complementing other evaluation methods. These tools have the potential for use in ecological forecasting, which can then be used in ex-ante assessments of forest cover, habitat quality, and carbon sequestration at a fine scale.

Using Remote Sensing to Measure Environmental Outcomes

Remote sensing and geospatial methods are useful, innovative tools for measuring

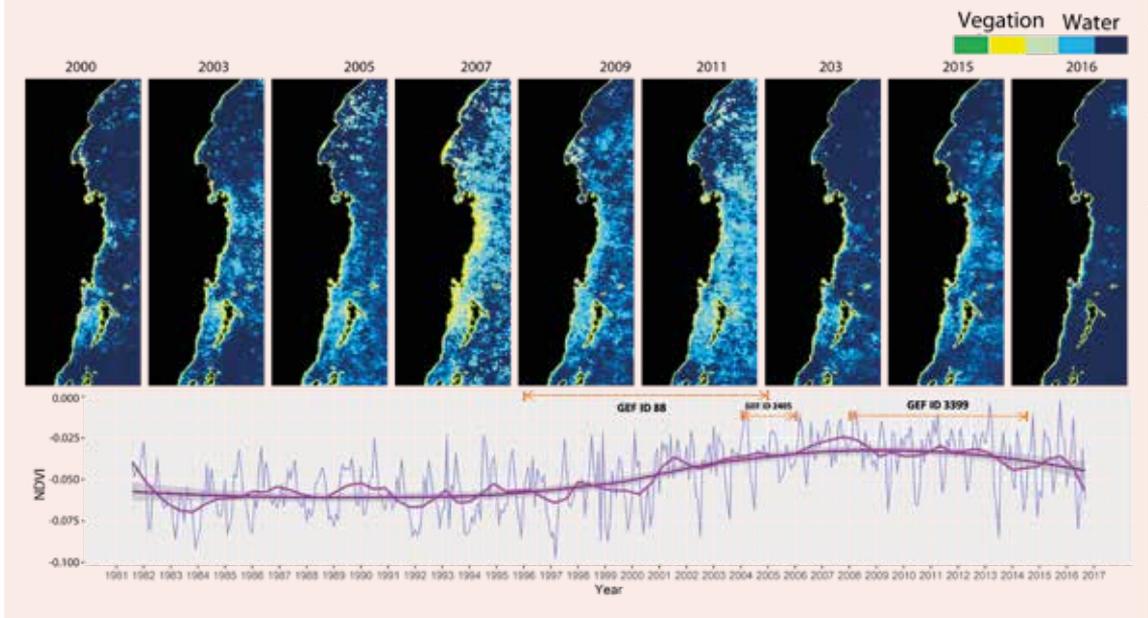
In the evaluation, remote sensing methods were used to observe changes in ►

BOX 1. IMPACT OF GEF INTERVENTIONS IN LAKE VICTORIA

Lake Victoria, with a surface area of about 68,800 km², is the second largest freshwater body in the world. It is a valuable transboundary resource shared by Kenya, Tanzania, and Uganda. In addition, Rwanda and Burundi are part of the upper watershed that drains into Lake Victoria through the Kagera River.

The water hyacinth is an invasive weed that was first reported in Lake Victoria in 1988. It quickly spread across the lake, cutting off communities and putting the economic and food security of millions at risk. Over the past two decades, the GEF has supported Lake Victoria communities by addressing major threats to the lake's ecosystem, including clearing the water hyacinth on site as well as nutrient load management in the upstream areas such that the nutrient load is lessened in the lake.

Figure 2: Vegetation productivity trend around Lake Victoria



► hyacinth infestation (figure 2). By the end of 2016, satellite data observed vegetation productivity - measured in terms of the normalized difference vegetation index (NDVI) – and showed that overall vegetation in Lake Victoria had entered a decreasing phase (IEO, 2016a).

The results show consistent improvement in vegetation cover around all the reserves included in the project. For example, in the Al Hashemiah reserve, the vegetation growth trend has visibly improved since 2013 (figure 2) with the average summer vegetation productivity (NDVI) in 2015 increasing by about 10 percent over pre-project 2012 levels. Overall, the vegetation significantly improved inside the range reserve as compared to outside the range (IEO, 2016 c).

Evaluating Transformational Change

The GEF 2020 vision and strategy identified the need for transformational change to address environmental pressures in order to enhance GEF’s impact. We developed an evaluative approach to assess GEF interventions which were transformative - defined as those which helped achieve deep, systemic, and sustainable change with a large-scale impact in an area of global environmental concern (IEO, 2017). We applied four criteria that permit a differentiation between transformational interventions from engagements that are “merely” highly successful, complex or large in size (World Bank Group, 2016): ►

BOX 2. IMPACT OF GEF INTERVENTIONS IN BIODIVERSITY IN JORDAN

The Badia region in Jordan is a desert ecosystem spanning 80 percent of the country’s area; it is administratively divided into northern, middle, and southern parts. The Badia Ecosystem and Livelihoods Project (BELP) is designed to enhance ecosystem sustainability and local livelihoods through a number of strategic interventions. These include investing in ecotourism and land use planning in the north, in addition to developing water harvesting infrastructure, rangeland reserves, and diversification of livelihoods in the south where livestock is the primary income-generating activity. To observe progress in the rangeland revegetation program around these reserves, dense time-series remote sensing data from NASA satellites was analyzed.

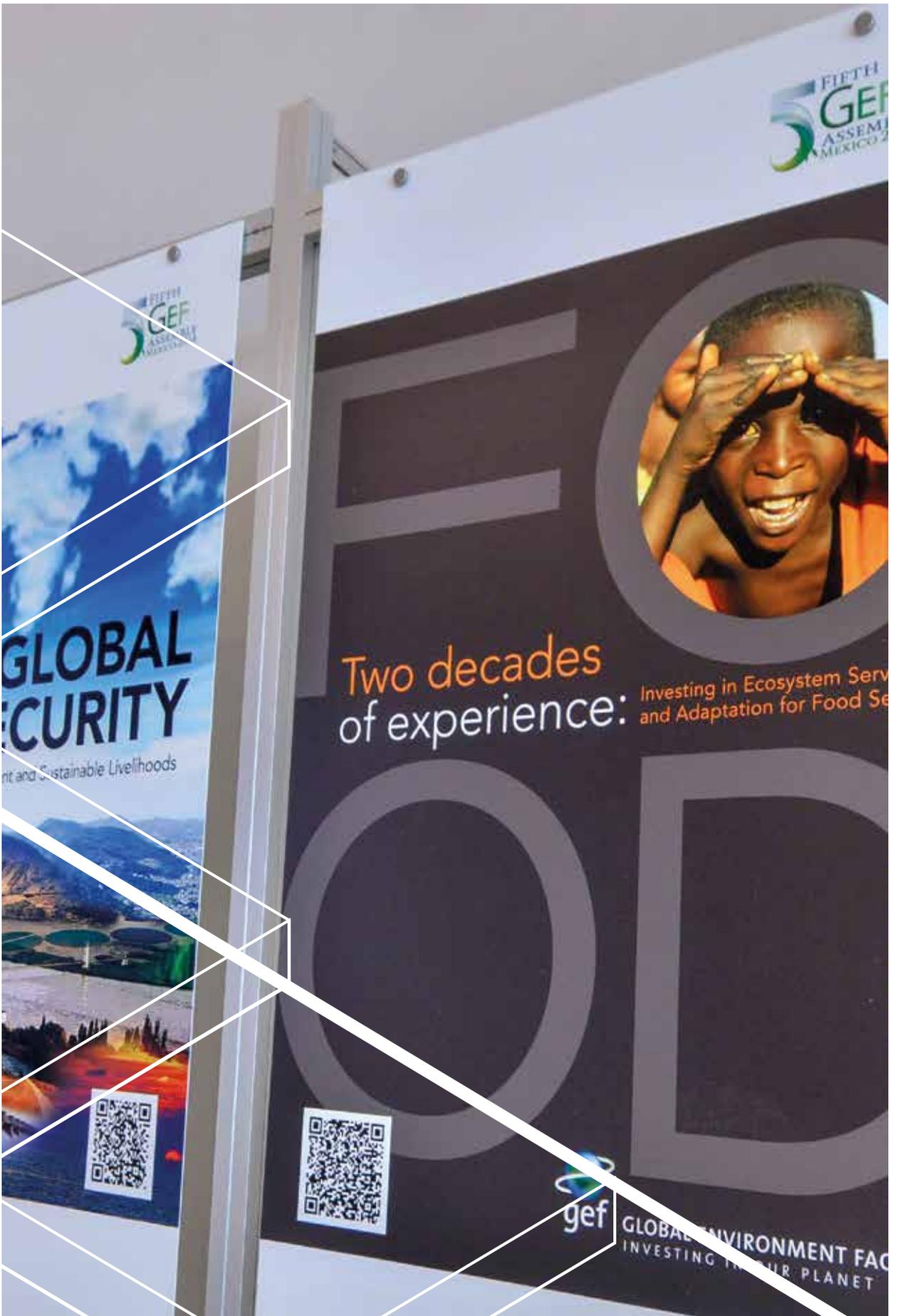
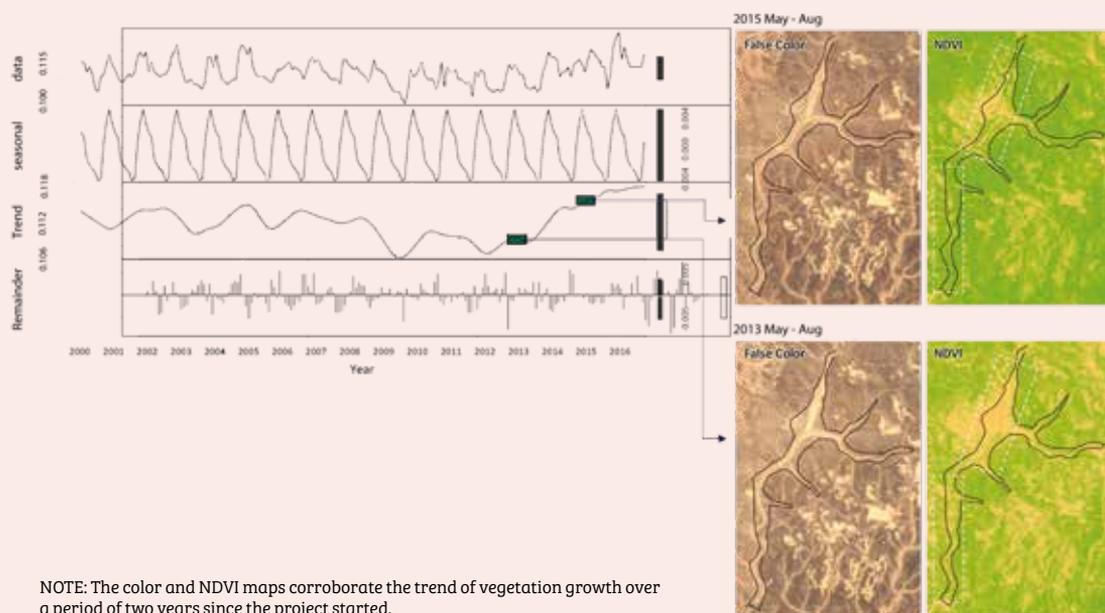


Figure 3: Vegetation growth trend around the Al Hashemiah reserve



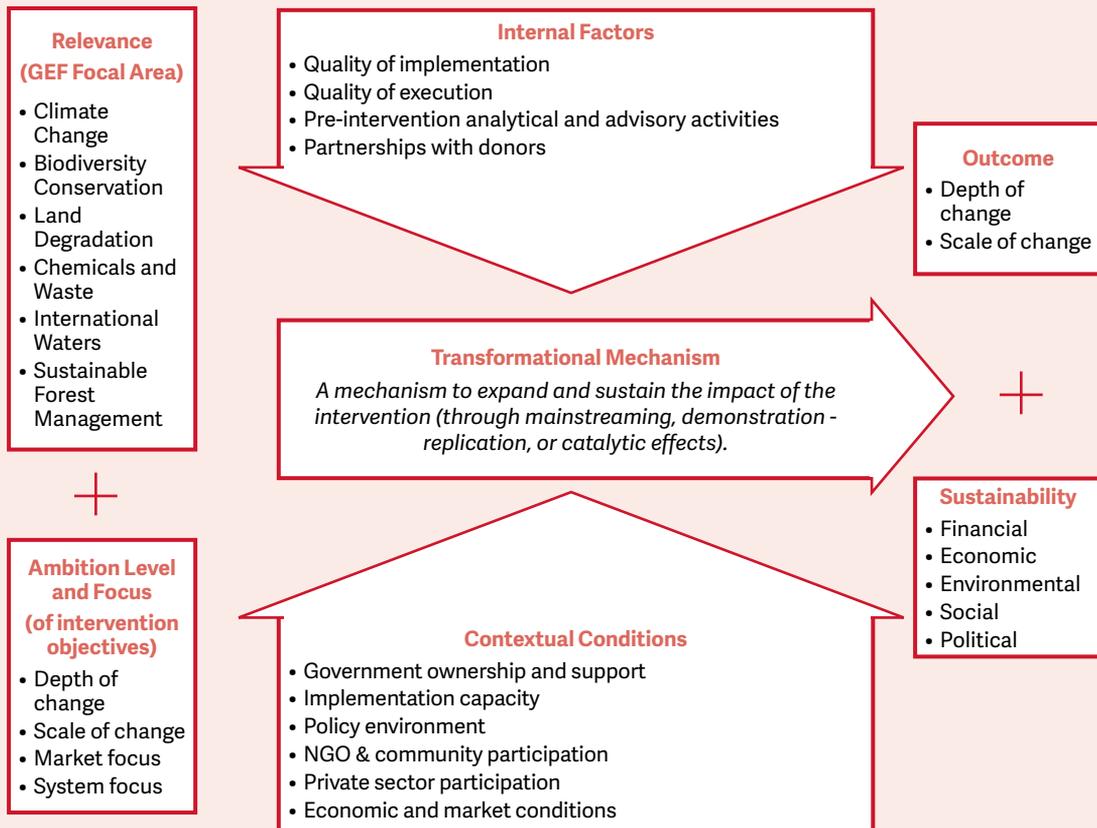
- ▶ ■ **Relevance:** the intervention addresses a global environmental challenge such as climate change, biodiversity loss, or land degradation.
- **Depth of Change:** the intervention causes or supports a fundamental change in a system or market.
- **Scale of Change:** the intervention causes or supports a full-scale impact at the local, national, or regional level.
- **Sustainability:** the impact is financially, economically, environmentally, socially and politically sustainable in the long term, after the intervention ends.

The underlying Theory of Change is that by strategically identifying and selecting projects that address environmental challenges of global concern and are purposely designed to ‘flip’ fundamental changes in key economic markets or systems, GEF interventions will be more likely to cause a large-scale and sustainable impact, subject to the quality of implementation/execution and supportive contextual conditions.

An outline of the Theory of Change, and the main causal conditions and indicators used for this study, are shown in Figure 4.

Applying the approach, the evaluation found a number of examples of transformational change generated by GEF projects:

- In 2016 Uruguay generated about 33% of its total electricity needs from wind power, up from 0% in 2008.
- Between 2005 and 2015, China’s wind power capacity increased from 1.3 GW to 129.3 GW thus producing about 3.3% of its electricity, and avoiding about 82.7 million tons/year of carbon emissions.
- Management effectiveness was improved in about 98% of Namibia’s protected areas, while estimated populations of lions, leopards, cheetahs and wild dogs doubled between 2004 and 2012.
- About 1.3 million households in remote, off-grid areas of Africa have ▶

Figure 4: Theory of Change for GEF Transformational Interventions

▶ purchased quality-certified solar PV lanterns at market prices through a market transformation scheme supported by the Lighting Africa program.

■ About 13 “strict protection” areas totaling 13.2 million hectares, and 30 “sustainable use” protected areas totaling 10.8 million were created with the support of the Amazon Region Protected Areas Program.

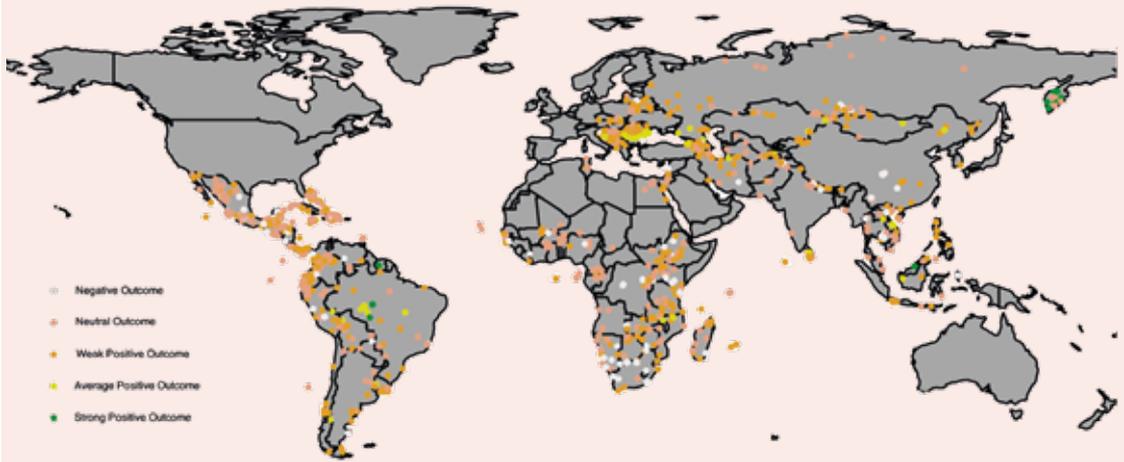
All of the above completed transformations involved a fundamental system change. They all established a demonstration-and-replication mechanism to trigger and scale up the supported activities and reforms. They were satisfactorily implemented and executed in addition to being adequately supported through good policy and a

positive economic environment. Finally, the evaluation found that ambition for transformation is important at the outset, but size is not important—medium-size projects can be just as transformational as major, multi-phase investment projects.

Does GEF Deliver Value for Money? Measuring the efficiency of GEF interventions in Land Degradation and Biodiversity

Increasingly donors are interested in understanding the efficiency of GEF investments. We developed a “value for money” analysis to examine the returns to GEF investments in land degradation and biodiversity interventions. This novel approach involved a multi-step approach of geocoding project locations and ▶▶

Figure 5: Impact of treatment in Biodiversity Projects



NOTE: Estimated impact of GEF biodiversity projects on NDVI. Strong outcomes are observed in Eastern Europe; neutral to negative outcomes tend to be clustered in Southern and Central Africa.

► combining outcome information on forest cover and forest fragmentation with other explanatory environmental variables such as temperature and proximity to infrastructure. Project locations were contrasted with geographic locations where no known intervention occurred (IEO 2016b) Propensity score matching, machine learning and causal tree approaches were used to understand the factors influencing environmental outcomes across project and control locations, impact estimates were constructed and valuations for carbon sequestration were estimated using a value transfer approach.

Land degradation analysis findings showed that the range of potential benefits from a single-focal area land degradation project is estimated at \$52–\$143/ha affected in terms of carbon sequestration alone. At the same time, soil retention promotes an additional value of \$10–\$43/ha, for a total valuation of \$62–\$186/ha across all land degradation projects. After all costs are accounted for, it is estimated that the per dollar return on investment for land degradation projects is approximately \$1.08 per dollar invested. However, this is likely to be an underestimate, as it only captures two

ecosystem services. In addition, the initial state of the environment is a key driver in GEF impacts, with GEF projects tending to have a larger impact in areas with a poor initial condition.

In the case of biodiversity, the results show that globally, GEF biodiversity projects tend to have a positive impact. A range of \$60–\$166/ha of affected area is estimated for carbon sequestration; an additional value of \$10–\$41 is estimated as attributable to soil retention benefits, for a total of \$70–\$207/ha. Geographically, impacts on forest cover were relatively homogeneous; however, significant geographic heterogeneity existed in the case of vegetation productivity (figure 5). On average, a return of \$1.04 per dollar was observed, in both land degradation and biodiversity. Impacts are observed after a time lag, and access to electricity is positively associated with outcomes.

Measuring the Sustainability of GEF Interventions

According to data from terminal evaluations conducted at project closure, 80 percent of completed GEF projects ►►

BOX 3. OUTCOME SUSTAINABILITY IN BA BE PROTECTED AREA USING REMOTE SENSING

The GEF project ‘Promotion of Sustainable Forest and Land Management in the Vietnam Uplands’ (GEF ID 3627), implemented from 2010 to 2013, supported sustainable forest management in Ba Be National Park. The project focused on local communities and piloted: improved systems for animal husbandry and conservation of sloping lands; bio-energy applications; payment for ecosystem services including participatory approaches to forest protection; and ecotourism. The threats to the forest included fragmentation of habitat, conversion of forest land for infrastructure and agriculture, illegal hunting and overuse for non-timber forest products. At completion, the project was rated “moderately satisfactory” for outcomes and “moderately likely” for sustainability. Figure 6 shows that although there was reported forest loss at the national level and in the buffer areas, remote sensing showed that protected forest cover in Ba Be was stable. Information gathered through interviews indicated that a community focused project design, due attention to income generation activities, and sustained support from provincial governments contributed to positive project outcomes.

► perform satisfactorily in achieving their expected outcomes (IEO, 2018b), but only sixty-two percent of the completed GEF projects were rated in the ‘likely’ range for outcome sustainability at project completion. This finding is comparable with other multilateral development organizations. It ranges from 52 percent in the African Development Bank to 66 percent in Asian Development Bank. Ratings for IFAD, the Inter-American Development Bank, the World Bank and the GEF are at 60 percent, 62 percent, 65 percent and 62 percent respectively. A higher percentage of projects in the countries with large GEF portfolios (Brazil, China, India, Mexico and Russia), and a lower percentage of projects in Least Developed Countries and in fragile states are rated in the ‘likely’ range for

outcome sustainability. The key factors that contribute to higher sustainability include high stakeholder buy-in, political support, availability of financial support for follow up, and sustained efforts from the executing agency. Box 3 provides an example of an analysis of sustainability in Vietnam. The analysis demonstrates the importance of combining geospatial data which provides the long term trend in outcomes, with the need for qualitative field verification to understand the reasons for changes observed.

Similarly, the GEF project in Maiko National Park in DRC has sustained environmental outcomes as there has been no increase in deforestation in the protected area despite the increase in the forest loss in the buffer. ►►

Figure 6: Ba Be: Sustainable Forest Management in Vietnam

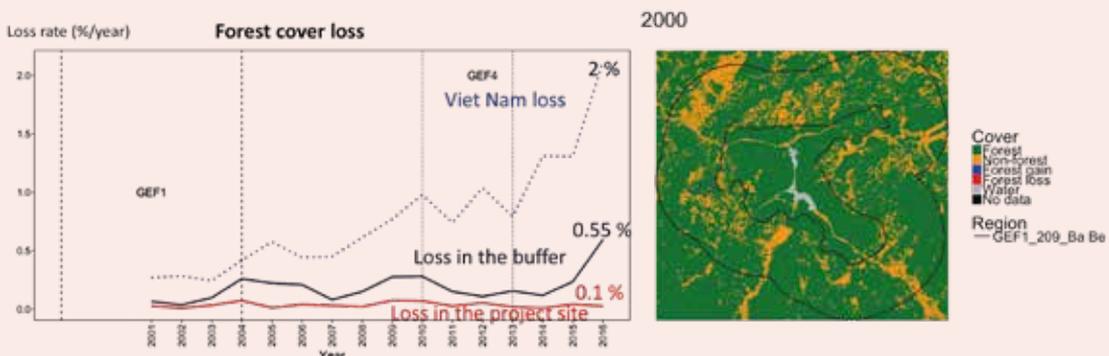
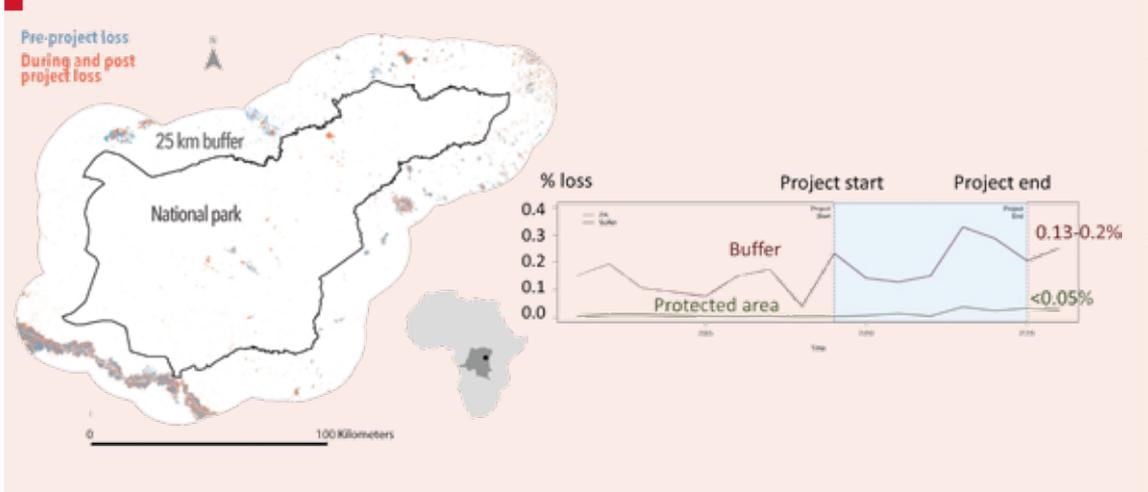


Figure 7: Maiko National Park: Sustainable Forest and Nature Conservation: DR Congo



► In contrast, there is no evidence of sustainable outcomes in the case of GEF interventions in the Cardamom mountains in Cambodia or the San Rafael National Park in Paraguay.

In each of these cases, remote sensing provided additional information that we would not have acquired using traditional evaluation methods, thereby strengthening the evidence base for the evaluation.

Conclusions

The methods and results presented in this article demonstrate recent efforts

by the IEO to adapt and address issues at the forefront of donor agencies during replenishment discussions. In particular, how is the GEF driving transformational change? Did the interventions provide value for money? Are the interventions sustainable in the long term? Addressing these questions has required a rethink of frameworks and tools applied in evaluation, as well as exploring the vast data sets now available to measure outcomes. The results reinforced the importance of using a combination of quantitative and qualitative approaches to evaluate environmental outcomes.



Figure 8: Cardamom Mountains: Integrated Protected Area System, Cambodia

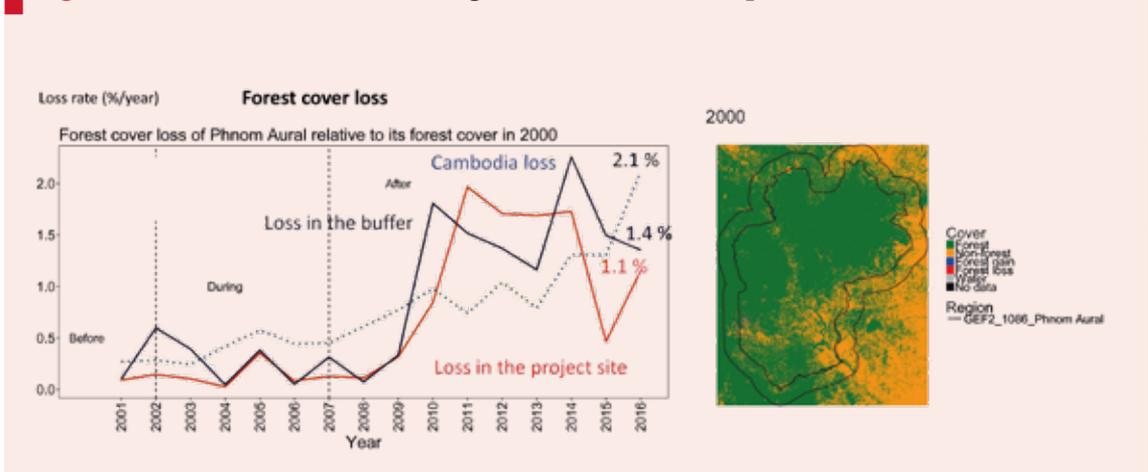
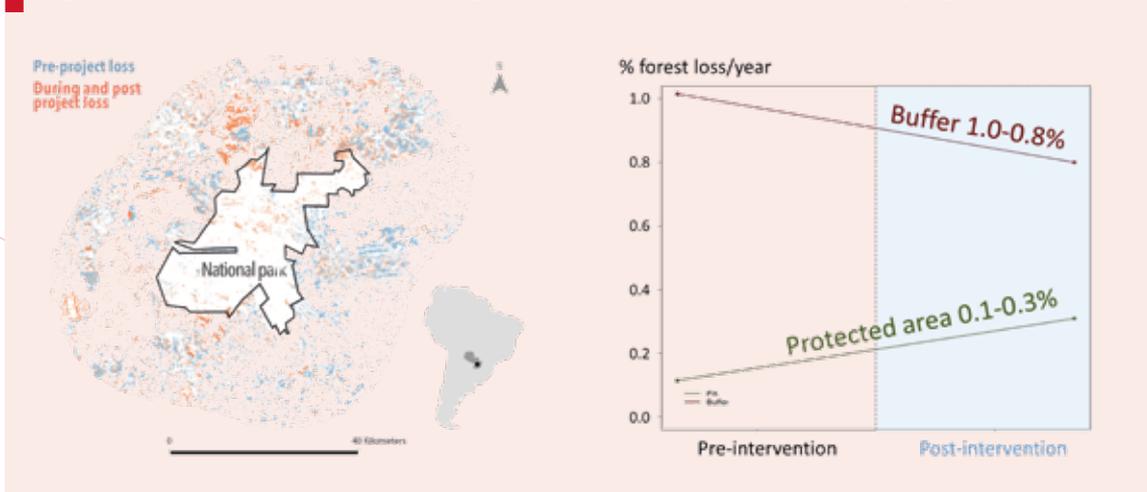


Figure 9: Conservation of Biodiversity, San Rafael National Park East Paraguay

References

- Donor Committee for Enterprise Development (DCED) 2013. Donor Partnerships with Business for Private Sector Development: What can we learn from Experience. DCED Working Paper.
- Gillenwater, Michael. 2012. What is Additionality? Part1: A Long-Standing Program, Greenhouse Gas Management Institute, ResearchGate Publications.
- IFC 2013. IFC's Additionality Primer. https://www.ifc.org/wps/wcm/././Additionality_Primer_Jan_2013_without_case.doc?
- IEO (GEF Independent Evaluation Office). 2016a. International Waters Focal Area Study. <http://www.gefio.org/evaluations/international-waters-iw-focal-area-study-2016>
- IEO. 2016b. A Value for Money Analysis of GEF Interventions in Land Degradation and Biodiversity Projects. IEO Brief. <http://www.gefio.org/sites/default/files/ieo/signposts/files/value-money-ld-2016-brief.pdf>
- IEO. 2016 c. Measuring Environmental Outcomes using Remote Sensing and Geospatial Methods. <http://www.gefio.org/signposts/measuring-environmental-outcomes-using-remote-sensing-and-geospatial-methods-brief>
- IEO. 2017. Evaluation of GEF Support for Transformational Change. <http://www.gefio.org/evaluations/evaluation-gef-support-transformational-change-2017>
- IEO. 2018a. Sixth Comprehensive Evaluation of the GEF. The GEF in the Changing Environmental Finance Landscape. http://www.gefio.org/sites/default/files/ieo/evaluations/files/ops6-report-eng_1.pdf
- IEO. 2018b. Annual Performance Report, 2017. <http://www.gefio.org/evaluations/annual-performance-report-apr-2017>
- IEO. 2018c. An Evaluative Approach to assessing GEF's Additionality. <http://www.gefio.org/council-documents/evaluative-approach-assessing-gef-s-additionality>
- IEO. 2018d. Evaluation of the GEF-UNIDO Global Cleantech Innovation Program. (<http://www.gefio.org/evaluations/evaluation-gef-unido-global-cleantech-innovation-programme-2018>)
- Lech, M., J.I. Uitto, S. Harten, G. Batra and A. Anand. 2018. Improving International Development Evaluation through Geospatial Data and Analysis. International Journal of Geospatial and Environmental Research, Volume 5, Number 2, Article 3. <http://www.gefio.org/sites/default/files/ieo/documents/files/international-development-evaluation-geospatial-data-2018.pdf>
- Valatin, Gregory. 2012. "Additionality and climate change mitigation by the UK forest sector". Forestry, Vol.85, No.4.
- World Bank Group (2016). Supporting Transformational Change for Poverty Reduction and Shared Prosperity – Lessons from the World Bank Experience. Learning Product. Independent Evaluation Group. Washington, DC.

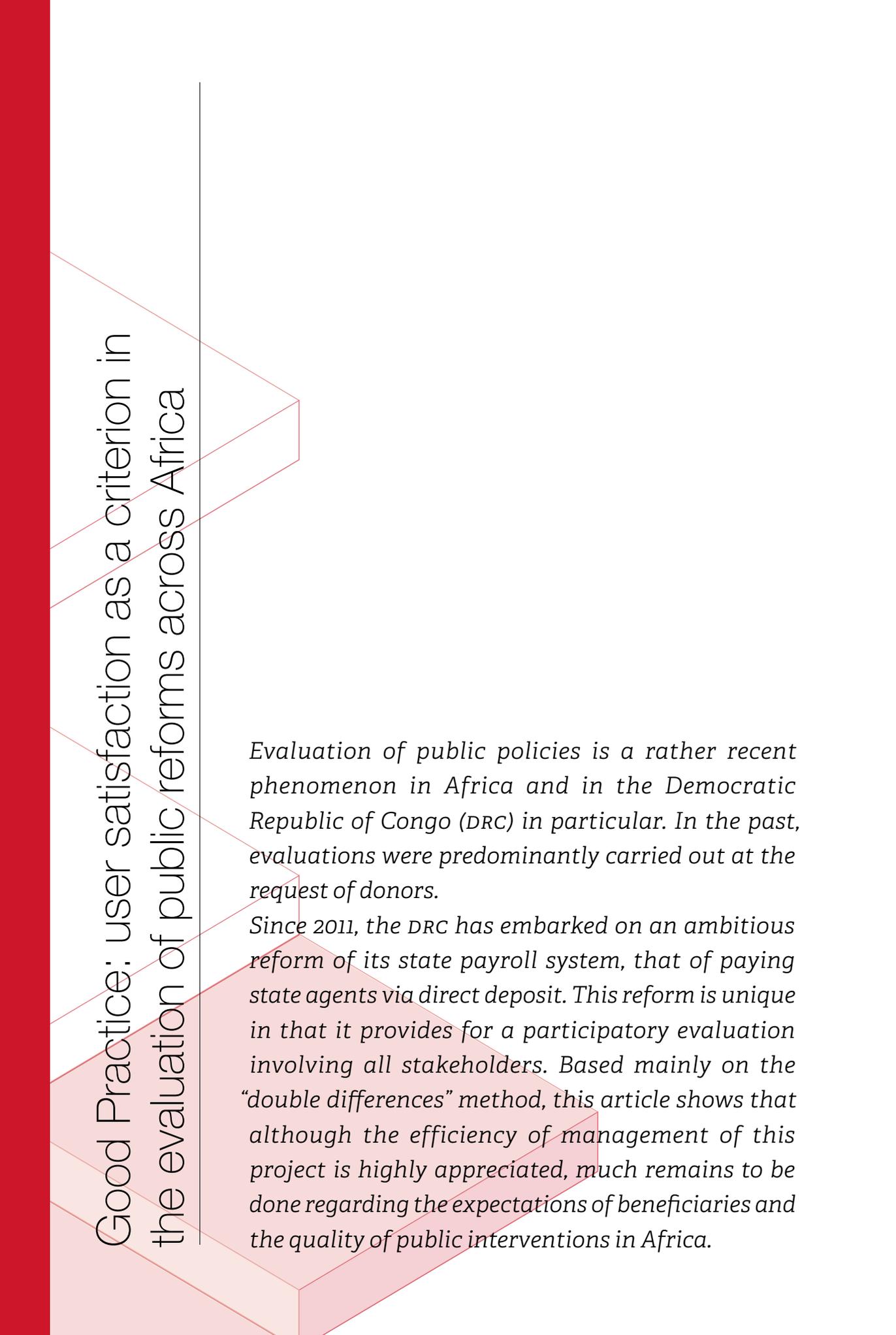
Authors' profile

Dr. Geeta Batra is Chief Evaluation Officer and Deputy Director of the IEO of the GEF, since June 2015. Prior to this position, Geeta was Manager of Country and Corporate Evaluations in the Independent Evaluation Group of the World Bank from 2010-2015. Prior to that, she set up and managed the Results Measurement System for Advisory Services in the International Financial Corporation (IFC), the private sector arm of the World Bank from 2004-2010. She was in the private sector development department of the World Bank from 1998-2004, responsible for investment lending and advisory operations. She has conducted and managed project, program and country level evaluations in a leadership role over the past 16 years. Geeta has a Ph.D. in Economics.



Dr. Juha I. Uitto is Director of the IEO of the GEF. He came to the position after being the Deputy Director of the IEO of UNDP from 2009 to 2014. Prior to that, he worked as an Evaluation Adviser at the United Nations Development Program (UNDP) and as Senior Monitoring and Evaluation Coordinator/Specialist with the GEF since 1999. He has conducted and managed programmatic and thematic evaluations of international cooperation at the global, regional and country levels, in particular related to environmental management and poverty-environment linkages. He served as the Executive Coordinator of the United Nations Evaluation Group (UNEG) from 2009 to 2012 and again in 2014. Dr. Uitto spent the 1990s with the United Nations University (UNU) coordinating the university's environment and sustainable development research and training programs. He was educated at the Universities of Helsinki and Lund, and holds a PhD in Social and Economic Geography. He has authored/edited several books and published more than 40 peer reviewed articles and book chapters on topics related to the environment, natural resources management, environmental hazards, and evaluation.





Good Practice: user satisfaction as a criterion in the evaluation of public reforms across Africa

Evaluation of public policies is a rather recent phenomenon in Africa and in the Democratic Republic of Congo (DRC) in particular. In the past, evaluations were predominantly carried out at the request of donors.

Since 2011, the DRC has embarked on an ambitious reform of its state payroll system, that of paying state agents via direct deposit. This reform is unique in that it provides for a participatory evaluation involving all stakeholders. Based mainly on the “double differences” method, this article shows that although the efficiency of management of this project is highly appreciated, much remains to be done regarding the expectations of beneficiaries and the quality of public interventions in Africa.

Bobo B. Kabungu, Central Bank of the Democratic Republic of Congo

Introduction

In order to become an emerging economy, Africa needs major reforms of its institutions and public policies. Indeed, the World Bank evaluation of policies and institutions in Africa for the year 2017 reveals that African countries on average score "slightly below the average of other countries receiving aid from the International Development Association"¹. (World Bank 2018)

Although the DRC is not top of the class in terms of development, since the beginning of the millennium, it has been committed to catching up, in order to meet the expectations of the Agenda 2030. It is in this context that a direct deposit system of pay for State employees and civil servants was launched in 2011. This project, which is part of other institutional reforms aimed at strengthening the efficiency of the State, aims at the non-accumulation of wage arrears and implies a harmonization to the system of the payment for employees and officials of the State.

Expected benefits of this reform, grouped according to targeted stakeholders, are as follows:

- **For the Government:** (i) control over the size of the workforce and (ii) of the wage bill;
- **For State Employees and officials,** the guarantee of: (i) regular payment of wages, (ii) receipt of the full salary and (iii) improved financial inclusion.

To this end, it was first necessary to carry out physical checks upstream (concerned administrations) and downstream (paying banks), and to ensure the visibility of operations in order to facilitate real time monitoring via an automated payroll system. In addition, there was need to significantly reduce the number of intermediaries and to promote the granting of credit with the domiciliation of the salary at the banking institution as its only collateral.

This reform relies on the rigor of banking institutions to control the size of the workforce and, by extension, the wage bill. Currently, payment of civil servants' wages and salaries is the largest public expenditure in the DRC, with more than one million employees spread over a geographical area 80 times larger than Belgium and with barely 15 active banks in major urban centres.

This article is a synthesis of a project completed thanks to financial support from the Congolese Initiative for Evaluation, Research, Welfare and Gender (ICEBERG asbl) and presented during an evaluation workshop held in Kinshasa from February 6 to 8, 2019. It attempts to present how the satisfaction of the beneficiaries of the direct-deposit payment system in the DRC was measured using the double-differences method. This way, the evaluation is able to reach normative conclusions through an assessment of achievements starting with empirical data and explicit criteria followed by supporting analysis. Moreover, this work is a pioneering exercise if we consider the fact that the approach essentially focused on the technical management of the reform, which emerges from three ➤

- ▶ earlier evaluation exercises carried out respectively in 2015, 2016 and 2017.

Brief presentation of the initial evaluations of the reform

Between 2011 and 2018, three evaluations were performed to assess the use of direct deposit for the payment of state employees and civil servants. The first, conducted in March 2015, was limited to recording the increase in bank enrolment of employees (73.9% at the end of December 2014) as a result of the reform and to identify challenges to address.

A second evaluation in April 2016 analysed the impact of the reform, and found a positive effect on the living conditions of beneficiaries, as well as the real need for the harmonization of the pay statistics of state employees and officials.

In June 2017, a third exercise to assess the management of the reform was carried out with input from all stakeholders. It revealed that the final outcomes of the reform deviated by about 10.0%, give or take, from the expected outcomes.

It is important to note that these evaluations failed to put the beneficiary of the intervention, or the user of the service, at the heart of a quality assessment approach. Indeed, the evaluation of public interventions also requires taking into consideration the expectations of the beneficiaries of an offered service, thereby capturing the satisfaction of users in addition to policy performance.

Conceptual Framework

User Satisfaction

In their research paper, Beau, Blanche, Garcin and Morice define public quality

as "the ability to meet the needs of users, whether expressed or implied needs" (Beau et al., 2005: 11). In essence, everything must be done to ensure that public services respond to end-user needs.

In the context of a public economy that advocates for the general interest and well-being of many, the evaluation of State public policies must integrate a criterion of "user satisfaction" with offered services in order to measure overall impact.

Evaluation of public interventions: purpose and approaches

Evaluation is strongly linked to the rationalisation of public action and aims to determine the extent to which a policy has achieved its set objectives in addition to expected impacts on the public concerned. It consists of an analysis of the results of a public intervention and a judgment of its value on the basis of its effective capacity to satisfy a need.

"In the end, civil society and donors are the stakeholders likely to push public authorities to construct a quality approach to public policies as determining criteria for good governance."

As far as methodology is concerned, a vast literature (Bertrand, Duflo and Mullainathan, 2004, De Vreyer, 2019, De Vreyer, 2019 b, De Vreyer, 2019, Gertler, Martinez, Premand, Rawlings and Vermeersch, 2016, Klerman, Olsho and Barlett, 2015) outlines several methods: random evaluation, simple difference, difference in difference (or double differences), multiple variable regression, statistical matching, regression on discontinuity, and so on. ▶



► The evaluator makes his/her choice according to the options available to him/her as well as the objective pursued; however, Kabungu (2018) mentions that where data and circumstances permit, two scenarios can be used: "before and after" and "with or without". The first technique is often complicated by the absence of baseline data, despite the contribution of specific techniques (memory, community mapping, etc.) in terms of identifying what the situation was prior to a given intervention. The second so-called counterfactual technique is to define a situation that would have prevailed in the absence of intervention.

Methodology

Introduction to the survey

The most common way to evaluate the quality of an intervention is to interview the beneficiaries, thus we conducted a survey of 102 State employees and civil servants. The resulting inputs allowed us to reach firstly, an appreciation of the quality of the reform through the construction of a composite index and secondly, a determination of the reform's own effects using the double-difference method.

To do this, we chose a stratified sample with a random draw. We based this on the main characteristics of the target population that was then broken down into subgroups

or strata according to their gender, age, background, category, etc. Subsequently, we randomly drew individuals from each stratum, ensuring its proportionality in the initial set in order to preserve the representativeness of the sample.

On the basis of a total population of 1,300,000 State employees and officials, a sample size calculator (available online at <https://fr.checkmarket.com>), advised to interview 385 people (assuming a response rate of 100.0%, a margin of error of 5.0% and a confidence level of 95.0%). After sorting the coded questionnaires, we selected 102, fifty-one being users of the public intervention and 51 non-users.

We were thus able to obtain 204 observations based on a "before and after" of the reform, which we analysed using SPSS software. Information on before the reform was completed on the same questionnaire, meaning that interviewees based their answers on memory in order to recall their situation prior to the project being evaluated.

Presentation of the double differences method

In the context of policy evaluation, this quasi-experimental, counterfactual method is useful in assessing the effect of an intervention by comparing the gap or difference between the recipient or treatment group and the non-treatment group (control group) before and after (or since) the (new) program. ►

Table 1: The doubles differences method in practice.

	Treatment group	Control Group	Observations
Before	$Y_{11} T$	$Y_{11} C$	
Since/After	$Y_{00} T$	$Y_{00} C$	
Observations	Comparison of the average of the outcome variable before and since / after $Y_{11} T - Y_{00} T$	Comparison of the average of the outcome variable before and since / after $Y_{11} C - Y_{00} C$	Subtracting the two averages ($Y_{11} T - Y_{00} T$) - ($Y_{11} C - Y_{00} C$)

Source : Vermeersch (2008, p. 5).

► Evaluation results

The results are presented in three stages: (i) the treatment group; (ii) the control group and (iii) the difference in differences. Tables 2 and 3 define and present the sub-indexes used and the encoding elements of the answers to the questions.

The results indicate that, during the reform period, the overall situation of the average beneficiary of the reform improved by 0.19 points, with the user satisfaction index having risen from 0.26 before to 0.44 after its implementation. In contrast, this index rose on average from 0.23 to 0.28 for the non-beneficiary, an increase of 0.0447 (rounded to 0.04). The difference between the two differences gives us a net effect of 0.14 that we credit to the reform. In other words, without the direct deposit of pay for State employees and civil servants, the expectations of the group of beneficiaries would not have obtained the same answer.

However, it should be noted that while the average for all sub-indexes shows a clear improvement (with difference scores of 0.55, 0.49 and 0.47 respectively for the

collection of full salary, timeliness and access to credit), the weak implantation by banks lessened opinion with regard to proximity, which depreciated by 0.25. Hence the need to intensify efforts (including incentives) to encourage banks to open branches in regions with difficult access. Another way would be to restructure the Caisse d' Epargne du Congo (Congo Savings Bank) along the example of the Postal Service.

Conclusion

A first observation is that, whatever the method used, evaluations of the reform are more than encouraging. Following the Public Expenditure and Financial Accountability (PEFA) model, the reform obtains a grade B (satisfactory) for its outputs defined after reconstitution of the logical framework. Its overall performance according to an assessment by the Independent Office of Evaluation of the International Fund for Agricultural Development IFAD (IOE) is 5 on a scale of 1 to 6, ranging from total dissatisfaction (1) to complete satisfaction (6). These ►►

Table 2: Presentation of sub-indexes

Sub-index	Definition
Optimum date	Conforming to payment calendar.
Proximity	Distance of the pay site from State employees and civil servants.
Timeliness	Timing of payment to beneficiary.
Access to credit	Capacity of employees and civil servants to obtain credit by the intervening financial institution.
Security	Beneficiary confidence in bank security and access to funds.
Rapidity	Speed of payments.
Confidence	User confidence in receiving payment.
Culture of savings	Propensity to save a portion of payment within the financial institution.
Access to information	Ability to be informed as to receipt of payment.
Perception of improvement	Extent of improvement.

Source : Author

Table 3: Coding of responses

Designation	Code
Age	Young (≥ 18 yrs < 50 yrs) = 1, Older (> 50 yrs) = 0
Sex	Male = 1, Female = 0
Location	With FI = 1, Without FI = 0
Sector	Monthly update = 1, without = 0
Revenue	More than the equivalent of USD 100 = 1, Less than = 0
Optimum date	from 15th to 25th of month M = 3, 26th to 30th M = 2, 1st to 10th M+1 = 1, after the 10th M+1 = 0
Proximity	Close (≤ 5 kms) = 3, nearby (> 5 kms ≤ 10 kms) = 2, some distance (> 10 kms ≤ 15 kms) = 1, long distance (> 15 kms) = 0
Timeliness	Punctual = 3, somewhat punctual = 2 or 1, non-existent = 0
Access to credit	Ease of obtaining a medium-term loan on the basis of the domiciliation of salary = 3, possibility of borrowing in the short term on the basis of the domiciliation of salary = 2, possibility of access to the credit for an additional guarantee = 1, no access to credit = 0
Security	Yes = 1, no = 0
Rapidity	Yes = 1, no = 0
Confidence	Yes = 1, no = 0
Culture of savings	Yes = 1, no = 0
Access to information	Yes = 1, no = 0
Perception of improvement	Yes = 1, no = 0

Source : Author

Table 4: Treatment group (beneficiaries) and control group (non-beneficiaries) opinions and determination of the difference in differences noted Diff

		on the due date	Proximity	Entirety	Access to credit	Security	Rapidity	Confidence	Culture of savings	Access to information	Perception of improvement	Isaqual
		3	3	3	3	1	1	1	1	1	1	1
beneficiary β	Before	0,87	1,43	0,98	0,37	0,43	0,27	0,14	0,08	0,08	0,08	0,26
	After	1,51	1,08	1,73	1,06	0,82	0,76	0,37	0,22	0,22	0,22	0,44
	Diff β	0,73	-0,35	0,75	0,69	0,39	0,49	0,24	0,14	0,14	0,14	0,19
non beneficiary η	Before	0,73	1,49	1,02	0,35	0,25	0,12	0,12	0,00	0,00	0,08	0,23
	After	0,96	1,39	1,22	0,57	0,31	0,22	0,18	0,00	0,00	0,12	0,28
	Diff η	0,24	-0,10	0,20	0,22	0,06	0,10	0,06	0,00	0,00	0,04	0,04
Diff δ		0,49	-0,25	0,55	0,47	0,33	0,39	0,18	0,14	0,14	0,10	0,14

Source : Author

► two dimensions come together with scores that, expressed as a percentage, reach 90.0% of achievements compared to the PEFA forecasts and 83.3% if we refer to the IOE.

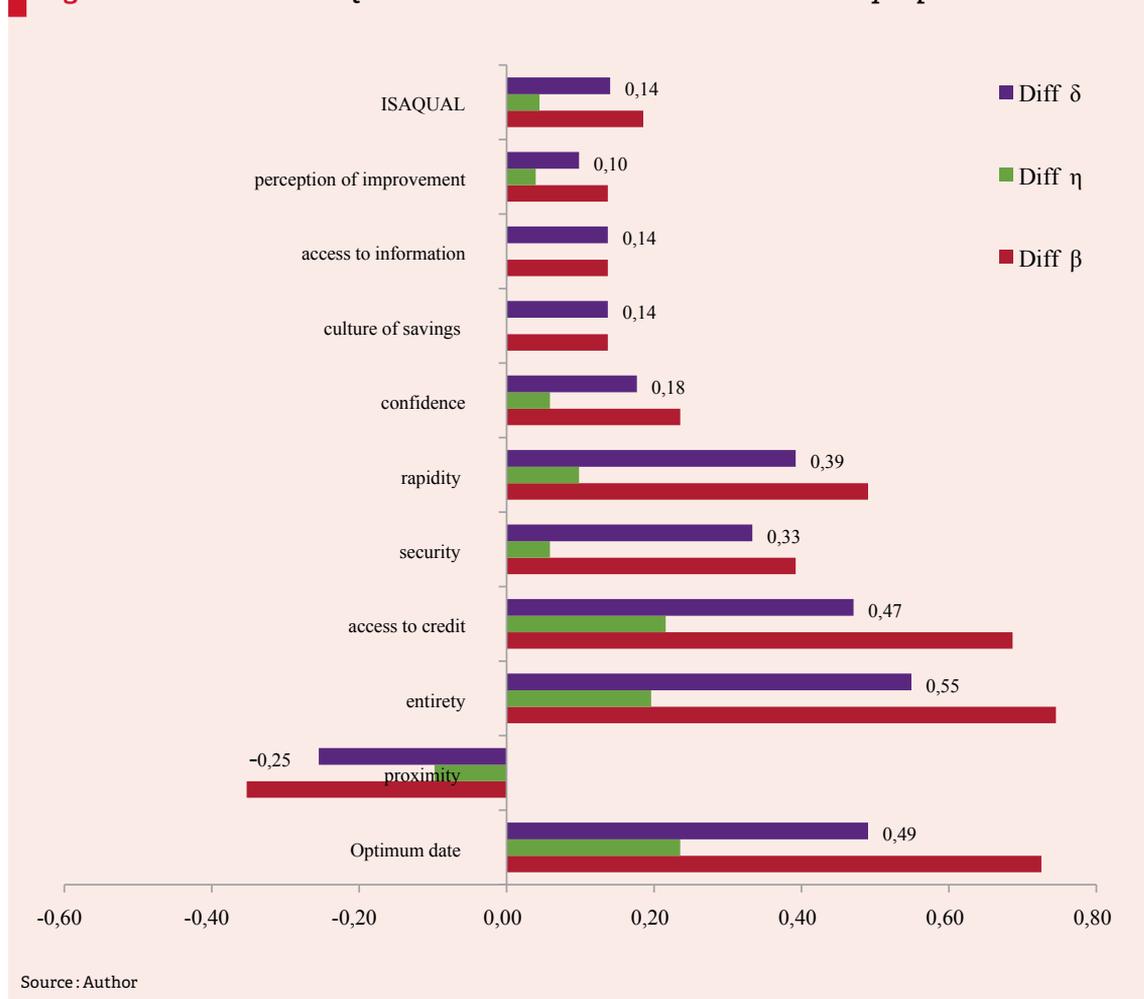
The last dimension is just as comfortable, as the overall assessment of the quality of payroll reform for State employees and civil servants (ISAQUAL) shows an improvement between before and after the intervention for both beneficiaries (0.19) and non-beneficiaries (0.04), thus reflecting, by the double-difference method, an index score of 0.14³. However, although positive, taking into account the expectations of beneficiaries represents only 14.0%, which still represents an important challenge

regarding the quality approach of the reform.

These results confirm those of the DRC Government's Capability Scan (CAP-Scan) in 2011 in line with monitoring and evaluation activities, according to which on one hand, all the sectors appropriate a success and reject a failure among them, and on the other, "the real needs of users of public services have difficulty to be seen as a priority and the responses may not be in line with the expectations of the beneficiaries". (Kabungu, 2018b, p.18). In the end, civil society and donors are the stakeholders likely to push public authorities to raise the quality approach of public policies as a determining criterion for good governance.



Figure 1: Difference in ISAQUAL² differences and doubles differences by expectation



Source : Author



© General Adolphe TANGAMO, Director of Human Resources at the General Secretariat for Defense, presents on payroll management within the National Defense Department, during the Kinshasa evaluation workshop on February 6-8, 2019. (Mr. Jean-Louis KAYEMBE wa KAYEMBE, Chairman of the Payment Monitoring Committee, is seated in the first row, 2nd from the right).



© Mr. Bobo B. KABUNGU, Coordinator of the Technical Secretariat of the Payroll Monitoring Committee, reporting on the results of a “mini-survey” of beneficiaries’ perceptions. Kinshasa, DRC February 6 – 8, 2019.

References

Beau, J., Blanche, F., Garcin, M. and Morice F. (2005). Quality approach and public satisfaction. Research paper. Diploma of Conservator of Library, Villeurbanne: ENSSIB.

Bertrand, M., Dufló, E. & Mullainathan, S. (2004). How Much Should We Trust Difference in Differences Estimates? *Quarterly Journal of Economics*. 119(1), 249–275.

De Vreyer, P. (2019). The double difference method. Introduction to impact evaluation methods. Paris: Université Paris Dauphine.

De Vreyer, P. (2019 b). The regression on discontinuity. Introduction to impact assessment methods. Paris: Université Paris Dauphine.

De Vreyer, P. (2019 c). The regression on discontinuity. Introduction to impact assessment methods. Paris: Université Paris Dauphine.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. & Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice*. 2nd edition. Washington DC: World Bank.

Independent Office of Evaluation [IOE] of IFAD (2016). *Evaluation Manual*, IFAD, Rome.

Kabungu, B. B. (2012). *Direct Deposit of pay for State employees and civil servants, Frequently Asked Questions*. Kinshasa: CSP.

Kabungu, B. B. (2018). *Governance, the evaluation of public policies and quality in the DRC, Case of the Direct Deposit of pay of state staff*. Master's thesis 2. Grenoble: UGA.

Kabungu, B. B. (2018 b). *Evaluative practice in the French and Congolese public sphere: comparative analysis and perspectives of institutionalization*. *Annales de l'UNIGOM*. (8)2, 3-25.

Klerman, J.A., Olsho, L., Barlett, S. (2015). *Regression Discontinuity in Prospective Evaluations: The Case of the FFVP Evaluation*. *American Journal of Evaluation*. 36(3).

PEFA (2016). *Public Financial Management Assessment Framework*, Washington: Office of the PEFA.

PRRAP. (2016). *Analytical study of the impact of Direct Deposit of pay for employees and civil servants of the State in the DRC*. Full report. Kinshasa: Ministry of Public Service – DRC.

Vermeersch, C. (2008). *Estimation of double differences (diff-in-diff) and panel data*. World Bank.

World Bank. (2018). *2017 Evaluation of Policies and Institutions in Africa (CPIA): A situation that is globally identical to that of 2016*. Press Release No. 2019/007 / AFR, [online], <https://www.worldbank.org/en/news/press-release/2018/09/12/world-bank-review-reveals-unchanged-quality-of-policies-and-institutional-performance-in-africa> accessed February 13, 2019.

Endnotes

1. This Country Policy and Institutional Assessment (CPIA) is an annual diagnostic tool used to measure the quality of public policy and institutional structures to support sustainable pro-poor growth.
2. ISAQUAL Quality of payroll reform for State employees and civil servants
3. Note: this is not an error but the result of rounding to two decimal places.

Author's profile

Bobo B. Kabungu is Assistant Director at the Central Bank of Congo and Coordinator of the Technical Secretariat of the Payroll Monitoring Committee in the DRC. Economist by profession, the author is a doctoral student in public administration and holds several masters including one in Evaluation and Management of Social Policies obtained at the University Grenoble Alpes. In the voluntary sector, he is a key promoter of the Congolese Initiative for Evaluation, Welfare, Research and Gender (ICEBERG asbl) since his participation in the International Program of Training in Development Evaluation (PIFED) of the National School of Public Administration (ENAP) in Quebec in 2016. He is also a researcher at the Centre for Research in Human Sciences (CRESH-DRC).



The Most Significant Change Approach in Evaluation: A quick guide and its application at the African Development Bank

On 7 June 2019, as part of its support to the gLOCAL Evaluation Week organized by the Centers for Learning on Evaluation and Results (see “News in Pictures” section), IDEV hosted a webinar on ‘Most Significant Change (MSC) in Evaluation: The experience from the application of MSC in the African Development Bank.’ The [webinar](#) was delivered by Madhusoodhanan (Madhu) Mampuzhasseril, Chief Evaluation Officer at IDEV, and drew on the seminal work by Dr. Rick Davies on the subject as well as the Evaluation of the African Development Bank’s Decentralization Strategy and Process (Davies 2005 and 2009; AfDB 2009). Madhu’s presentation gave an overview of the MSC approach and described how the Bank has used this approach in the past. The following is a summary of the key points discussed during the webinar and the question-and-answer session following the presentation.

Madhusoodhanan (Madhu) Mampuzhasseril, IDEV, African Development Bank

Description of the Approach

Msc is a systematic, transparent and participatory method of collecting and analyzing qualitative information about change. This information is in the form of stories of change, obtained through individual interviews, which are then subject to selection by different stakeholder groups (Davies R. and Jess Dart (2005). The MSC technique focusses on the collection and systematic analysis of significant changes caused by an intervention. It is worth noting that MSC is not intended to be a stand-alone technique for monitoring and evaluation. MSC combines well with other evaluation methods such as short surveys and focus group discussions.

Steps in applying MSC

Step 1: Start by raising interest

- Explain the evaluation methodology to selected individuals or groups.
- Start small. Begin as a pilot to see what works and what does not.
- Identify key “champions” who will motivate people, explain the technique, facilitate the collection of stories, ensure feedback, and

maintain confidentiality where/when necessary.

Step 2: Define the domains of change

- Domains are broad categories of Significant Change (sc) stories.
- Domains of change are not indicators. Allows different interpretations of what constitutes change in the area.
- Ideally 3 to 5 domains are manageable.
- Use predetermined domains as well as open window domain.
- Can have negative change domains.
- Not necessary to be predetermined.

Step 3: Define the reporting period

- Monitoring: Frequency as per the reporting requirements.
- Evaluation: Usually a one-off exercise

Step 4: Collect significant change stories

- Open question: “During this abc period, what do you think was the most significant change in the xyz domain of change?”
- Interviews and note taking ▶

MSC CAN BE USED FOR PROGRAM EVALUATION...	AND FOR ORGANIZATIONAL LEARNING
<ul style="list-style-type: none"> ■ MSC can render judgments, facilitate improvements and/or generate knowledge. ■ MSC can be built into a summative evaluation as a preceding activity. ■ Can be used to identify and aggregate large scale stakeholder views. ■ Can be combined with a theory based evaluation. ■ Rank stories rather than selecting only the MS. 	<ul style="list-style-type: none"> ■ Can influence the values of stakeholders including staff within the organization. ■ Dialogue built into the process facilitates learning.

- ▶ ■ Through group discussions
- Beneficiary writes the story
- Use of videos
- Each story should specify: a) who collected; b) description; and c) significance

Step 5: Select MSC stories

- Iterative process of selecting and then pooling stories.
- Can use the hierarchy of existing organization.
- Decide: number of levels, how many processes, how many stories, who should participate.
- The selection process should be open debate. Read in group, discuss, select most significant, show reason.
- Majority vote, iterative voting, rating, secret ballot can be used.
- Participants: Beneficiaries, field staff, line management and people with advisory capacities can be involved in selection.
- Document the selection process.
- Filtered out stories have value at certain levels though they are not the most significant.

Step 6: Feed back the results of the selection process

- Feedback given to the providers of the story. This completes the feedback loop

and creates ongoing dialogue on what is significant change.

- Use different media to provide the feedback. The Bank used a specially designed blog.

Step 7: Verify the stories

- Needed to ensure the change actually happened. There could be fictions, exaggerations and misunderstandings.
- Verify only stories selected as most significant. Not randomly selected ones.

Step 8: Quantify whenever possible

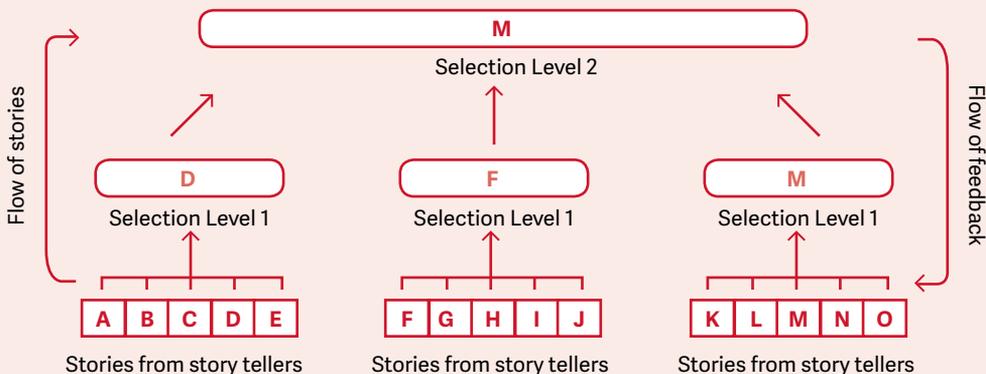
MSC emphasis is on qualitative reporting but quantification is possible.

- Within individual stories: no. of people, activities...
- During feedback stage: specific quantitative data on specific phenomena observed
- Analysis of full set of stories, frequency of occurrence

Step 9: Perform secondary analysis and meta-monitoring

- Examine, classify and analyze content of all stories.
- Keep record of all stories and the process in a retrievable format.
- Meta-monitoring: a) number of stories reported and trend; b) who is writing and who is not; c) whose stories are selected and whose not; d) recommendations acted upon. ▶▶

Figure 1: Selecting MSC stories - Flow of stories and feedback in MSC



Source: Guide to MSC



- ▶ ■ Secondary analysis can contribute to summative evaluation. Contributing to the analysis of outcomes.

Step 10: Revise the system to fit your evaluation needs

- Almost all organizations that have used MSC made changes in the content and the way in which MSC was implemented.
- Names of domains change
- Frequency of reporting
- Types of participants
- Structure of meetings

When and when not to use it?

Useful to evaluate

- Complex programs that produce diverse and emergent outcomes
- Large programs with several organizational layers
- Programs focused on social change
- Programs whose evaluation is participatory in ethos and design
- When the evaluation focuses on learning more than accountability

Less useful for

- Evaluations that aim to capture expected change
- Retrospective evaluations of programs that are completed
- Quick and low-cost evaluations

Key Enablers of MSC

- Organizational cultures that accept failures
- Champions with good facilitation skills
- Having the time to run several cycles of the approach
- Commitment by senior management

The MSC approach in practice

In 2008 the MSC approach was used as part of a formative evaluation of the

decentralization strategy and process of the African Development Bank. The evaluation aimed to: a) identify unexpected changes; b) provide a more in-depth description of the events monitored; c) identify areas of agreement and disagreement about the expected outcomes and process; d) promote a dialogue between the Bank headquarters and Field Offices about the decentralization process; and e) identify, through this process, innovations, risks to performance and strategic choices to approach decentralization.

The evaluation used a mixed-methods approach to gather both qualitative and quantitative evaluative evidence. It is in this context that the MSC technique was used to collect qualitative information about perceptions of changes- as a result of the decentralization strategy- in the following areas:

1. Changes in AfDB's country dialogue: with government, donors, private sector, and civil society
2. Changes in the management of the AfDB country portfolios
3. Changes in the relationship between AfDB country office and AfDB headquarters

Sixty-three (63) Significant Change stories were collected through 21 interviews conducted in four country offices (Burkina Faso, Nigeria, Sierra Leone and Tanzania), and analyzed to arrive at the most significant changes due to the decentralization process. Through the selected stories, the following most significant changes were highlighted:

Changes in the Bank's country dialogue

- AfDB mobilizes donors to address the government's high priority needs

Changes in the management of the Bank country portfolios

- Disbursement speeded up through better document management ▶▶

- ▶ **Proactive portfolio review reduces processing time by 1-3 months**

Changes in the relationship between the Bank's country office and AfDB headquarters

- Better communications means a more responsive FO
- Great trust in the FO is increasing the efficiency of Bank operations
- Locating a task manager in the FO speeds up implementation

The results of the entire MSC exercise are documented on a blog Decentralising AfDB¹ accessible to all AfDB staff. Video summaries are also provided, along with summaries of the discussions of the stories by the FO teams and four stakeholder groups at AfDB Headquarters. The exercise was instrumental in helping inform the mid-term review of the 11th African Development Fund (ADF) in 2009, as well as the ADF replenishment discussions conducted in 2010.

Challenges of using MSC to evaluate the AfDB decentralization strategy

One of the main challenges reported with using this approach was the ability to secure the availability of respondents and program staff to attend the interviews and other related activities to select the most significant stories. It required good planning and coordination from the evaluation team to keep and abide by a tight schedule while on site. The approach was also considered relatively costly (about 40% of the evaluation budget) by some stakeholders.

Concluding observations

To recap, the Most Significant Change (MSC) technique is a tool most suited for collecting, discussing and selecting stories about the significant changes that people experience as a result of a particular program or intervention. It involves people at ▶



► different levels of an organisation discussing the stories and then selecting the stories they consider most significant. This process aims to promote ongoing dialogue and learning about programs and/or interventions and how they can be improved to better meet their aims. It also useful in enabling discussion about the unexpected or negative changes that may have happened as a result of a program and/or intervention.

In our case study, the MSC approach was well received by AfDB stakeholders. It generated greater awareness amongst stakeholders about the complexity and benefits of the Bank's decentralization process, leading to the decision at the Senior Management level to accelerate

"The Most Significant Change (MSC) technique is a tool most suited for collecting, discussing and selecting stories about the significant changes that people experience as a result of a particular program or intervention".

the process. From the evaluation team's perspective, the MSC approach contributed to the formative evaluation of the Bank decentralization process by supplementing other evaluative information, and it served its purpose well.

eVALU

References

AfDB. (2009). Independent Evaluation of the Decentralization Strategy and Process at the African Development Bank. Tunis: African Development Bank. Available at: <http://idev.afdb.org/sites/default/files/documents/files/Independent%20Evaluation%20of%20the%20Decentralisation%20Strategy%20and%20Process%20at%20the%20African%20Development%20Bank.pdf>

Davies R. and Jess Dart. (2005). The Most Significant Change (MSC) Technique: A Guide to its Use. Available at: <https://mande.co.uk/special-issues/most-significant-change-msc/>

Davies, R. et al. (2009). Evaluating Decentralisation in the AfDB: The use of Most Significant Change stories. Tunis: African Development Bank.

June Lennie. (2011). Equal Access Participatory Monitoring and Evaluation toolkit. Available at : https://www.betterevaluation.org/sites/default/files/EA_PM%26E_toolkit_MSC_manual_for_publication.pdf

Endnote

1. Decentralizing AfDB blog is managed by Dr. Rick Davies. It is accessible by AfDB Bank staff at: <http://decentralisingadblog.wordpress.com>

Author's profile

Madhusoodhanan (Madhu) Mampuzhasseril is Chief Evaluation Officer at IDEV, the division responsible for Agriculture, Human and Social Development, Country and Corporate Evaluations.



Prior to joining the African Development Bank in 2009, Madhu worked with IDEV as a long-term consultant, involved in sector-level evaluations of health, education, and agriculture and rural development, and in the corporate evaluations of the Bank's decentralization strategy and process, and the quality-at-entry of ADF operations and strategies 2005-2008. Other corporate evaluations that he has completed in recent years include those of the Bank's operational procurement policy and practice, administrative budget management, human resource management, and the Bank's quality assurance process across the project cycle. He has also managed the Country Strategy and Program Evaluations for Ghana and Zambia.

IDEV Annual Report 2018 summarizes a year of delivering knowledge, engaging with stakeholders, and responding to evaluation needs and demand to improve the lives of the people in Africa

On 21 May 2019, IDEV presented its 2018 Annual Report to the AfDB's Committee on Development Effectiveness. Covering the year in which evaluation and evaluators worldwide were called upon to "speak truth to power", the report entitled Delivering, Engaging and Responding focuses on how IDEV delivered evaluative knowledge; engaged with Bank operations staff, stakeholders and partners to strengthen its products and build a culture of evaluation; and responded to evaluation needs, demand and changes.

The report also highlights how IDEV continued to improve the evaluation function at the AfDB, including completing both a self-assessment and an Independent Peer Review, in order to draw lessons and further enhance the quality and impact of its work.

IDEV's noteworthy achievements in 2018 include:

- [15 evaluation products](#) completed and presented to the AfDB Board and/or Management, including its first ever mid-term evaluation.
- The institutionalization of capitalization workshops at AfDB to discuss (with

operations staff) findings, lessons and recommendations from evaluations and foster improved project designs and strategies.

- The organization of AfDB Development [Evaluation Week 2018](#), which attracted over 450 participants.
- Preparing a discussion on the [Strategic Directions](#) of the Evaluation Function of the AfDB.

● <http://idev.afdb.org/en/news/delivering-knowledge-engaging-stakeholders-and-responding-evaluation-needs-and-demand-improve>



IDEV contributes to gLOCAL Evaluation Week 2019

Organized for the first time in 2019 by the Centers for Learning on Accountability and Results (CLEAR), gLOCAL Evaluation Week is a global platform for evaluation knowledge-sharing and networking events. The week features both physical and online events organized by public, private, and academic institutions and organizations that produce, use, or promote evaluations to strengthen development programs. As part of its contribution to gLOCAL Evaluation Week which took place from 3 to 7 June 2019, IDEV facilitated three events – one workshop and two webinars. In total more than 270 events across the five continents (in 37 countries) were hosted for the benefit of the evaluation community.

Events hosted by IDEV during gLocal Evaluation Week 2019:

1. A workshop on 'Optimizing the AfDB's Program-Based Operations Support as a Package'
2. A webinar on 'Promoting Evaluative Evidence Use: Opportunities and Constraints'
3. A webinar on 'Most Significant Change in Evaluation: The experience from the application of MSC in the African Development Bank'

<http://idev.afdb.org/en/news/idev-contributes-glocal-evaluation-week-2019>



IDEV Acting Evaluator General attends AfDB Annual Meetings 2019

The 2019 Annual Meetings of the Boards of Governors of the African Development Bank Group were held in Malabo, Republic of Equatorial Guinea from 11 to 14 June 2019 under the theme of “Regional Integration for Africa’s Economic Prosperity”.

Ahead of this strategic gathering of leaders and policy makers, IDEV acting Evaluator General, Karen Rot-Munstermann shared with the meeting participants key lessons learned and recommendation from four IDEV evaluations on regional integration.

At the Annual Meeting itself, she presented the results of the [independent evaluation of the Bank Group’s Development and Business Delivery Model](http://idev.afdb.org/en/news/addressing-regional-integration-challenges-africa-learning-past), which had been requested by the Boards of Governors at their previous Annual Meeting, to the Governors’ Consultative Committee which is considering a General Capital Increase for the Bank.

🌐 <http://idev.afdb.org/en/news/addressing-regional-integration-challenges-africa-learning-past>



Evaluation Cooperation Group (ECG) meets to discuss and share experiences on current evaluation issues

The Evaluation Cooperation Group (ECG) 2019 spring meeting took place in Thessaloniki, Greece from 24–25 June 2019, hosted by the Black Sea Trade & Development Bank.

The event brought together about 25 participants from international financial institutions with an independent evaluation function. The meeting was organized around a number of thematic sessions, including: results frameworks for the evaluation function; the role of impact evaluations; processes and key elements

of bank-wide evaluation policies; the revision of the OECD-DAC evaluation criteria; and evaluations of PPPs and the mobilization of private sector finance in the development context. Members heard presentations, exchanged views and shared their experiences on each of these topics.

At the meeting, Karen Rot-Münstermann, the Bank acting Evaluator General, shared with the ECG the highlights of IDEV's recently completed evaluation of the AfDB's utilization of the Public Private Partnership Mechanism (2006–2016) during the session on PPPs and private sector finance.

<https://idev.afdb.org/en/news/idev-participates-evaluation-cooperation-group-spring-meeting-2019>





What can we learn from evaluations to achieve better results in implementing the AfDB's High 5s?



Join the conversation at:
#EvalMatters

Access the blog at:
idev.afdb.org/en/evalmattersblog



Evaluation of the Middle-Income Countries Technical Assistance Fund (MIC-TAF)

In 2018, at the request of the Board of Directors, IDEV launched an evaluation of the AfDB's Middle-Income Countries Technical Assistance Fund (MIC-TAF) in order to inform the Board's annual decision-making on the net income allocation of the Bank. The objective of this evaluation was to examine the extent to which the MIC-TAF has achieved its original goal of delivering development results in beneficiary countries during the entire period of the Fund's existence (2002 to 2018).

The evaluation found that:

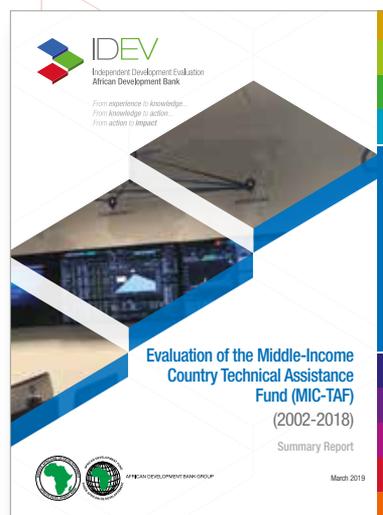
- The Fund and its operations are relevant, though the Fund lacks a clear strategic focus.
- The Fund is effective, based on its ability to achieve one of its main objectives, namely improving the Bank's portfolio in MICs; albeit with an appreciation of the Fund's limited capacity to generate development outcomes.
- The Fund is inefficient. The long delays in responding to, and processing, MIC-TAF requests have had an adverse impact on the timely completion of

investment projects and capacity-building initiatives.

- In terms of governance, the lack of strategic focus in the evolution of the Fund caused it to turn it into a financing instrument that supplements the Bank's administrative budget.
- On sustainability, the evaluation could not provide a final rating due to the limited number of Project Completion Reports available.

Some recommendations to be considered by AfDB Management include:

- Clarify the institutional arrangement of the Fund and establish an effective management;
- Enhance the financial sustainability of the



Fund and set-up a Project Preparation Facility specifically for MICs;

- Improve the Fund's guidelines and establish a stronger quality assurance process for MIC grants;
- Increase support to ongoing Bank lending

operations and consider Bank execution of selected projects when necessary.

<http://idev.afdb.org/en/document/evaluation-middle-income-countries-technical-assistance-fund-mic-taf>

Evaluation of the African Development Bank Group's Program Based Operations: Energy Governance Cluster

The cluster evaluation of Program Based Operations (PBOs, also known as budget support) in energy governance was one of seven components of a broader evaluation on the use of PBOs by the AfDB in 2012-2017. It examines eight PBOs focused on energy, which the AfDB approved and implemented in five countries (Angola, Burkina Faso, Comoros, Nigeria and Tanzania). The evaluation drew the following lessons:

1. With the exception of Crisis Response Budget Support, PBOs should be structured as medium-term operations, based upon three to four tranches over the same number of years, and should largely be a part of a sequence of multi-year PBO operations.
2. Maximizing the effects of PBOs' contribution to fiscal space requires the design

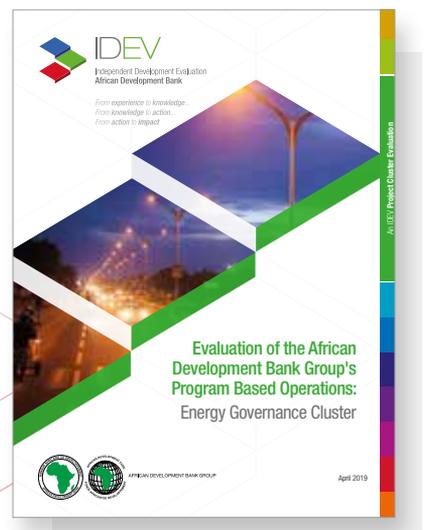
and programming of PBOs to take into account the country's immediate financial needs, and appropriate medium- and long-term structural reforms.

3. The contribution to fiscal space is the most obvious benefit of PBOs and should therefore be used strategically to ease structural constraints in support of longer-term reforms.
4. Strong and sustainable results are achieved by building upon a well-established program of reform, to which the Bank has contributed over a number of years through investment lending, technical assistance and policy dialogue.

These lessons have relevance both for the organization of the Bank's work in the energy sector as a whole, as well as for

the design and management of PBOs in general.

<http://idev.afdb.org/en/document/evaluation-african-development-bank%E2%80%99s-program-based-operations-energy-governance-cluster>



Evaluation of the African Development Bank Group's Program Based Operations: Private Sector Enabling Environment Cluster

The objective of the Private Sector Environment (PSE) Cluster Evaluation was to assess the relevance, effectiveness, efficiency and sustainability of the Program Based Operations (PBOs) of the AfDB that focused on the PSE. The PSE Cluster Evaluation is one of the seven components of a broader evaluation on the use of PBOs by the AfDB in 2012-2017.

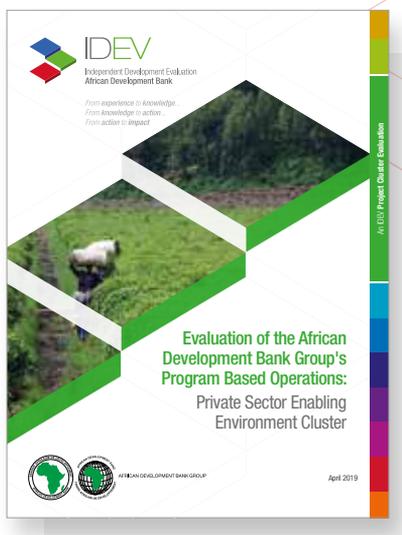
design, programming, implementation and post-implementation.

- 4. Appropriate dialogue and technical support are important for the systematic and strategic use of the PBO instrument.
- 5. Successful government implementation of complex reforms in key areas requires adequate collaboration among development partners including the Bank.

<http://idev.afdb.org/en/document/evaluation-african-development-bank-group%E2%80%99s-program-based-operations-private-sector>

Lessons Learned

- 1. PBOs are relevant, and an integral part of the Bank's portfolio, as they can play a strategic role in satisfying the development objectives of the Bank, and those of African countries and Development Partners.
- 2. Creating a conducive private sector environment starts with ensuring a stable macroeconomic context, strengthening public sector governance (including procurement rules) and improving access to key (e.g. energy) infrastructure.
- 3. The achievement of medium- and long-term structural reforms requires multi-level and sustained support in PBO



Evaluation of the AfDB Country Strategy and Program in Eswatini

IDEV has conducted an evaluation of the Bank's Country Strategy and Program in Eswatini for 2009-2018. The evaluation covered a Bank portfolio of 14 projects representing a total of UA 119 million in seven sectors: agriculture; water supply and sanitation; transport; finance; power/energy; environment; and multi-sector (governance).

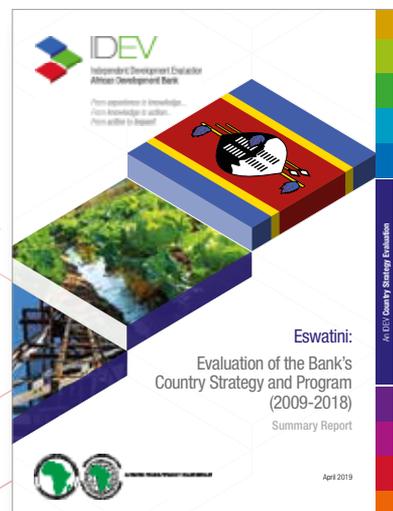
Overall, the evaluation found that the AfDB's interventions in Eswatini were relevant to the country's needs, development challenges and priorities. However, the evaluation found some shortcomings, including overreliance on the farmer company model in agriculture; a focus on highways rather than feeder roads in transport; and on using Swaziland Development Finance Corporation's existing practices to reach Small and Medium Enterprises. It also emerged that interventions were often developed in silos, which increased risk of costly duplication and re-invention of wheels. Spreading the Bank's interventions across seven sectors rendered their delivery ineffective and compromised their impact, as it reduced the amount of attention that

could be provided to each of the sectors.

The evaluation proposes **five key recommendations** that could improve the development outcomes of the Bank's interventions in Eswatini, including:

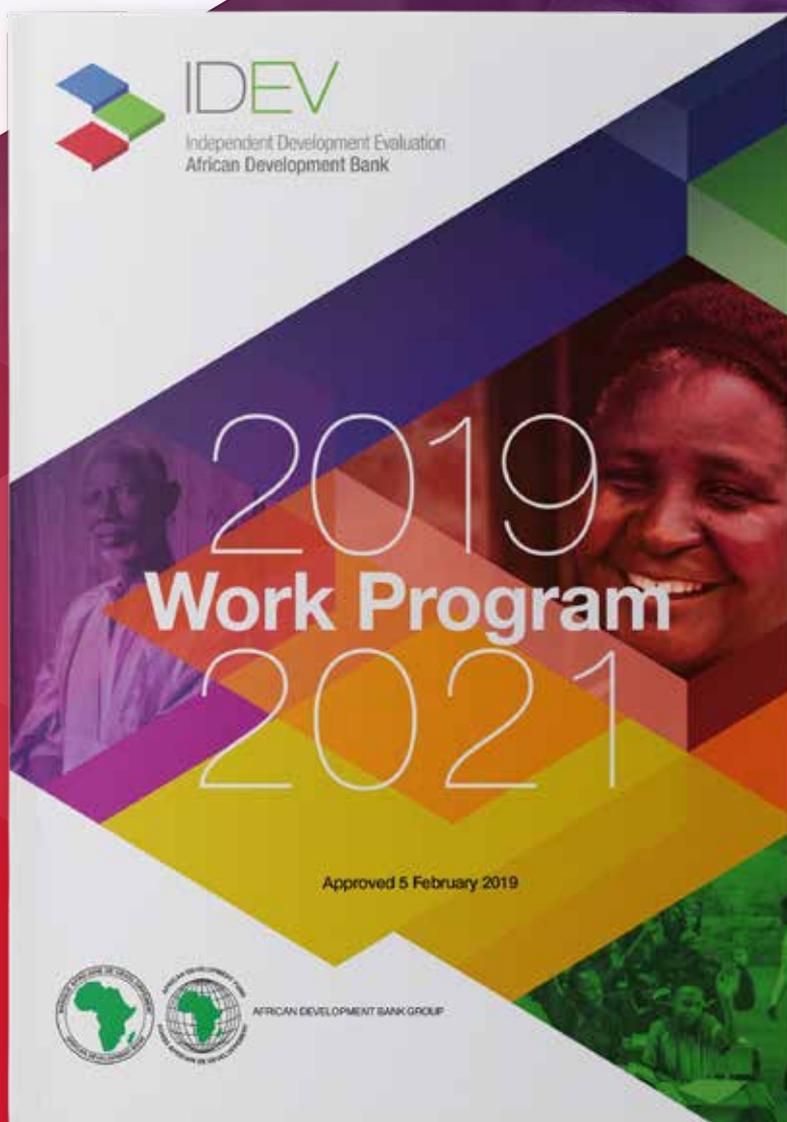
1. enhancing selectivity and portfolio design for more effective and sustainable results;
2. improving quality at entry to improve efficiency of implementation of the interventions;
3. managing for development results to improve overall aid effectiveness and reduce aid transactions costs on the Royal Government of Eswatini;
4. safeguarding development benefits to improve sustainability of Bank's interventions in Eswatini; and
5. boosting policy dialogue and knowledge management to improve project design, implementation and dissemination of results.

🌐 <http://idev.afdb.org/en/document/afdb-country-strategy-and-program-evaluation---kingdom-eswatini>

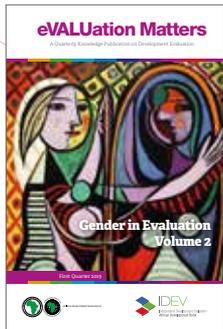


IDEV 2019–2021 Work Program now available

Find out more!



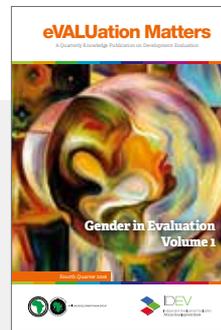
Past issues



First Quarter 2019: *Gender in Evaluation Volume 2*

Women continue to suffer significant economic, political, legal, social and cultural disadvantages in almost all societies. Evaluations of projects, programs and policies must take into account these disadvantages and provide stakeholders with sound and compelling evidence to better inform the planning and implementation of future development interventions. This edition complements Evaluation Matters Quarter 4 2018 by providing examples of how selected individuals and institutions have been able to concretely integrate Gender Responsive Evaluation approaches into their work.

<http://idev.afdb.org/en/document/gender-evaluation-volume-2>



Fourth Quarter 2018: *Gender in Evaluation Volume 1*

This edition of Evaluation Matters seeks to contribute to the debate around some of these questions, including: what types of approaches and methods that meaningfully include gender in evaluation have shown promising results? What type of information should an evaluation seek in order to assess the different impacts of development interventions on women and men at all levels? How could evaluation approaches support the change in mindset required to achieve wider societal impacts (transformative gender equality and women's empowerment practices)?

<http://idev.afdb.org/en/document/gender-evaluation-volume-1>



Third Quarter 2018: *Evaluation Week Special*

Strengthening Development Impact was the theme selected for AfDB Development Evaluation Week 2018. This edition of Evaluation Matters captures the images, discussions and knowledge shared during the three-day knowledge event on the crucial role of evaluation in facilitating the achievement of Africa's transformation agenda.

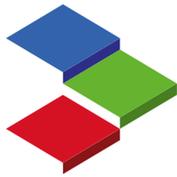
http://idev.afdb.org/sites/default/files/documents/files/EMQ3_EN_WEB.pdf



Second Quarter 2018 : *Building supply and demand for evaluation in Africa Vol. 2*

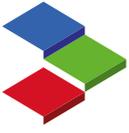
This edition of Evaluation Matters also focuses on the theme of developing a supply and demand for evaluation in Africa. While Evaluation Matters Quarter 1 2018 looks at the critical role played by evaluation in the effective implementation of good governance structures in Africa, in promoting accountability, learning, development effectiveness, and sustained and rapid economic growth, this edition showcases experiences from various evaluation stakeholders, focusing on peer learning and different views on building the supply and demand for evaluation at country and institutional levels.

<http://idev.afdb.org/en/document/building-supply-and-demand-evaluation-africa-vol-2>



IDEV

Independent Development Evaluation
African Development Bank



IDEV

Independent Development Evaluation
African Development Bank



idev.afdb.org

African Development Bank Group
Avenue Joseph Anoma, 01 BP 1387, Abidjan 01, Côte d'Ivoire
Phone: +225 20 26 28 41
E-mail: idevhelpdesk@afdb.org