

eVALUatiOn Matters

A Quarterly Knowledge Publication on Development Evaluation

Third Quarter 2015

EMERGING SOLUTIONS TO

DEVELOPMENT CHALLENGES

Volume 1

VOLUME 1

Responding to a Changing Development
Agenda: Challenges for Evaluators

Karen Jorgensen, OECD

From MDGs to SDGs: Evaluating Global
Environmental Benefits

*Geeta Batra, Juha Ilari Uitto and Lee Cando-
Noordhuizen, GEF*

Accounting for Happiness in Policy
Evaluation

B. Essama-Nssah, Independent Consultant

VOLUME 2

Inclusive Growth and Transitioning to
Green Growth in Country Strategy Papers:
Progress and Challenges

*Independent Development Evaluation, African
Development Bank*

Emerging Trends in Redesigning Results
and Alignment Tools for Country Strategies
by Multilateral Development Banks

Richard Schiere,

Scaling Up African Development Bank
Country Strategy Papers: Challenges
and Opportunities

*Independent Development Evaluation, African
Development Bank*



AFRICAN DEVELOPMENT BANK GROUP

Published by
IDEV
Independent Development Evaluation
African Development Bank
idev.afdb.org



Contents

4 Emerging Issues in Evaluation

Rakesh Nangia, African Development Bank

A huge potential and equally huge challenges await the evaluation community—which must continue to build on its new-found momentum and be bold in forging ahead. This will help bring about the qualitative changes necessary for economic transformation and sustainable development.

8 Responding to a Changing Development Agenda: Challenges for Evaluators

Karen Jorgensen, OECD

Recent and coming changes to the development co-operation agenda will require evaluators and the evaluation profession to evolve: Evaluators will need to be ready to evaluate whole of government approaches, increasingly integrated policies, global public goods, a more diverse range of partnerships and increasingly complex forms of development assistance. The key questions to be asked are: Are we fit for the future? What will it take? How will we get there?

16 From MDGs to SDGs: Evaluating Global Environmental Benefits

Geeta Batra, Juha Ilari Uitto and Lee Cando-Noordhuizen

Transitioning to a new development agenda, including to the SDGs, can be a challenge for evaluators. Evaluations will need to be strategic and focused to inform policymakers of progress towards the SDGs.

24 Mobilizing Impact Evaluations: Mobile Survey Micro-Experiments for Sustainable Development

Kweku Opoku-Agyemang, University of California at Berkeley

Fast-growing mobile markets in developing countries mean that mobile surveys can complement standard surveys in evaluations. Can mobile phone surveys play a role in impact evaluation and help us understand the development preferences of citizens around the world?

30 Investing in Development Impact Evaluations: Towards a Decision Model

Debazou Y. Yantio, Consultant, African Development Bank

Investment in impact evaluations represents a sizeable amount of development finance, which is growing despite shrinking development budgets. The central question is whether the value of the outcomes [evaluation projects] as aggregated social benefits is worth the costs, compared to other available projects.

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: Felicia Awwontom (fawwontom@afdb.org)
Produced under the guidance of the AfDB Evaluator General, Rakesh Nangia.

The team that contributed to this issue includes Deborah Glassman (consultant), Mireille Cobinah (coordinated printing); and Karen Rotmunsterman, who provided comments.

© 2015—African Development Bank (AfDB)
African Development Bank Group
Immeuble du Centre de commerce international d'Abidjan (CCIA)
Avenue Jean-Paul II
01 BP 1387, Abidjan 01
Côte d'Ivoire
Phone: +225 20 26 44 44
Fax: +225 20 21 31 00
Internet: www.afdb.org

Design & layout: Créon and Felicia Awwontom
(based on original design by Phoenixdesignaid)

38 Accounting for Happiness in Policy Evaluation

B. Essama-Nssah, Consultant

The pursuit of happiness is emerging as the ultimate goal of public policy. Policymakers should strive to identify and implement interventions that have the greatest potential for enhancing citizens' wellbeing or happiness.

49 Ethics in Evaluation and a Global, Integrated Framework for Evaluating Development Policies

Olivier Sossa

The absence of a framework for a global and integrated evaluation of the various MDGs and their interrelationships suggests that the MDGs contribute to development on an individual basis and does not take into account their synergy and interdependence. We need a global, integrated framework for evaluating development policies and for including ethics as a dimension of evaluative approaches.

54 Whither Multilateral Aid? A Bilateral Perspective on Non-core Funding to the Asian Development Bank and the World Bank

David Slattery, Department of Foreign Affairs and Trade, Australia

The emergence of non-core funding as a significant feature of the aid landscape reflects its advantages over more traditional forms of aid. Is there a risk that this funding might undermine rather than enhance aid effectiveness? ODE's evaluation of non-core funding to the ADB and World Bank suggests ways to obtain the best possible results while minimising the risks to aid effectiveness.

63 Building Africa's Innovative Capacity: the Role of the AfDB

Foster N. Ofori, African Development Bank

The somewhat successful MDGs were optimistic attempts to improve the lives of Africans. The newly ratified and adopted Sustainable Development Goals are an even more ambitious attempt to improve the lives of the poor, most of whom reside in Africa. The goals are smart. The approach to achieving them is not as smart. What can be done differently in the approach towards achieving the SDG's that was not considered in the MDGs?

68 Innovation: Africa's Forte

Shikoh Gitau, African Development Bank

The goals are smart. The approach to achieving them is not as smart. What can be done differently in the approach towards achieving the SDG's that was not considered in the MDGs?

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

Evaluator General: Rakesh Nangia, r.nangia@afdb.org

Managers:

Samer Hachem, S.hachem@afdb.org
Karen Rot-Munstermann, k.rot@afdb.org
Rafika Amira, ramira@afdb.org

Questions?

Telephone: +225 2026 2041

Web: <http://idev.afdb.org>

Write to us: EvaluationMatters@afdb.org
idevhelpdesk@afdb.org

Copyright: © 2015—African Development Bank (AfDB)



From the AfDB Evaluator General

Rakesh Nangia

Emerging Issues in Evaluation

Welcome to eVALUation Matters. In this issue, we focus on emerging issues in development and in development evaluation.

Good evaluators always look in the rear-view mirror to avoid pitfalls as they chart out the way forward. In my rear-view mirror, I observe great progress as well as rapid and significant changes taking place in the still nascent field of development evaluation. Thus, to me, it is no surprise that 2015 was declared the International Year of Evaluation. A key objective of designating 2015 the International Year of Evaluation is to advocate and promote evaluation and evidence-based policy making at international, regional, national and local levels. This is important, for despite the progress made, development evaluation and use of evidence to make policy decisions is still in its infancy.

Evaluation is generally recognized as a key feature of every learning organization. In Africa in particular, it is growing, albeit slowly, as a full-fledged discipline and practice within governments and organizations. This evolution positively affects the development process, pushing further the effectiveness of initiatives by measuring success, assessing progress, enabling mid-course correction, fostering the learning of lessons, focusing most importantly on evidence. In the last edition of eVALUation Matters, I mentioned the increasing interest in impact evaluation largely due to its ability to assess changes (intended and unintended) that can be attributed to an initiative.

Assessing impact through rigorous scientific methods remains key to maintaining the development process on sound and solid foundations. Impact evaluation thus seems to be one of the (re) emerging issues that pose questions around the opportunity and costs of applying experimental or quasi-experimental methods to inform policy decisions.

Beyond this traditional debate on scientific methods, I am keen to see an evaluation community that has greatly diversified its areas of interest by opening up evaluative thinking to various political and developmental themes. With the shift from the Millennium Development Goals (MDGs) to the Sustainable Development

Goals (SDGs), I believe the issue of evaluation of global policies will mark the world over the next decade. Despite largely positive results, many developing countries, particularly in sub-Saharan Africa, have lagged behind with respect to the MDGs. Although progress against the MDG targets could be assessed easily, evaluating and understanding reasons for the trends both locally and globally has been a much greater challenge. With the adoption of the SDGs, including a comprehensive M&E framework, global public policy will take on a new dimension with an increased consideration for important development factors such as human and environmental rights, gender, equity, sustainability, inclusion, and so on. The prescribed use of evaluation in the SDG framework will put greater emphasis on the relevance and effectiveness of public policies by responding to the dual question: “Are we doing the right things and are we doing them the right way”?



The strength of Evaluation remains its ability to focus on value, including timeliness. Whether looking at economic value, value for money, social or societal values, evaluation remains a fascinating discipline that adapts to new development challenges.”

In this context, questions such as happiness or quality of life induced by public policies gain a central place. Whether evaluating the operations of the African Development Bank or the national policies of a state, one of the major concerns is now the ultimate beneficiary: the citizen. It is henceforth the ability of interventions and policy to change lives for the better that is sought.

The strength of evaluation remains its ability to focus on value, including timeliness. Whether looking at economic value, value for money, social or societal values, evaluation remains a fascinating discipline that adapts to new development challenges. It also has the ability to make the most of the potential of communication technologies to improve its tools and methodologies and thus constantly strengthen the participatory approach by taking into account the views of all key stakeholders.

The use of mobile phone-based communications has proved useful for evaluation and it is certainly an emerging practice to be encouraged as networks spread fast throughout Africa. The challenge for the near term is how we can continue to expand horizons and use technology to bring increased participation and value for money without compromising quality and robustness. Lack of imagination and creativity is the only barrier to strengthening development outcomes and increasing value for money.

Another concern is raised by the boom of **Public-Private Partnerships** (PPPs) as a privileged instrument of development financing by states. After several decades of experimentation by states and development institutions, there are now numerous studies on the impact of PPPs, their successes and failures. In a structural context of limited public resources, evaluators and leaders share a common interest in knowing the real ability of PPPs to provide a solution for infrastructure deficits and the optimal conditions for their use. Of course, PPPs as such can only be as good as the environment in which they are undertaken, and this should spur countries to undergo the structural reforms that are needed for better governance and an enabling business environment. A body of evidence will continue to be built in order to provide more precise answers to these questions.

At this time, the question is no longer how to evaluate a particular project or development initiative, because much expertise has been acquired in this field. The major challenge for countries and organizations such as the African Development Bank is to demonstrate how broadly their policies have brought a positive change in the lives of African people. This calls for deployment of more appropriate (complex) methodologies to address global concerns about the validity and legitimacy of public action itself.

Methodological issues

In my opinion, **methodological issues** will persist in debates in the evaluation world as the development agenda progresses in our countries. The imperatives of accountability demand that we push the methodological frontiers to the farthest possible point and ensure the credibility and validity of evaluation. In this respect, with several forthcoming evaluation reports, we will see more and more debates on the quality of evaluations and their utilization, as well as on meta-synthesis methodologies to help make further sense of an increasing body of evaluative evidence.

Where we can expect further innovation in the post-2015 evaluation agenda

Mostly, the debate will continue to be about reengineering — how we evaluate our strategies and policies — triggering innovative solutions to focus more on the questions that matter to end-users in areas such as inclusion, gender, equity and efficiency. Our ability to strengthen the learning process through evaluation, to ensure positive change and knowledge sharing, will also be put to test, with out-of-the-box and technology-informed thinking needed to reach the next level.

Communication of evaluation results remains a challenge. Impressive results have been achieved lately by evaluators opening up and communicating results in an efficient manner using social media. However, this is the first step – raising awareness. We need to do better and foster greater utilization of all generated knowledge by decision makers and the general public. These challenges are important not only for development organizations but also within developing countries where evaluation is steadily gaining importance.

Last, but not least, we need **innovation to build strong national evaluation systems** that promote an evaluation culture of knowledge sharing, learning and accountability towards the institutions and the public. It is important for an institution like the African Development Bank to evaluate the effectiveness of its initiatives. But there is clear evidence that the size of development budgets is shrinking fast. In some countries, this has now become a “rounding error”. It is therefore paramount to ensure that public resources in general are not invested for limited results and without prior comprehensive appraisal of the relevance and chances of success of policies.

Thankfully, there is growing demand for strong evaluation systems and evaluation capacity building from governments and civil society organizations. Indeed, development of evaluation capacity in developing countries constitutes an additional tool to strengthen good governance in these countries. This is also recommended by the United Nations General Assembly resolution passed in December 2014 whereby the UN recognizes that evaluation is an important component of the development process and that it can strengthen and support development results. The resolution stresses the need for multilateral and bilateral cooperation to strengthen national capacity to conduct evaluation activities at the country level.

Looking ahead, and glancing in the rear-view mirror, I see huge potential and huge challenges. The evaluation community needs to continue to build on its new-found momentum and be bold in forging ahead. This is critical to making a difference and bringing about the qualitative changes necessary for economic transformation and sustainable development.

ABOUT THE AUTHOR

Rakesh Nangia is the Evaluator General at the African Development Bank. Prior to joining the AfDB, he spent 25 years at the World Bank where he held several positions including Director of Strategy and Operations for the Human Development Network and Acting Vice-President for the World Bank Institute. He attended the Indian Institute of Technology in Delhi and Harvard University and holds degrees in business administration and engineering.

Recently Completed IDEV Evaluations

- Evaluation of Bank Assistance to Small and Medium Enterprises (2006–2013)
- Independent Evaluation of Bank Group Equity Investments
- Independent Evaluation of General Capital Increase VI and African Development Fund 12 and 13 Commitments: Overarching Review
- Independent Evaluation of Policy and Strategy Making and Implementation.
- Administrative Budget Management of the African Development Bank
- An Independent Evaluation



Karen Jorgensen

Responding to a Changing Development Agenda: Challenges for Evaluators*

Recent and coming changes to the development co-operation agenda will require evaluators and the evaluation profession more generally to evolve. Commissioners or evaluation users and evaluators will be called upon to assess new areas and broader cross-government initiatives. To do so effectively, evaluators must be ready to rise to the challenge by working now to develop and adopt new tools and the approaches needed for the future of development co-operation. The key questions to be asked are: Are we fit for the future? What will it take? How will we get there?



Learning and innovation go hand in hand. The arrogance of success is to think that what you did yesterday will be sufficient for tomorrow.”

~William Pollard

Major changes in the development co-operation agenda include: 1) the shifting context of development co-operation in relation to political and economic challenges facing OECD member countries; 2) partner countries' evolving needs for assistance and co-operation; and 3) global game-changers. This article briefly explores some of the drivers for change in development co-operation and the implications that these changes have for evaluation.

* I thank Susanna Morrison-Métois for her help in drafting this article. Thanks also go to Poul Engberg-Pedersen and Ole Winckler Andersen for their comments on early drafts.

Evaluators will need to be ready to evaluate whole-of-government approaches, increasingly integrated policies, global public goods, a more diverse range of partnerships and increasingly complex forms of development assistance. To do so evaluators may need to increase focus on intermediate outcomes, build pertinent theories of change, prudently strive to make use of new sources of data, and ensure the evaluability of programmes from the onset. At the same time, evaluators must work to establish continuous feedback loops so that evaluation findings can be used in a timely manner in decision making. This continues to be a challenge despite the efforts already made.

Managers and decision-makers involved in development co-operation will need to be clear about the relationships between planning for results, ongoing monitoring of progress towards reaching results, and evaluation as an independent means of verifying how well they have fared in achieving their expected results.



Change is inevitable. Change for the better is a full-time job.”

~Adlai E. Stevenson

OECD Members' Shifting Domestic Contexts

Official Development Assistance (ODA) flows reported from member countries of the Development Assistance Committee (DAC) of the OECD totalled (net) USD 135.2 billion in 2014, equal to the previous all-time high of USD 135.1 billion reached in 2013.¹ ODA remains, for the time being, a stable source of financing for development despite the domestic economic challenges faced by several OECD member states.² At the same time, economic challenges and concerns in OECD countries and in global financial markets in recent years have led to discernible shifts in the domestic contexts of many DAC members. Beholden to public demands for increased attention to national economic concerns, many OECD countries have faced the challenges of undertaking public sector reforms and implementing deficit reduction measures. High levels of unemployment and high debt levels have forced some of the traditional providers of international development aid to become more inward-looking.

Related to changes in domestic contexts, OECD countries have become more upfront about the need for their development co-operation agenda to be an integral part of their foreign policy objectives and about development co-operation being, in part, a tool to further their political and economic diplomacy. Hence the **'national interest' dimension of development co-operation** has become more explicit. This shift is reflected in recent changes to institutional set-ups for development co-operation in OECD member states such as Australia, Canada and New Zealand - which over the last three years have seen the integration of development 'agencies' into ministries of Foreign Affairs - and in the deeper integration of foreign policy and development agendas in already integrated systems. Development co-operation must respond to the joint imperative of: 1) helping the poor, and 2) ensuring that this first goal is consistent with overall foreign policy objectives.

1 OECD (2015), "Development aid stable in 2014 but flows to poorest countries still falling", OECD, Paris, <http://www.oecd.org/dac/financing-sustainable-development/Addis%20flyer%20-%20TECH%20NOTE.pdf>.

2 Nevertheless, a worrying recent trend is the decreasing percent of ODA going to some of the poorest countries. For more information, please see the source listed above or OECD's financing for sustainable development homepage: <http://www.oecd.org/dac/financing-sustainable-development>.

Finally, there has been increased public and parliamentary scrutiny of development co-operation budgets and **calls for greater accountability** of aid programmes. This trend predates the financial crises experienced in many member states, as demand for accountability has grown in relation to increasing aid budgets over the last 15 years.³ The heightened scrutiny of public spending and development aid budgets has led to greater attention to issues such as value for money, achievement of results and potential corruption or diversion of development assistance funds. The focus is on ensuring that not a euro, a dollar, or a krone is lost.



Photo: Ray Witlin, World Bank

Partner Countries' Evolving Needs

At the same time, OECD/DAC countries have made commitments which compel them to consider the evolving needs and interests of developing partner countries. The OECD DAC's **'Fit for the Future' project** is designed to look at how DAC members' approaches to development co-operation may need to adapt in the future to be more responsive to partner country governments'

3 "In the past 15 years, net ODA has been rising steadily and has increased by 66% since 2000." Please see: OECD (2015), "Development aid stable in 2014 but flows to poorest countries still failing: Detailed summary", OECD, Paris, <http://www.oecd.org/dac/stats/documentupload/ODA%202014%20Technical%20Note.pdf>

needs and expectations.⁴ A survey of 40 partner countries found that partner countries want donors to recognise and have greater confidence in their capacity to manage their own affairs and suggested that partner countries continue to have concerns about the predictability, flexibility and alignment of ODA funding.

There is a growing demand for new forms **and increased levels of technical co-operation based on knowledge sharing and policy dialogue**; partner countries are increasingly requesting that more policy advice be shared ministry to ministry. This trend is not entirely new, but partner countries have become more vocal about their preference for this form of assistance. Overall, there is recognition that **partner countries want greater predictability** about transitions in relationships and in financing, in particular as they graduate from low to middle income status.

We have been hearing similar stories when we talk to OECD members and other partners, confirming the survey's findings. Donors, of course, have for some time recognised that local political will and ownership are necessary preconditions for successful development. As development increasingly becomes partner country-led, evaluators must therefore ask: How do we work alongside partners to further support their own initiatives? How, as evaluators, can we contribute to the ownership and alignment of aid?

Another area in which there is high demand from partner countries is the use of ODA to leverage private financial flows to developing countries. Increasingly, development initiatives in partner countries have been working through **Public-Private Partnerships (PPPs)**; these and other forms of partnership must deepen, evolve, and continue moving forward towards achieving common objectives while fostering increased levels

4 OECD (2015), "The Agency of the Future Project: Making development co-operation fit for the future", presentation at DAC meeting, 10 March 2015, OECD, Paris, <http://olisweb.oecd.org/portal/site/olisnet/menuitem.cabc5d6a61781359bb2845c643c066a0/>.

Global Game-changers

of trust, respect and mutual accountability.⁵ Work is currently ongoing at the DAC on how development assistance providers can best promote private sector initiatives and help encourage more enabling environments for investments in developing countries. Increasingly, development co-operation must focus on using ODA to catalyse investments and incentivise financial flows from the private sector. We know too little about what has been effective and why, a gap evaluators can help close.

Finally, there has been an important, progressive recognition from both partner countries and donors of the critical need to **tackle the challenge of climate change** and to **address effectively the consequences of global warming** that are being experienced now. The need for international co-operation on climate change issues is high on the agenda of developing countries as they are faced with the substantial economic costs of the consequences of global warming such as rises in sea levels, changes in rain fall patterns, extreme weather events, increased frequency of natural disasters or desertification, to name a few.⁶ If these challenges are left unaddressed, countries may suffer set-backs of hard earned development achievements.

There is much work still to be done in reaching consensus about **what works best in combating climate change** while respecting a country's domestic political and economic contexts.⁷ Evaluators will continue to be called upon to evaluate environmental, energy, natural resource management and various climate change co-operation initiatives, and may need to familiarise themselves better with the current science, relevant issues and evaluative methods being used in these fields.

This is an exciting year for the global development community. The Third International Conference on Financing for Development held in Addis Ababa in July, the adoption of the Sustainable Development Goals in September and the COP21 event in Paris in December make 2015 a **milestone for international co-operation**, a critical step on the journey to sustainability started even before the Rio Conference in 1992. There is now greater acceptance of the imperative for sustainability, equality, carbon neutral development pathways and inclusive growth - all of these ambitions are reflected in the 17 objectives adopted at the United Nations Sustainable Development Summit this September in New York. What is new about the **Sustainable Development Goals (SDGs)** is that they are universal in two ways: applicable to all countries and within each country a "whole-of-government" matter. The questions we face now are: How will we put the SDGs into action? How will we measure and monitor the progress? And, how will we evaluate the outcomes and eventual impacts of development assistance in meeting the ambitious objectives outlined in the SDGs?

There are some real challenges and opportunities for evaluators in ensuring that the 17 goals are translated into action with attention given to the expected results, impacts, and measurements of gradual progress in order to **bring more evidence into decision making**. Implementing the SDGs will require different types of partnerships, new aid modalities, across-government and across-sector approaches. Hence evaluation may need to take place in a whole-of-government, cross-sectoral way. In addition, to achieve the SDGs in all countries will also require a zealous effort to ensure that the 'unintended consequences' or negative externalities of a given country's foreign or domestic policies do not undermine or diminish the potential of other countries to achieve the SDGs. In other words, we still need to focus on policy coherence for development. As partner countries are increasingly asking for

5 OECD (2015), *Development Co-operation Report 2015: Making Partnerships Effective Coalitions for Action*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/DCR-2015-en>.

6 For some examples of how climate change is impacting small island developing states, see: OECD (2015), "Small island developing states (SIDS) and the post-2015 development finance agenda", OECD, Paris, <http://www.oecd.org/dac/financing-sustainable-development/Addis%20Flyer%20SIDS%20FINAL.pdf>.

7 For more information: OECD (2015), "Environment and development finance: capturing synergies for sustainable development", OECD, Paris, <http://www.oecd.org/dac/financing-sustainable-development/Addis%20flyer%20-%20ENV.PDF>.

policy coherence “to be made real,” we must meet the challenge of providing evidence of impact and measures of progress towards positive change.

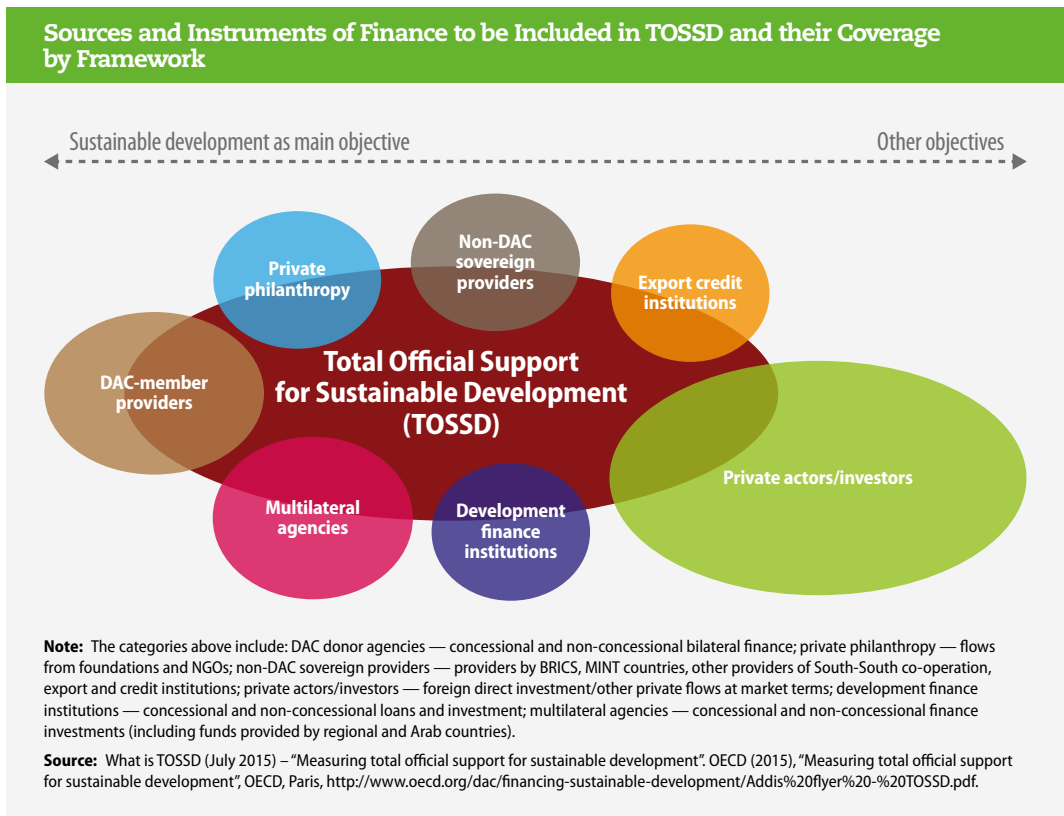
In order to implement the SDGs effectively we **will need to mobilise a wide array of domestic and international resources from both public and private actors**. Currently, the OECD is helping to advance innovations in financing for development through its policy and statistical measurement tools.⁸ The DAC’s new, proposed measurement framework, Total Official Support for Sustainable Development (TOSSD), for example, is aimed at helping to capture and incentivise financing above and beyond aid.

While negotiations around TOSSD and future financing mechanisms for development continue, we do not yet know the final form they may take; what is certain is that calls for and new initiatives to put private sector investments, increased regional and international trade flows, remittances, domestic resource mobilisation/broadened tax bases, and FDI to work for development will continue and will likely be amplified in coming years. Capturing and evaluating these flows and assessing the use of ODA to leverage other forms of financing will present future challenges to development actors, individual evaluators and to the evaluation community more generally.

Implications for Evaluation

Having discussed OECD member countries’ changing domestic contexts, partner countries’ evolving needs, and development co-operation

⁸ For more comprehensive information on the OECD’s inputs and contributions to the Third International Conference on Financing for Development event held on 13-16 July in Addis Ababa, Ethiopia, please see the OECD website: <http://www.oecd.org/dac/financing-sustainable-development/ffdandtheoecd.htm>.



game-changers, we can now address some implications that these changes will have for evaluators. Some of the most obvious consequences of recent changes will be the need for evaluation practitioners to: 1) be ready to evaluate **global public goods** or “bads”; 2) be able to **evaluate integrated policies**, and 3) be prepared to develop and employ new tools to **assess effectively a more diverse range of partnerships**.

Evaluations of partnerships should include the direct involvement of all parties, and as partner countries' confidence and capacities increase, donors must strive to foster and support their **'ownership' of the evaluation** as well. The DAC has been actively promoting joint and collaborative evaluations for over 20 years and first raised the importance of involving all stakeholders in evaluations of development activities in 1991 with the publication of 'Principles for the Evaluation of Development Assistance' (OECD, 1991).⁹ Still, our work is not done.

Joint and collaborative forms of evaluation must evolve to remain administratively light and flexible, and must aim to include mechanisms for building partner countries' evaluation capacities. Denmark, the Netherlands, Norway as well as Sweden and Switzerland have developed promising joint-light models for conducting collaborative evaluations. These new collaborative models hold future potential if they can be modified to include partner countries. Taking seriously principles of more effective aid should also lead evaluators to think more about how they can rely on each other's work through division of labour - including international organisations. New models may be needed to include non-state actors - especially important when working with the private sector.

As international development co-operation programmes and partnerships become more

complex and inclusive, there is greater need to ensure from the outset that programmes and other international co-operation initiatives can be evaluated, and to ensure that cross-cutting issues (such as gender, sustainability, and equality) are included in the evaluation Terms of Reference. Evaluators must, therefore, also be part of the movement to strengthen collaboration across policy communities and different 'sectors', which may currently lack a common evaluation culture. This is related to the need to **increase demand for evaluation from policy makers** and other development partners - creating a framework of joint ownership and mutual accountability for development results among all stakeholders.

As development co-operation seeks to incite private sector initiatives and investments, evaluators must be ready to **evaluate innovative pilot programmes** and **Private-Public Partnerships (PPPs)** in a rigorous way prior to their scaling-up. This may require that policy makers invest more in research and evaluation budgets and make conscious, selective decisions about what to evaluate in order to build a strong evidence base on what works. Increased focus on the private sector will force development evaluation practitioners to develop more precise technical definitions and to develop new evaluation tools. For example, evaluators will increasingly be called upon to define and evaluate **'development additionality'**.¹⁰ At the same time, evaluators will have to recognise and **assess what levels of risk tolerance** are acceptable for various organisations and Public-Private Partnerships (as well as actors' risk aversion related to financing in fragile states). If ODA is to be used to provide initial financing to leverage further private sector funds, then these programmes must be evaluated in a credible, independent manner with specific indicators, plans for evaluation, and theories of change

⁹ As pointed out in OECD's recent flagship publication: OECD (2015), *Development Co-operation Report 2015: Making Partnerships Effective Coalitions for Action*, OECD, Paris, <http://dx.doi.org/10.1787/DCR-2015-en>.

¹⁰ For more information and discussion on development additionally and measuring additionally both 'ex ante' and 'ex post', please see: DCEC (2013), "Donor partnerships with business for private sector development: What can we learn from experience?" Melina Heinrich, DCEC Working paper, Cambridge, 2013, <http://www.enterprise-development.org/page/download?id=2147>.

defined at the onset. Evaluating development programmes and investments may present even more of a challenge as these become increasingly partner owned and executed. Evaluators will therefore be faced regularly with the challenge of measuring and defining “knowledge creation” and knowledge sharing among partners. There will be even greater need to support partners in conducting their own impact evaluations to get evidence to influence domestic policies as well as the policies and approaches of development co-operation providers.

Calls to increase the professionalisation, training and recognition for evaluation as a specific ‘metier’ or profession will surely continue, not least in response to the **use of new technologies** and the **increasing availability of data**.¹¹ At the same time, operational staff will continue to participate in and be called upon to conduct evaluations. The evaluation profession will continue to balance these competing needs and trends and will need to evolve to continue to keep pace with general changes in development co-operation.

With more complicated programmes and partnerships, managers of development co-operation and evaluators will also have to focus more attention on what they need to know and, as a result, be more strategic about what is evaluated, when and for what purpose. Monitoring, as part of ongoing management towards results, will help inform managers and decision-makers about when evaluations are needed. Whereas monitoring is closely related to moving us to results, evaluations must continue to be separate and independent in order to help us understand why approaches are working and whether or not we have reached our goals as intended.

11 PARIS21 (2015), “Financing a country-led data revolution”, OECD, Paris, <http://www.oecd.org/dac/financing-sustainable-development/Addis%20flyer%20-%20P21%20Financing%20a%20country-led%20data%20revolution.pdf>.

Regardless of the technology or methods to be used, there will be a continued, strong demand to demonstrate results and to identify and manage risks. The call for more transparency and greater rigor may lead to increased competition to get evaluations out in a timely way. Hence, we need to ask how evaluators can meet or even drive the political agenda through **timely evaluation**. To do this, the evaluation community may have to **grapple with various trade-offs** such as between timeliness and rigour. Other trends, such as the move towards greater integration of evaluation and development co-operation into ministries of Foreign Affairs could help to make evaluation a core part of development co-operation organisations. However, the trend towards greater integration and the need for a whole of government approach may also challenge traditional concepts of independence or may require new institutional set ups.

Will the broadened policy agenda and need for engagement across government, give rise to the creation of more independent evaluation institutions (such as ICAI in the United Kingdom or DEVAL in Germany)? Would these units be charged with ensuring overall accountability and cross-government policy coherence, with a focus and mandate to look at policies and programmes going far beyond development co-operation? Meanwhile, could one imagine that development evaluation units would be used increasingly for evidence and learning purposes? The DAC Network on Development Evaluation - a body that unites evaluators from across DAC member countries and multilateral development banks and institutions - has recently launched a major new review of evaluation units’ set-ups across its membership. Hopefully this will provide insights into how DAC members are currently structuring and using their evaluation units as well as what we will need in the coming years.

Conclusion

Evaluators will continue to be called upon to contribute to the learning, the accountability and the results agendas. Evaluators and evaluation units should use the current pressure for results to raise the profile of evaluation within their organisations and focus on the added value they can bring in regards to helping to define and clarify processes of change and complexity. In the context of the 2030 agenda, development co-operation evaluators will need to widen their field of vision, adjust their toolboxes to accommodate new partnerships and approaches and develop new ways of working across the whole of government to continue to ensure that development co-operation is driven by evidence.

Suggested further reading

OECD/WTO (2015), “Aid for trade at a glance 2015: Reducing Trade Costs for Inclusive, Sustainable Growth”, OECD Publishing, Paris, http://dx.doi.org/10.1787/aid_glance-2015-en.

OECD (2015), *Development Co-operation Report 2015: Making Partnerships Effective Coalitions for Action*, OECD, Paris, <http://dx.doi.org/10.1787/DCR-2015-en>.

OECD (2015), “Development finance: Total Official Support for Development –an emerging concept in support of the POST-2015 Sustainable Development Goals”, OECD, Paris, [http://www.oecd.org/dac/DCD-DAC\(2014\)48-En.pdf](http://www.oecd.org/dac/DCD-DAC(2014)48-En.pdf).

OECD (2015), “Financing in crisis? Making humanitarian finance fit for the future”, Rachel Scott, OEC Development Co-operation Directorate, Paris, <http://www.oecd.org/dac/OECD-WP-Humanitarian-Financing-Crisis%20.pdf>.

OECD (2015), “Measuring total official support for sustainable development”, OECD, Paris, <http://www.oecd.org/dac/financing-sustainable-development/Addis%20flyer%20-%20TOSSD.pdf>.

OECD (2014), *Development Co-operation Report 2014: Mobilising Resources for Sustainable Development*, OECD Publishing, Paris, http://www.keepeek.com/Digital-Asset-Management/oecd/development/development-co-operation-report-2014_dcr-2014-en#page73

ABOUT THE AUTHOR

Karen Jorgensen Heads the Review, Evaluation, and Engagement Division in the OECD Development Co-operation Directorate.

Prior to joining the OECD, she spent 15 years at UNDP, where she was in charge of the multi-country office in Samoa, Senior Deputy Country Director in Afghanistan, Chief of the Management and Investigation Section of the Office of Audit and Performance Review, and Assistant Director of the Sustainable Energy and Environment Division and Special Assistant to the Executive Director of UNEP in Nairobi. Previously, she was Chief of Section in the Norwegian Ministry of Environment and served as Norway’ Chief Negotiator for the Global Convention on Trans-boundary Movements of Hazardous Waste. She holds a B.A. in Political Science and International Relations and a Masters of Arts/Ph.D. in Political Science from the University of Minnesota.



Geeta Batra, Juha Ilari Uitto, and Lee Cando-Noordhuizen

From MDGs to SDGs: Evaluating Global Environmental Benefits



As we transition from the Millennium Development Goals (MDGs) to the Sustainable Development Goals (SDGs), the evaluation community has opportunities, particularly in evaluating global environmental benefits. It can come together to invest in new methodologies to evaluate complex systems that deliver environmental and socio-economic benefits, encourage innovative assessments of impacts, and broaden the adoption of these methodologies through replication and scale up by developing partnerships and sharing knowledge.

Introduction

The year 2015 is an important one for the development community. Several summits are setting the stage for international cooperation over the coming decades. In September 2015, the MDGs expired. The UN General Assembly subsequently adopted SDGs, which come with a broad, universal, ambitious, and action-oriented development vision and agenda. They recognize the linkages and interdependence of the challenges of eradicating poverty, combating inequality, creating sustained and inclusive economic growth, and preserving the environment.

In July 2015, the Addis Ababa Action Agenda (AAAA) was adopted at the Third International Conference on Financing for Development. It provided a foundation for implementing the post-2015 agenda. The AAAA included a new global framework for financing sustainable development by aligning all financial flows and policies with economic, social and environmental priorities. It guides future actions by governments, international organizations, civil society, business and philanthropies. The various commitments of resources made in the context of the conference are estimated to be between half a trillion and a trillion USD to be invested in sustainable development. Much of this money will move through development banks. The private sector is expected to provide additional resources.

In December 2015, France will be hosting the 21st Session of the Conference of the Parties to the United Nations Framework Convention on Climate Change (COP 21/CMP 11). On the table will be a new, legally binding international climate agreement about global warming that is binding on all countries.

How can these interrelated goals of improving peoples' lives and achieving economic growth while addressing environmental challenges such as climate change and food security be achieved? What have we learned and what do we need to know? What do we need to do better? What do we need to do differently?

Evaluation plays a crucial role in addressing these questions. It does so by providing evidence about (1) what works, for whom, and under what circumstances in development investments; (2) the effectiveness and efficiency of policies and programs and, most importantly; (3) the outcomes and impacts on the most disadvantaged communities, families and individuals, women and men.

From MDGs to SDGs: a smooth transition?

First, there were eight goals. However, the MDGs did not specifically address economic development, make any mention of human rights, or refer to monitoring, evaluation, or accountability. Now there are 17 goals, 169 targets and 304 proposed indicators. The SDGs are intended to be universal with a common global vision of progress towards creating a safe, just, and sustainable space for all human beings to thrive. They reflect moral principles that no person or country should be left behind and that all should be regarded as having a common responsibility for playing their part in delivering the global vision (UN Stakeholder Forum 2015).

The SDGs come with acknowledged challenges. Three of these are discussed below.

Data and measurement

Indicators are the backbone for monitoring progress towards the SDGs at local, national, regional and global levels. A sound indicator framework can turn the SDGs and their targets into a management tool to help countries develop implementation strategies, allocate resources, issue a report card to measure progress and ensure the accountability of all stakeholders (UN SDN 2015). The many goals, targets and indicators of the SDGs, each with its measurement complexity, create challenges, particularly when data is already lacking or weak in many countries, making measurement and tracking even more difficult. For multilateral institutions to incorporate SDG indicators into results frameworks will require commitment and robust monitoring. Even with good monitoring systems, indicators are limited in their inability to capture spillover, the inter-relationships across implementation goals or other complexities and provide insights into why things happen. Evaluation will need to fill this gap.

Operationalization

SDGs are universal in character but need to be adapted to the specific constraints and opportunities of national contexts. Countries must specify their global commitments to create an enabling environment for sustainable development worldwide. Adapting global goals into national targets ensures their ownership and facilitates answerability, thereby promoting an accountability framework that is inclusive, transparent and participatory (UN 2015). A recent evaluation of the World Bank's Results and Performance Report offers some lessons for operationalizing the MDGs in country strategies with varying degrees of focus. The report found that Bank support was often followed up through knowledge and support programs and the results measurement system for IDA incorporated MDG indicators into country results. However, a clearly articulated results chain for MDGs and a robust M&E system would have provided stakeholders and partners with better information on institutional responses.

The Global Environment Facility (GEF) is currently the only multi-donor funding mechanism specifically designed to help tackle key global environmental issues. It would also need to adapt to respond to the SDGs.

Measuring country level contributions to global benefits and addressing complex challenges of sustainability at all levels—institutional, environmental, socio-political and economic—require a more integrated, holistic framework, while working with global conventions with more specific goals. Demonstrating the impacts of environmental programs has become far more important while the impacts of individual projects may never deliver significant global environmental benefits. Indeed, it is unreasonable to expect otherwise. Unlike most large-scale interventions, environmental outcomes and impacts are often difficult to determine because of the timescale for achieving them and the inherent problem of inadequate metrics.

Fragmentation

The 1987 Brundtland Report¹ introduced sustainable development – a new paradigm for environmental sustainability, economic growth and social equality – to the international community. However, the effective integration of environmental and social issues into economic decisions remains challenging. The concept of public goods serves as a good reference point for addressing these questions.

Public goods are defined in economic terms as ‘non-rival’ and ‘non-excludable,’ that is, impossible to trade (Van den Berg 2011). Global public goods have a fairly universal impact on a large number of countries (more than one group of countries or regions), people (several and preferably all population groups), and generations (current and future, or at least current generations without jeopardizing the development options and opportunities of future generations).

Some goods, such as the moonlight or the atmosphere, have always been considered global public goods. Others have changed over time from being considered local or national public goods (or «bads», if they have negative effects) to global in their benefits or costs. Communicable diseases such as HIV/AIDS, for example, have spread across borders. The evolving concept of global public goods or bads indicates that we are facing just one major challenge: how to rethink and reorient public policy-making to catch up with today’s new realities of interdependence and globalization (Kaul et al. 2003).

Environment as a public good tends to get short-changed, being viewed in the short-term as an externality with perceived trade-offs between economic development and environmental protection (Uitto 2014). National environmental policy is often fragmented and lacking in coherence as environmental issues often fall under the purview of several authorities or departments.

¹ In 1987, the World Commission on Environment and Development published “Our common future,” a report that came to be known as the Brundtland Report after the Commission’s chairwoman, Gro Harlem Brundtland. It developed guiding principles for sustainable development, as it is generally understood today.

Similar challenges exist at the international level. There are a large number of issues and a wide range of institutions with environmental agreements with overlapping roles. Bilateral donors, international institutions, development banks and export credit agencies are exposed to fragmentation risk when their sectoral development programs and policies do not adequately take into account the broader sustainable development perspective (UN 2012). Policy relevance and effectiveness would require systematic monitoring and evaluation.

Treating climate change adaptation independently of agriculture, water management, or health care may make sense for some as these areas fall under different institutions or ministries. Yet, for families and communities, a web of linkages and ripple effects connect these issues. Many poor communities in many countries rely on the environment and natural resources for their livelihoods. It is imperative to take a broad-based approach to understanding the multiple linkages across the different sectors and strategies, as is seeing these dimensions through a broader systems lens. Evaluators have a certain responsibility to provide policymakers with feedback on environmental benefits within the broader landscape of sustainable development and to improve program design and implementation: they must do things differently.

Moving Beyond Assessing Individual Interventions

Evaluation must move beyond assessing individual interventions in isolation. It must contribute to understanding how to better incorporate environmental concerns into national and global development efforts. This does not imply abandoning project evaluation, which plays an important role for accountability and in improving specific interventions (Uitto 2014). As environmental evaluation becomes part of a broader discussion about development effectiveness, it will contribute to the broader understanding of the reasons for the micro-macro paradox of treating the environment

as a global public good an externality to economic development.

Why don't achievements observed in interventions translate to the national level? How can a country have many interventions that score moderately to highly satisfactory on achievement of outcomes and yet not reflect the same level of achievements in its national development? (Uitto 2014). How does one evaluate social, economic and institutional results in a program with a global environmental mandate?

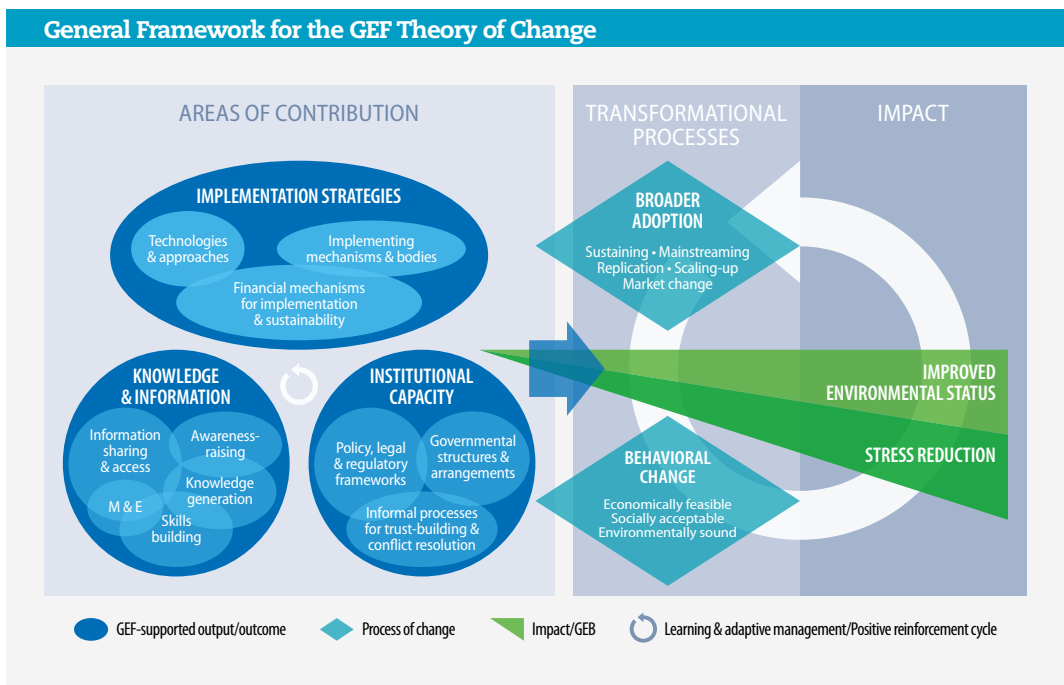
It is important to define results and impacts at different levels: direct results attributable to an intervention, a subsequent level of results that may reduce environmental stressors, to a final ecosystem impact. The relevance of interventions could then be evaluated against whether or not the changes lead to the ultimate impact of environmental improvement and increased wellbeing rather than to the alignment of activities with national or global priorities (Van den Berg 2011).

Beyond contributing to an evidence-based analysis of results and explaining what works, why, and

under what circumstances, evaluation can help bring improved integration and mainstreaming of environmental concerns into development programs (Van den Berg 2011).

Mixed Methods to Mixed Approaches in Complex Systems

An impact evaluation's primary task is to determine which impacts were caused by an intervention, as distinguished from those caused by other actions. In complex systems, where elements are interconnected and interdependent, feedback loops shape how change occurs; often unpredictable behavior emerges from the interaction of the parts. Complex systems often address environmental and socio economic outcomes. They require a different way of asking questions. Interventions in complex systems are prone to manifesting less obvious types of impact (unintended, indirect and secondary), which may be a means to assess the role of interventions in complex systems. Paying more



attention to these may help achieve the longer-term, large-scale impacts that a development organization targets (Garcia and Zazueta 2015).

The GEF addresses different levels of impact of climate change phenomena – local, national, regional and global. Evaluation questions that need to be asked, in addition to the underlying project’s theory of change for the given intervention, must understand the system that the intended activity was trying to influence, including its limits, components, their interactions and emergent properties characteristic of each system.

Quality Assurance and Safeguard Mechanisms

Quality assurance and safeguard mechanisms are important for advancing the integration of the environmental, economic and social dimensions of sustainable development. They aim to avoid, mitigate, or minimize the adverse social and environmental impacts of programs and projects and to produce positive outcomes for people and the environment. Consequently, while these mechanisms do not define strategic orientations or programmatic priorities for sustainable development, they contribute to the achievement of sustainable results.

The WBG and the GEF have carried out pioneering work in this regard. A key GEF principle is that GEF-financed operations that achieve benefits in one area should not lead to adverse environmental or social impacts in other areas. The GEF applies good-practice fiduciary standards and has established high standards for environmental and social safeguards, gender mainstreaming, and engagement with civil society organizations and indigenous peoples (UN 20124). Evaluations need to be able to address the multiple global environmental and socio-economic benefits generated by these interventions and pay attention to cross cutting issues such as gender mainstreaming.

Partnerships and Stakeholder Engagement

Sustainable development is not possible without including local governments and local community organizations at the forefront of the agenda. This requires accountability and a strong engagement with partners having diverse skills. Complementarities can be achieved through a careful, adaptive approach, ensuring that duplication of efforts is avoided.

Despite the continuing importance of and need to increase official development assistance and other public sector funds, it is clear that there will be huge financial demands on the private sector, whose important role has been emphasized in the SDGs. The past few years have seen a number of innovative public-private partnerships designed to reduce investment risks, optimize the use of public and private finance, and pool human resources and strategic capabilities. Practitioners recognize increasingly that such partnerships could play a pivotal role in scaling up sustainability efforts in both developing and developed countries. Evaluating the effectiveness of partnerships across these sectors and the private sector would be very important.

Knowledge Management

The effectiveness of multi-stakeholder partnerships for capacity building and learning at local, national, regional and global levels will be tied increasingly to the ability to manage and share knowledge and expertise about the issues, processes, and solutions being promoted. Traditional publications, websites, newsletters, implementation registries, knowledge hubs, and technology-sharing platforms can help to disseminate knowledge and expertise. These products and practices also allow multi-stakeholder partnerships to effectively track progress, and provide the ultimate stakeholders with ownership, transparency and accountability (UN DESA 2015).

Knowledge management is an integral part of the evaluation process and can strengthen

interdisciplinary collaboration, effectively coordinate big datasets, and fill gaps. The use of modern information and communications technology such as text messaging make monitoring, evaluation, and the delivery of real time data possible. If monitoring and evaluation frameworks were to draw on these new possibilities, the data could be collected, interpreted and integrated into an overall evaluation based on sound scientific methods (UN Scientific Advisory Board 2014). Methods to mainstream participation, partnerships, gender, and private sector involvement across evaluations will have to be developed without overburdening evaluations.

Conclusion

More than 20 years after the Rio Earth Summit, the challenge and opportunities of sustainable development are more relevant than ever. With increasing clarity and acceptance, economic growth, environmental protection and social equity are interlinked and belong to one and the same agenda – the sustainable development agenda. Progress in one area depends on progress in other areas.

Evaluation plays an important role in the achievement of the SDGs. It is needed to understand why things happen or not, what trade-offs are required or must be considered, what the accelerating factors are, and what works for whom and under what circumstances.

Transitioning to a new development agenda, including the SDGs, can be a challenge for evaluators. Evaluations will need to be strategic and focused to inform policymakers of progress towards the SDGs. Some guiding principles for this to happen:

1. Maintain strategic focus and set priorities: Be mindful of the risk of extending evaluations into too many areas because of the broad agenda.

2. Draw on past lessons for new projects and develop good practice results frameworks to guide the operationalization of projects and programs.
3. Develop a more integrated and holistic evaluation framework while working with global conventions with more specific goals that demonstrate the clear linkages between environmental, economic and social pillars.
4. Move beyond assessing individual interventions in isolation and contribute to understanding how to better incorporate environmental concerns into development efforts in the national and global contexts.
5. Ensure quality assurance and safeguard mechanisms. These are important to advance the integration of the environmental, economic and social dimensions of sustainable development, to avoid, mitigate, or minimize adverse social and environmental impacts of programs and projects and produce positive outcomes for people and the environment.
6. Strategic partnerships will require accountability and strong engagement with partners with diverse skills. Take a careful and adaptive approach to achieve complementarity, thereby ensuring that duplication of efforts is avoided while achieving the most productive synergies.
7. Knowledge management is an integral part of evaluations. It can strengthen interdisciplinary collaboration, effectively coordinate big datasets, and address evaluation gaps on issues such as gender, partnerships, and private sector participation strategically.

References

- Garcia J. and A. Zazueta, 2015. "Going Beyond Mixed Methods to Mixed Approaches: A Systems Perspective for Asking the Right Questions," *IDS Bulletin* 46 (1), 30-43.

- Kaul, I., P. Conceicao, and K. Le Goulven (Eds.), 2003. *Providing Global Public Goods: Managing Globalization*. New York and Oxford: Oxford University Press.
- Uitto, Juha I., 2014. "Evaluating environment and development: Lessons from international cooperation," *Evaluation* 20(1): 44-57.
- UN, 2015. *Mainstreaming the three dimensions of sustainable development throughout the United Nations system (A/69/79-E/2014/66)*
- UN Department of Economic and Social Affairs, 2015. *Transitioning from the MDGs to the SDGs: accountability for the post-2015 era*.
- . *Sustainable Development in Action: Special report on voluntary multi-stakeholder partnerships and commitments for sustainable development*.
- UN High-level Panel on Global Sustainability, 2012. *Resilient People, Resilient Planet: A future worth choosing*. New York: United Nations.
- UN Scientific Advisory Board: *The Crucial Role of Science for Sustainable Development and the Post-2015 Development Agenda*, 2014. UN Stakeholder Forum: *Universal Sustainable Development Goals*, 2015.
- UN Sustainable Development Solutions Network: *Indicators and a Monitoring Framework for the Sustainable Development Goals*, 2015.
- Van den Berg, Rob, (2011). "Evaluation in the context of global public goods," *Evaluation* 17(4): 405-15.

ABOUT THE AUTHORS

Juha I. Uitto is Director of the GEF Independent Evaluation Office. Prior to this position, he was Deputy Director and Evaluation Adviser in the UNDP; Monitoring and Evaluation Coordinator/Specialist with the GEF and coordinator of the United Nations University's environment and sustainable development research and training programs. He has also held positions in other development organizations and in academia.

Juha 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 30 peer reviewed articles and book chapters on topics related to the environment, natural resources management, environmental hazards, and evaluation. He has conducted and managed a large number of programmatic and thematic evaluations related to environmental management and poverty-environment linkages. His book "Evaluating Environment in International Development" was published by Routledge-Earthscan in May 2014. In October 2012, the European Evaluation Society awarded him for a "distinguished contribution to evaluation practice."

Geeta Batra is Chief Evaluation Officer and Deputy Director in the GEF Independent Evaluation Office. She was previously a manager in the Independent Evaluation Group (IEG) of the World Bank Group; Global Head of Results Management for the International Finance Corporation's Advisory Services; a Senior Private Sector Specialist in the World Bank and Senior Manager for Risk at American Express. Geeta holds a PhD in Economics.

Lee Cando-Noordhuizen is a consultant in the GEF Independent Evaluation Office. She has worked for various UN agencies including the Executive Office of the UN Secretary-General, UNDP, UNICEF and UNFCCC in the fields of climate change and sustainable development. Recently, Lee was posted in Indonesia where she worked as a Climate Change Communications Coordinator for the Center for International Forestry Research (CIFOR), and as a Senior Advisor to the German International Cooperation's (GIZ) ASEAN-German Programme on Response to Climate Change.



Kweku Opoku-Agyemang

Mobilizing Impact Evaluations: Mobile Survey Micro-Experiments for Sustainable Development

The unanswered question of how to reconcile sustainable development and subjective well-being remains urgent. The Sustainable Development Goals represent a new and exciting chapter in development practice, but it is unclear how evidence-based policy can play a defining role. Related dilemmas facing the next generation of policymakers include how to address international migration, conflict and global health concerns. At the core of all of these issues is the fact that we still know little about the preferences of people in developing countries and how those preferences evolve over time.

Although the impact evaluations revolution coincided with the past Millennium Development Goals, it is clear that these research tools will need to evolve in scale and scope for the sustainable development agenda to be more inclusive and incorporate the feedback of developing country citizens into policy endeavors. Although many evaluations have been carried out in Africa and other developing regions, relatively few academics and organizations in these areas can afford to independently design such research programs. Impact evaluation budgets can reach the hundreds of thousands of US dollars, effectively excluding most people in developing regions from paying for and carrying out such projects. Although innovative programs have valiantly attempted to create research partnerships across developed-developing contexts, their sustainability is prone to similar cost concerns. For similar reasons, it is rarely feasible for impact evaluations to incorporate the perspectives of developing country citizens who live in the contexts under study. However, such financial constraints precede the arrival of modern impact evaluations in development studies.

This problem may negatively affect the sustainability of development through indirect mechanisms: many local experts and citizens have contextual knowledge from which impact evaluation designs would arguably benefit, but scarcely exploit, possibly leaving gaps in identification while perhaps placing a theoretical ceiling on the usefulness of policies that emerge from research investigations. The specific financial constraint for scaling up the use of impact evaluations has been data access, a key ingredient of any impact evaluation. Although science, technology, and capacity building are pillars of the Means of Implementation of the Post-2015 Agenda, it is unclear how evidence-based research can be operationalized in innovative ways to become more widespread.

The innovative data collection approach discussed in this paper involves the use of mobile phone surveys to access information efficiently and cheaply and serve the needs of development policymakers and citizens alike. Fast-growing mobile markets in developing countries mean that mobile surveys can complement standard surveys in evaluations. Can mobile phone surveys play a role in impact evaluation and help us understand the development preferences of citizens around the world?

This paper describes how mobile phone surveys can be used to conduct impact evaluations by using innovative randomized survey experiments, or micro-experiments. The advantages of mobile surveys over traditional impact evaluations go beyond financial costs. For example, traditional experiments may take several years to be implemented from baseline to outcome measurements, whereas mobile survey experiments can be completed in a matter of weeks. The data can reflect thousands of observations in days — with the robust statistics that researchers require for rigorous evaluations. Mobile phone-based surveys can accelerate the data collection process in developing countries, make impact evaluations more inclusive, and make policy-making more innovative.

Given growing global access to mobile phones, mobile survey experiments may help both impact evaluation research and policymaking make significant strides towards incorporating the voice

of the poor and marginalized in evaluations. Mobile platforms offer the potential to support economic development programs and initiatives to succeed in developing countries by engaging citizens in new and exciting ways. Strategic partnerships with mobile network providers or non-governmental organizations in the technology and development space can play a greater role in implementing impact evaluations for sustainable development. Since developing countries are striving to mobilize more domestic resources for economic progress, such approaches are highly relevant for the sustainable development era.

This paper describes how novel mobile survey experiments can be implemented in such contexts, drawing on the author's recent research in Ghana. It then discusses how these can be used for mobile survey impact evaluations and nest them in open data programs and policy innovations that are occurring across Africa with implications for sustainable development.

In Ghana, many policymakers regularly participate in radio programs to share their policy visions with their communities and receive feedback on policy endeavors. Such programs came into their own during the second wave of Ghanaian democracy in the 1990s, when telephone booths were the norm. Many people would queue for hours for a chance to call into a radio program to engage with relevant policy concerns. Since the turn of the century, the stratospheric rise of mobile telephony has provided surprisingly limited improvements – many more Ghanaians are able to call into radio stations but the abilities of radio stations to receive calls has not changed significantly. The demand to engage in discussions on poverty and other development topics has thus appeared to outstrip the supply over time. Although such arrangements have served a significant purpose in bringing people closer to the officials that serve them over time, they still have certain shortcomings. For example, since community radio programs are relatively short (under half an hour), most callers are never

able to speak directly to an official even if such interactions are theoretically possible.

One constraint of the status quo may be particularly familiar to impact evaluation researchers. It is never clear to most listeners in the audience whether questioners to policy makers are cherry-picked by radio hosts, or if the callers are randomly chosen and do indeed represent the majority of Ghanaian perceptions. Radio hosts must continually assure listeners of their integrity so as not to lose their audiences to competing stations. A perception of significant variation in journalists' ability to be fair may be one reason why there is often little engagement on various topics such as poverty.

A related observation is that the discourse on international development has suffered from a similar self-selection bias (Kweku 2015). Despite several policy efforts, the voices of many in developing countries are inadvertently excluded from discussions on international development and global poverty. As a result, we know little about the preferences of developing country citizens with respect to development processes.

Mobile surveys provide a unique opportunity to understand the origins of policy preferences. The key hypothesis of the paper is that priming study participants to think of international development might affect their preferences for local issues and help policy makers understand how these perspectives evolve. Several awareness programs have been implemented in the areas of health, governance, sanitation and other areas, but unfortunately, there is still a limited sense on the subjective needs of most poor people. With the assistance of VOTO Mobile, a software solution based in Ghana, I designed and implemented automated mobile survey experiments in the summer of 2014 to shed light on the origins of citizen preferences for policy issues.

VOTO Mobile is an easy to use software solution that relies on voice-based surveys to access Ghanaians across the country for research. Although there is an encouraging trend in the growth of mobile

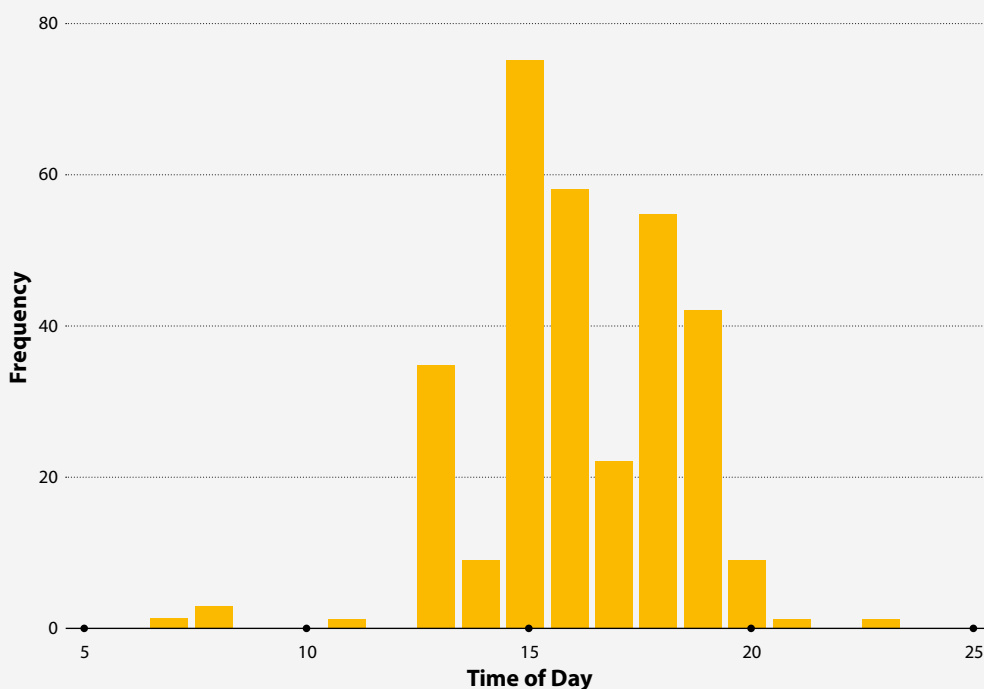
survey programs across Africa and the developing world in general, VOTO Mobile is unique. Many solutions rely on SMS (short message services) text messages that are often limited to a minority of literate people while the ultra-poor must rely on limited radio programming for access to policy information. Voice-based mobile surveys in local languages and dialects on the other hand, reach relatively representative populations for arguably better external validity.

Mobile survey evaluations are also distinct from standard evaluations in terms of their implementation. The surveys are executed randomly using an algorithm that is able to conduct randomized mobile calls from a national database of study participant phone numbers. Investigators do not see actual phone numbers so as to blind research projects. The VOTO Mobile solution optimizes for times when users are likely to be away from their phones or taking other calls, or other factors that

might lead to self-selection among users that respond to survey questions. The system is able to record how long a participant takes to complete a survey, which question was answered when, and provide both audio responses for qualitative data and quantitative information embedded in a familiar spreadsheet. Both treatment and endline surveys were completed in two-and-a-half weeks. Respondents could take breaks and complete surveys at their convenience throughout the day and on their own terms so that the survey was only minimally disruptive to work and life schedules. Interestingly, many participants answered the surveys in the afternoon, perhaps after work duties (Figure 1).

Although information technology makes the process of evaluation more efficient than is usually the case by automating the survey process, complementing the surveys with in-person interactions was critical to understand citizens' subjective perceptions of the various policy topics to suit the evolving context.

Figure 1: Mobile Survey Adoption (Completion) Based on Time of Day



The horizontal axis shows the time of day surveys were completed and the vertical axis shows the frequency of survey completions.

During the research, several policy makers and radio hosts complained about the difficulties of engaging citizens on important policy issues, an issue that appeared to be of significant concern. Paradoxically, many citizens simultaneously complained that they felt left out of the governance processes in their communities. For a survey experiment to understand the origins of policy preferences, a policy question exogenous to the listless local discourse had to be included.

In the treatment instrument of the mobile survey, Ghanaians were asked what they thought about the international aspect of development in Ghana. Audio responses were recorded and respondents could contribute for as long as they desired, in the hope of minimizing whatever discomfort might occur during traditional in-person semi-structured interviews. Many Ghanaians had favorable impressions of various non-governmental organizations working in their communities. The control group participants were recruited for the study but not exposed to the treatment survey question. In the final mobile survey, respondents were asked to rank policy performance and provide their preferences in the areas of democracy, sanitation, and anti-malaria policy. To gauge economic progress in the country, study subjects also discussed whether they thought personal effort was sufficient for success in Ghana. Finally, participants were asked how happy they were so as to isolate subjective wellbeing more broadly.

Being exposed to the treatment survey did not significantly shift respondents' preferences in terms of the subjective importance they attached to the outcome topics. That is, being primed and encouraged to think about the international character of development did not necessarily motivate participants to change their perceptions of policy accountability. This outcome conflicts with a mature discussion in the behavioral economics of development, which argues that financial incentives significantly affect economic preference rankings. However, this may not be the case with respect to

preferences for development, and more research is needed.

The international development financial incentives significantly affected citizen participation in the policy discussions. Respondents who were asked about the international aspects of Ghanaian development were significantly more likely than the control group to discuss and engage with all localized topics. Mobile surveys on global topics may therefore encourage participation in local policy discussions that are objectively important but receive little engagement.

Another innovative aspect of the study is the estimation approach used. Although most impact evaluations use Ordinary Least Squares (OLS) to understand causal impacts, these rely on linearity and additivity assumptions that might become less salient as data size grows with mobile use. In addition to OLS, the paper used statistical learning methods that solve standard regression problems without requiring strong parametric assumptions. Since mobile data is often large and complex in structure (relative to standard datasets that emerge from traditional evaluations), such approaches may become more important in future evaluations for policy.

Such mobile survey evaluations can help policy makers understand subjective issues that have eluded quantitative evaluations and been the preserve of qualitative research. Topics relevant for indigenous populations and slum dwellers for example, may often draw on subjective phenomena such as violence, which have been absent from most evaluations but may be important in the era of sustainable development. Mobile survey experiments are not meant to replace standard evaluations, but they can help complement existing approaches based on in-person surveys. Standardization remains critical so that researchers can use multiple data sources of various formats, and although many policy experts are trained in economics traditionally, knowledge in computer science is likely to be complementary for such evaluations in many cases. New partnerships with

the private sector should harness mobile network operators to support innovative forms of impact evaluations.

It is worth remembering that many people still do not have access to mobile phones. In Ghana, the recent removal of tariffs on smartphones may portend improved access. When participants have limited access and must own phones with others, researchers can allow participants to answer mobile surveys at times that are convenient for them. The distribution in Figure 1 implies that this is important. Since incoming calls are free of charge, study subjects can flash into the study service and be recalled automatically with the necessary content and resume where they left off during the survey.

Going forward, safeguarding the privacy of study participants' responses will be a key issue. This study did not require the actual phone numbers of study subjects and relied on encryption to protect their identities. In most cases, such information is usually not significant and codes may be preserved to create panel datasets while protecting the identity and confidentiality of study participants in the future. Another topic that is likely to grow in importance is the need for open data structures in machine-readable formats while further easing the access of developing country researchers to data in their own countries. Mobile surveys may be merged with open data sources for impact evaluations to understand how natural experiments affect the wellbeing of citizens over time for example.

Although poverty and inequality are not necessarily limited to developing country environments,

impact evaluations are only beginning to receive the policy attention they warrant in wealthier nations. Since the sustainable development agenda is global, there is an exciting possibility for organizations based in developing countries to conduct future research in developed nations for comparative analyses of preferences using mobile impact evaluations. In a similar vein, mobile survey micro-experiments could easily analyze the preferences of minority groups with respect to various policy concerns. Also, the anonymity provided by mobile survey impact evaluations can allow policymakers to address controversial issues that are rarely discussed. For example, illegal international migration from West Africa to Europe through North Africa is a topic that would be difficult to study via in-person surveys and impact evaluations given the context spanning several countries and sub-regions. On the other hand, mobile survey experiments of the kind described here could draw attention to why citizens are taking such serious risks and how the various underlying problems may be addressed relatively easily. Much more research is needed to advance this new approach to impact evaluations. By harnessing the ability of phones to gauge local preferences and their intersection with global issues, relatively affordable, efficient mobile survey impact evaluations may help policy makers implement inclusive policies for sustainable development.

References

- Opoku-Agyemang, Kweku, 2015. "Preference Gravitation: A Theory with Evidence from 'Hands-Free' Mobile Survey Micro-Experiments." Working Paper.

ABOUT THE AUTHOR

Kweku Opoku-Agyemang holds a B.A. from the University of Ghana, an M.A. from Ohio University, and a Ph.D. from the University of Wisconsin-Madison. He is currently a Development Impact Lab Postdoctoral Fellow at the Blum Center for Developing Economies at the University of California, Berkeley where his research focuses on political economy and technological change. He acknowledges the support from the Development Impact Lab (USAID Cooperative Agreements AID-OAA-A-13-00002 and AID-OAA-A-12-00011) and the Blum Center for Developing Economies, University of California at Berkeley.



Debazou Y. Yantio

Investing in Development Impact Evaluations: Towards a Decision Model

Abstract: In 2006, the Center for Global Development (2006¹) observed that “billions of dollars were spent each year on thousands of programs to improve health, education and other social sector outcomes in the developing world, with almost no studies that could determine whether or not they actually made a difference. This absence of evidence not only wastes money but denies poor people crucial support to improve their lives”. Since the “Call for Action” that followed and the subsequent inception of 3ie in 2008, there has been “increased pressure on funding organizations that their activities be based on ‘rigorous’ evidence about ‘what works’”. Since then there has been a surge in investments in rigorous impact evaluations².

Millions in USD are now committed to impact evaluations, diverting a corresponding amount of funds from development programs to improve social sector outcomes. In the face of shrinking budgets for development, what is the social return on these evaluation investments? Is “Ignorance more expensive than impact evaluations?” Likewise, Gertler et al. (2011) advises that “[...] the cost of conducting an impact evaluation must be compared to the opportunity costs of not conducting a rigorous evaluation and thus potentially running an ineffective program”. “Is the value of the outcomes as aggregated social benefits worth the costs, compared to other available projects?” (Pritchett et al. 2012).

The question is one of investment decision making. To answer it, one needs to explicitly and systematically define an adequate “theory of change”, that is, a complete, coherent, and correct causal model from funding to inputs and activities to outputs to outcomes and impacts.

This paper reviews cited objectives and expected outcomes for carrying out impact evaluations, in order to identify possible benefits and common cost categories to consider in the valuation exercise. It then goes on to review impact evaluation valuation methods and finds that existing ones are very restricted in their simple assumptions, or incompletely cover the full range of benefits and cost, or that knowledge is not available. While expecting a more comprehensive method from further academic work, this paper ends with recommendations to evaluators on designing impact evaluations and writing proposal to facilitate investment decision making by donors and ensure more effective impact evaluations in terms of social welfare gains.

1 CGD, 2006, When Will We Ever Learn? Improving Lives Through Impact Evaluation, Report of the Evaluation Gap Working Group, May 2006, Center for Global Development

2 It is important to note that not all rigorous evaluations use RCTs nor are all RCTs actually ‘evaluations’ of actual projects. That is, many of the current RCTs are ‘field experiments’ that are designed and implemented by researchers for the purposes of research on techniques rather than evaluations of actual development projects

Introduction

The development community has committed huge amounts of resources to programs and is willing to continue despite downsizing in developed countries. According to the Organization for Economic Co-operation and Development (OECD 2015³), net official development assistance (ODA)⁴ amounted⁵ to 97.677 billion US dollars in 2004. Overseas development assistance (ODA) from OECD countries represented 20% of external financial flows of developing countries in 2006 compared to 60% during the 1970s (Frot et Santiso, 2008⁶). Funding of development programs by recipient developing countries themselves and foundations do not account for those figures. This investment is expected to improve development outcomes for better lives of billions of people in developing countries. However, the Centre for Global Development (CGD) Evaluation Gap Group established in 2004 observed that there were “almost no studies that could determine whether or not they actually made a difference. Evaluations of impact on outcomes typically lacked any coherent counter-factual for evaluating the causal impact of project outputs on the outcomes for intended beneficiaries. This absence of evidence not only wastes money but denies poor people crucial support to improve their lives (CGD 2006).

The creation of the International Initiative for Impact Evaluation (3ie)⁷ in 2008 resulted in “a massive rise in the use of Rigorous Impact Evaluation (RIE) techniques, including Randomized Control Trials (RCTs)⁸ and increased pressure on funding organizations that their activities be based on ‘rigorous’ evidence about ‘what works’ (Pritchett et al. 2012:2). Since its inception, 3ie spending has ranged from about 1 million USD in 2009, to a peak amount of 13.2 million USD in 2013, to a slight decline since 2014. Between 2008 and 2015, cumulative expenditure stood at 58 million USD, but commitments at about 88.5 million USD (3ie 2015)⁹. According to 3ie (2015), 12.9 million USD was spent on IE in 2014, that is only 0.01% of the 134,382 million USD of net ODA disbursed in the same year (OECD 2015). As of December 2014, a total of 146 impact evaluations in several sectors had been funded since 2008 (3ie 2014:25¹⁰). The corresponding expenses has increased 12-13 fold in only 6 years of operation.

As evidenced by the figures above and the about-to-be adopted Global Evaluation Agenda of the Sustainable Development Goals (SDG) of the United Nations, the market demand for impact evaluations of development interventions is growing, not only from governments and government departments, but also from civil society and the NGO communities (Conlin and Stirrat 2008:200). In addition to 3ie achievements that are only part of the global picture, Stéphanie Pamies-Sumner (2014¹¹) reports that more initiatives to support the production of impact

3 OECD (2015), Net ODA (indicator). doi: 10.1787/33346549-en (Accessed on 02 September 2015).

4 Official development assistance (ODA) is defined as government aid designed to promote the economic development and welfare of developing countries. Loans and credits for military purposes are excluded. Aid may be provided bilaterally, from donor to recipient, or channeled through a multilateral development agency such as the United Nations or the World Bank. Aid includes grants, “soft” loans (where the grant element is at least 25% of the total) and the provision of technical assistance. The OECD maintains a list of developing countries and territories; only aid to these countries counts as ODA. The list is periodically updated and currently contains over 150 countries or territories with per capita incomes below USD 12 276 in 2010.

5 The major source of international development funding in addition to foundation grants and recipient countries domestic investments.

6 Frot, Emmanuel, and Santiso, Javier. 2008. “Development Aid and Portfolio Funds: Trends, Volatility and Fragmentation” (December 1, 2008). OECD Working Paper No. 275. Available at SSRN: <http://ssrn.com/abstract=1277885> or <http://dx.doi.org/10.2139/ssrn.1277885>.

7 In order to respond to the CGD call for “more and better” impact evaluations, the Development Assistance Committee (DAC) Evaluation Network of the OECD joined with Evaluation Capacity Group (ECG) and United Nations Evaluation Group (UNEG) in a ‘Network of Networks on Impact Evaluation (NONIE)’.

8 It is important to note that not all rigorous evaluations use RCTs nor are all RCTs actually ‘evaluations’ of actual projects. That is, many of the current RCTs are ‘field experiments’ that are designed and implemented by researchers for the purposes of research on techniques rather than evaluations of actual development projects.

9 Access on September 2, 2015 at <http://www.3ieimpact.org/en/about/performance-metrics/>

10 International Initiative for Impact Evaluation, 2014, Annual Report, New Delhi, India.

11 Stéphanie Pamies-Sumner, 2014, *Les évaluations d’impact dans le domaine du développement : état des lieux et nouveaux enjeux*, Collection À Savoir n°27, Département de la recherche, Agence Française de Développement.

evaluations are ongoing or starting in several development agencies in all regions of the globe. The World Bank manages the Spanish Impact Evaluation Fund (SPIEF) that was created in 2007 with 14 million USD and 1.5 million USD grants from the Government of Spain and the United Kingdom Department for International Development (DFID), respectively. Fifty-one impact evaluations in developing countries were funded under this window between 2007 and 2012, along with numerous regional trainings related to impact evaluation. Terminated in 2012, its successor was the Strategic Impact Evaluation Fund (SPIEF)– with a DFID grant of about 40 million USD, of which 25 million USD was earmarked to fund 50 impact evaluations over the period 2012-2017.

The Development Impact Evaluation Initiative (DIME) was also put in place in 2005. The Africa Impact Evaluation Initiative (AIM) was established in 2005 to improve the quality of the World Bank's operations and strengthen evidence-based policy in government agencies. In the Human Development Network (HDN) a

series of Bank Netherlands Partnership Program (BNPP)-financed impact evaluations were developed to generate knowledge on specific development interventions. By 2008, HDN had well established program areas and a large number of ongoing evaluations. In 2010, the United States Agency for International Development (USAID) initiated the Development Innovation Venture (DIV) to support experimentation of innovative approaches in developing countries. It can finance up to 24 million every year. This limit was raised to 28 million USD in 2014. About 60 initiatives are funded in 8 sectors in 22 countries. The Global Development Innovation Ventures (GDIV) was set up in 2013 by USAID and DFID to replace the DIV in the framework of the National Impact Initiative. The Inter-American Development Bank (IADB) also has a portfolio of impact evaluations.

The 3ie impact evaluation repository has 2,657 entries as of 2 September 2015¹², including non-3ie-funded evaluations. At the World Bank Group, the production of IEs has also grown rapidly, from an average of 16 initiated per year in

Table 1: Average cost, or cost range of an impact evaluation study

Bilateral and Multilateral Donor Organization	Average cost, or cost range of an impact evaluation study
The World Bank	500,000 USD ^{a, b}
Millennium Challenge Corporation /USAID	1 – 2 million USD ^c
KfW	200,000 Euros
Japan International Cooperation Agency (JICA)	100,000 – 1 million USD
Agence Française de Développement (AFD)	35,000 – 800,000
Islamic Development Bank (ISD)	43,000 USD
SIEF	300,000 USD
3ie	250,000 USD
DIV	300,000 USD
DIV	75,000 – 1,5 million USD

a It represents 1.4% of the average cost of projects funded by the World Bank (IEG 2012)

b Based on another sample of World Bank programs in the Social Protection Sector and Labor Unit, Gertler et al. (2011:161) provide higher figures: average total cost of IE (968 750 USD); IE cost range (130 000 – 2 million USD); average % of IE Cost to Program Cost (4.5); range of the % of IE Cost to Program Cost (0.2 – 13.3)

c These agencies completely out source their impact evaluation projects while the World Bank only outsource data collection as reported by Pamies-Sumner (2014).

Source: Pamies-Sumner (2014)

12. Accesses on September 2, 2015 at <http://www.3ieimpact.org/en/evidence/impact-evaluations/>

the period 1999–2004 to an average of 62 per year in 2005–10 (IEG 2012¹³). In the same institution, Legovini (2010:6¹⁴) reported that “in 2004, there were 28 ongoing impact evaluations, and a total of 159 completed impact evaluations. By 2008, that number had grown seven-fold, as a result of DIME and two concomitant efforts in the Africa region and in the Human Development Network. Between 2004 and 2008, the Africa region evaluation portfolio grew forty times over”. The investment in impact evaluations represents a sizeable amount of development finance, and it is growing, despite shrinking development budgets.

The Problem

In this paper, we concur with Pritchett et al. (2012:3) that [impact] evaluation of a development project is itself a development project.¹⁵ The increasing commitment and outlays of development finance to impact evaluations is equivalent to less money available to other development projects that deliver more traditional social products and services to the needy target groups—this can be perceived as a potential problem. Finding a solution will require answers to the following questions: Did the surge in impact evaluations result in increased social welfare gains in the target communities? Is it true that ignorance is more expensive than impact evaluations as CGD (2006) claims? If that is true, to what extent should one invest in more and better impact evaluations? As Pritchett et al. (2012:7) put it: Is the (evaluation) project an effective use of resources for its intended purposes? Is the scope of the (evaluation) findings sufficient to justify time and cost of evaluation? Did evaluation results have an impact on beliefs

or behaviours of key actors? Hence, the central question is *whether the value of the outcomes [evaluation projects] as aggregated social benefits was worth the costs, compared to other available projects*. In a nutshell, it is all about the value of evaluations. As Barr and Rinnert (2015¹⁶) put it, “It is timely to focus on making a value-based case for evaluation, in this Year of Evaluation”. How then do we systematically assess the value of evaluations, particularly impact evaluations?

According to Weyrauch and Díaz Langou (2011:44), there are both conceptual and technical challenges in measuring the impact of evaluations on policy change. Newman and Evans (2014¹⁷) have reviewed the literature for methods to calculate the rate of return to investment in research. Currently, there is no systematic approach to valuing evaluations (Barr and Rinnert 2015). However, action needs to be taken now because “those who do research with important social benefits may lose out in the increasingly competitive battle for research funds (Smith 2001:1¹⁸).” He reports “pilots by the Dutch to produce a fully worked out method for assessing the social impact of research”. Barr and Rinnert (2015) used ex-post and cost benefit analysis with various scenarios (distributional weights/counterfactuals) to estimate the net present value of a randomized control trial of a social cash transfer programme in Zambia. Todd and Wolpin (2006¹⁹) have used behavioural methods for ex-ante evaluation of the impacts of social programs, using examples of wage subsidy programs, conditional cash transfer programs, and income support programs. They argue that ex-ante evaluation “is useful for designing programs that achieve some optimality criteria, such as maximizing impact for a given cost”. It

13 IEG. 2012. World Bank Group Impact Evaluations: Relevance and Effectiveness. Washington, DC: World Bank. doi: 10.1596/978-0-8213-9717-6. License: Creative Commons Attribution CC BY 3.0

14 Arianna Legovini. 2010. Development Impact Evaluation Initiative: A World Bank Strategic Approach to Enhance Developmental Effectiveness¹.

15 “Evaluation uses funds to finance inputs and activities (collection of data, analysis of data) that produce outputs (reports, research papers, policy advocacy) by an implementing agency (in this case an evaluation organization) with the ultimate intention of producing better developmental outcomes for intended beneficiaries.” (Pritchett 2012).

16 Julian Barr and David Rinnert. 2015. Maximising the Value of Evaluations: A Conceptual Framework, DFID Year of Evaluation Event, 18th June 2015, London.

17 Kirsty Newman and Will Evans. 2014. What is the evidence on the impact of research on international development? DFID Working Paper, Evidence into Action, Research and Evidence Division.

18 Richard Smith (editor). 2001. “Measuring the social impact of research: Difficult but necessary”, BMJ Volume 323 8 September 2001.

19 Todd, Petra and Wolpin, Kenneth I., Ex Ante Evaluation of Social Programs (May 2, 2006). PIER Working Paper No. 06-022. Available at SSRN: <http://ssrn.com/abstract=931393> or <http://dx.doi.org/10.2139/ssrn.931393>.

is clearly obvious that most impact evaluations are not designed to achieve an optimality criteria. Duflo (2002²⁰) affirms that the marginal rate of productivity of public funds is very far from being optimized. Can we retain this optimality criteria for the appraisal of impact evaluation projects? Picciotto (1999:9²¹) suggests a measure of the contribution of evaluation to society: the incremental value of actual outcomes compared to the ‘counterfactual’ which would have materialized in the absence of evaluation”.

Towards a Model of Benefits and Costs of Impact Evaluations

According to Essama-Nssah (2013²²), an evaluation question is a request for information by decision makers and other stakeholders about the performance or results of an intervention. Evaluation can be viewed as an inferential process designed to produce credible evidence to answer a set of well-posed policy questions (Gertler et al. 2011²³). Effective evaluation can produce reliable information on what works, what does not, and why. Program evaluation is “the application of systematic research methods to the assessment of program design, implementation and effectiveness (Chelimsky 1985²⁴)”. Thus, it is of the nature of a research project. “In the end, evaluation is of little more than academic interest unless it feeds into policy-making and has a positive impact on the development process” (Conlin and Stirrat 2008:3²⁵). Impact evaluations are more than research that only seeks to develop the knowledge base; they aim to influence, directly or indirectly, public

policy-making by informing the decisions of donors and policy makers in developing countries with evidence. Policy makers may use this information to modify or cancel ineffective programs, or design new ones and thus make the most of limited resources (Grossman 1994²⁶). In addition, a frequent aim of impact evaluations is to contribute to the development, adoption and amendment of policy (Weyrauch and Díaz Langou 2011:1). According to Duflo (2002), clean estimation of particular components of policy is difficult in the real world because most policies come as packages and packages are not necessarily well-motivated, and costs/benefits of various policies is widely different from one program to the next.

As in most development interventions, I believe this also applies to impact evaluations. This section of the paper attempts to identify generic benefits (outputs and outcomes/impacts) and costs (inputs) of any impact evaluation project. Systematic approaches and methods to value those benefits and costs will then be identified in the literature to pave the way for a generic appraisal model. In this regard, Barr & Rinnert (2015) distinguish between two approaches to valuing evaluations: ex-ante and ex-post. The ex-ante valuation is for decision-making and the ex-post is for accountability and learning.

Estimating the Benefits of an Impact Evaluation Project

For LSE (2011²⁷), “a research impact is a recorded or otherwise auditable occasion of influence from academic research on another actor or organization: academic impacts, and external impacts. Academic impacts from research are influences upon actors in academia or universities, e.g. as measured by citations in other academic authors’ work. External impacts are influences on actors outside higher education, that is, in business, government or civil society, e.g. as measured by references in the trade press or in government

20 Duflo, Esther, 2002, Policy Evaluation IIE Anniversary Symposium.

21 Picciotto, R. (1999), 'Towards an economics of evaluation', *Evaluation* 5(1): 7–22.

22 B. Essama-Nssah. 2013. What Questions Do Evaluations Answer? PREM Notes N°25 Special Series on notes The Nuts & Bolts of M&E Systems.

23 Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: World Bank

24 Chelimsky, E. (ed.) (1985) *Program Evaluation: Patterns and Directions*. Washington, DC: American Society for Public Administration.

25 Sean Conlin and Roderick L. Stirrat 2008. “Current Challenges in Development Evaluation”, *Evaluation*, SAGE Publications (Los Angeles, London, New Delhi and Singapore) DOI: 10.1177/1356389007087539 Vol 14(2): 193 – 208.

26 Grossman, Jean Baldwin. 1994. “Evaluating Social Policies: Principles and U.S. Experience.” *World Bank Research Observer* 9 (2): 159–80.

27 LSE. “Maximizing The Impacts Of Your Research: A Handbook For Social Scientists” Consultation Draft 3: LSE Public Policy Group.

documents, or by coverage in mass media. It is not the same thing as a change in outputs or activities as a result of that influence, still less a change in social outcomes.” Citing Carol Weiss, Weyrauch and Díaz Langou (2011:4) underscore that social science research has traditionally had the objective “to better understand reality” and more recently “to influence public-policy”. According to Court et al. (2004²⁸), a better use of research-based evidence in development policy and practice can help save lives, reduce poverty and improve the quality of life. Concerning the DIME program, Legovini (2010:64) reports that data has value beyond the purpose for which it was originally collected—“repurposing” of data, hence improving returns on impact evaluation investments.

However, it is challenging, yet not infeasible, to establish verified causal links from one impact evaluation to output changes or to social outcomes (social welfare gain), given the current state of knowledge. The conclusion that “Improving our knowledge of primary impacts as occasions of influence is the best route to expanding what can be achieved here” will be followed here but along with tracing what is called secondary impacts i.e. social welfare gain beneficial to society in a way (LSE 2011; Scoble et al. 2009²⁹). In order to

define the objectives towards secondary impacts of policy influence interventions, Weyrauch lists three categories: (i) capacity building; (ii) political and policy change with three subtypes: Expanding policy capacity, broadening policy horizons, and affecting policy regimes; and (iii) Citizen engagement. Reisman, Gienapp, and Stachowiak (2007:18-20³⁰) provide a richer list of six outcome categories: shift in social norms; strengthened organizational capacity; strengthened alliances; strengthened base of support; improved policies; and changes in impact.

Based on the objectives of impact evaluations as stated in the policy documents of various sponsors, expected benefits can be substantiated. According to Legovini (2010), the DIME program has three objectives: improve the quality of the World Bank’s operations through iterative learning; strengthen client institutions for evidence-based policy-making; and generate knowledge on critical development questions. “Evaluation results may be used to assess a program’s effectiveness, identify how to improve performance, or guide resource allocation” (GAO 2012³¹). They are all specificities of what Picciotto (1999) sets out as the single purpose of evaluation: to help

Table 2: Impact Categories in Policy Influence

Classification	Categories	Definition
Impact within research sector (academic impact)	Knowledge	Explicit and codified knowledge Papers, books and book chapters can be used as a proxy
	Impacts on future research	Generation of new research questions; development of new methods and/or datasets; capacity building; career development.
Impact outside sector (non-academic impact)	Impacts on policy	Effects of research on policy at many levels, for example: national policy; the policy of professional bodies; the policies of departments of organisations Includes effects on the ability, and propensity, of policy makers to use research.
	Impacts on practice	Effects on individual behaviour, which may or may not be in line with the policies of the organisation, or group to which the individual belongs.
	Wider social and economic impacts	Social or economic effects that change society, including impacts on public opinion. Media coverage can be used as proxy for impact on public opinion.

Source: Scoble et al. (2009:10)

28 Court, J., Hovland, I. and Young, J. (2004) Bridging Research and Policy in International Development: Evidence and the Change Process, ITDG.

29 Rosa Scoble, Keith Dickson, Justin Fisher, and Stephen R Hanney. 2009. Research Impact Evaluation, a Wider Context: Findings from a Research Impact Pilot, Working Paper, Brunel Research Impact Device for Evaluation (BRIDE), Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

30 Jane Reisman, Anne Gienapp, Sarah Stachowiak. 2007. A Guide to Measuring Advocacy and Policy, Organizational Research Services, Annie E. Casey Foundation.

31 GAO. 2012. Designing Evaluations 2012 Revision, Applied Research and Methods, GAO-12-208G, United States Government Accountability Office.

organizations and individuals achieve their objectives, based on societal values and norms.

There are other ways to disaggregate evaluation goals that may serve to draw greater attention to the knowledge generation or ‘cognitive’ processes that evaluation is a part of. Berriet-Sollicie et al. (2011³²) suggest three categories: to learn: the evaluation is primarily designed as a collective learning process; to measure: the evaluation is designed to assess programme performance and impact; to understand: the evaluation identifies and analyses the mechanisms by which the programme under evaluation can produce the expected outcomes or may create adverse effects.

In attempts to develop a methodology for evaluating the impact of research on health care, Grant et al. (2000:1107-08³³) piloted an approach disaggregating the research process and assessing the “payback” at each stage. They found that it is possible to use applied bibliometric techniques to “link” research funding organizations with both primary (publications in the serial peer reviewed literature) and secondary (evidence-based clinical guidelines) outputs.

Various kinds of evaluation have different impacts on the costs associated with search and information, bargaining, decision, monitoring and enforcement. The reduction of such transaction costs attributable to evaluation – both in the short run and, through adaptive efficiency, in the longer run – represents the benefits of investment in evaluation. Such benefits must be weighed against the direct costs involved in setting up and operating the evaluation function (Picciotto 1999:20³⁴).

32 Berriet-Sollicie, M.; Labarthe, P.; Laurent, C.; Baudry, J. 2011. Empirical validity of the evaluation of public policies: models of evaluation and quality of evidence; CESAER, Working Paper 2011/4.

33 Jonathan Grant, Robert Cottrell, Françoise Cluzeau, Gail Fawcett. 2000. "Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study", BMJ Volume 320 22 April 2000.

34 Picciotto, R. (1999), 'Towards an economics of evaluation', *Evaluation* 5(1): 7-22

Estimating the Costs of an Impact Evaluation Project

To justify an impact evaluation, Gertler et al. (2011) propose that “the cost of conducting [it] must be compared to the opportunity costs of not conducting a rigorous evaluation and thus potentially running an ineffective program”. These authors present a detailed account of budgeting the preparation and implementation of an impact evaluation. “Budget items include staff fees for at least one principal investigator/ researcher, a research assistant, a field coordinator, a sampling expert, survey enumerators, and project staff, who may provide support throughout the evaluation. These human resources may consist of researchers and technical experts from international organizations, international or local consultants, and client country program staff. The costs of travel and subsistence (hotels and per diems) must also be budgeted. Resources for dissemination, often in the form of workshops, reports, and academic papers, should also be considered in the evaluation planning”.

Implications for Donors, Policymakers and Evaluators

In a context of policy making primarily dominated by politics, Duflo (2002) argues that knowing what works makes no difference. Although politics will remain part of the process, it is advocated that development policy making be informed by evidence. Research-based evidence, that is, rigorous impact evaluations can contribute to development policies and practices that have a dramatic impact on peoples’ lives, even though influencing public policy is difficult to attain (Weiss 1999³⁵). Despite the difficulties and the absence of a systematic and comprehensive approach to value impact evaluations at this

35 Carol H. Weiss. 1999. "The Interface between Evaluation and Public Policy," *Evaluation*

point in time, evaluation managers and teams should make the benefits and costs of their evaluations more explicit. In designing more effective [impact] evaluations, evaluators should apply the recommendation they themselves provide to their clients: develop “an adequate ‘theory of change’—a complete, coherent, and correct causal model from funding to inputs and activities to outputs to outcomes and impacts” (Pritchett et al. 2012:1³⁶). We argue that a well-designed impact evaluation should clarify what are its purposes: increase the knowledge base of what works and what doesn't or bring about “changes in organizational outputs and social outcomes”. In developing the influencing objectives of impact evaluations, keeping as SMART as

possible is key as well as following Young and Mendizabal's (2009³⁷) recommendation to focus on the type of change that is targeted: Discursive changes (changes in language usage); Procedural changes (changing how something is done); Content changes (actual changes in written policy); Attitudinal changes (changes in perception of key stakeholders); and/or behavioural changes (sustainable changes in the way something is achieved or approached).

Overall, the evaluation community still demands a generic, comprehensive and systematic method of computing the economic or social rate of return on impact evaluation investment to support decision making by funding organizations.

ABOUT THE AUTHOR

Debazou Y. Yantio is a policy and development evaluation specialist. He has been a consultant with Independent Development Evaluation at the AfDB since April 2014. Since 1994, he has participated or led program evaluation related missions in 19 countries across Africa, 6 in Europe, 3 in Asia, and 3 in North America. He has provided evaluation consulting services to major UN agencies and multilateral development institutions, and nongovernmental organizations. He is the President of the Cameroon Development Evaluation Association (CaDEA), and a Board Member of the International Development Evaluation Association (IDEAS) where he chairs the Thematic Interest Group on Impact Evaluation.

An Ingénieur agronome specialized in agricultural economics, he holds a M.A. in Economic Policy Management (2001), and attended the International Program for Development Evaluation Training (IPDET) at Carleton University (2003).

36 Lant Pritchett, Salimah Samji and Jeffrey Hammer. 2012. *It's All About MeE Using Structured Experiential Learning ('e') to Crawl the Design Space*.

37 John Young and Enrique Mendizabal. 2009. *Helping researchers become policy entrepreneurs: How to develop engagement strategies for evidence-based policy-making*, Briefing Paper n°53, ODI



B. Essama-Nssah

Accounting for Happiness in Policy Evaluation

- *The pursuit of happiness is emerging as the ultimate goal of public policy.*
- *The identification, implementation, and evaluation of public policies must be based on the extent to which they enhance the subjective wellbeing or happiness of citizens.*
- *Valuation, which is a core component of evaluation, is a two-step inference that involves asking a sample of people from the target population a set of questions about their preference or life satisfaction and using their answers to infer the social value of policy outcomes.*
- *The credibility of the conclusions and recommendations from this inferential process is seriously threatened by factors such as confounding, miscommunications and misrepresentation.*
- *In the end, accounting for happiness in policy evaluation is all about making sure to measure the right thing the right way at the individual level, and to aggregate individual values into an indicator of social state which is statistically meaningful and respects the prevailing social values and norms.*

Introduction

The pursuit of happiness is emerging as the ultimate goal of public policy. Policymakers should therefore strive to identify and implement interventions that have the greatest potential for enhancing citizens' wellbeing or happiness. This article discusses the implications of this trend for policy evaluation. It provides a brief review of the issues around the definition and measurement of happiness and the valuation of policy outcomes. Valuation is a two-step inferential process of asking a sample of people from a target population a set of questions about their preferences or satisfaction with their lives, and using the answers to infer the social value of policy outcomes. The credibility of this process can be undermined by factors such as miscommunication, misrepresentation and confounders (i.e. factors that might mask a causal relationship thus rendering its identification difficult). In the end, accounting for happiness in policy evaluation entails assessing individual advantages and social progress in terms of satisfaction judgments. It is a matter of making sure that the right thing is measured the right way at the individual level, and aggregating individual values into an indicator of social conditions that is statistically meaningful and also respects prevailing social values and norms.

Over the past three decades or so, governments worldwide have come under mounting pressure to adopt an *evidence-based* approach to decision-making to promote transparency and accountability in their policymaking, which is conducted by the bureaucracy in the name of society at large (Quade 1982). To be able to hold policymakers to account, citizens must understand the process of choosing policies and how the implementation of these policies will affect them. Policy evaluation has an important role to play here as a process designed to produce *evidence* to answer key policy or programmatic questions about which decision-makers and other key stakeholders care (Gertler et al. 2011). These questions reflect the information needs of decision-makers that may arise during the policy cycle. They focus the evaluation and determine the appropriate evaluation design and methods¹.

At the design stage, policymakers would like to know *the right thing* to do and *the right way* of doing it. This requires a sound understanding of the problem that the policy is supposed to resolve and of plausible strategies for solving it. The understanding comes from a *situation analysis* or needs assessment that explores the nature and extent of the problem, including its root causes and social consequences. The final policy to be implemented emerges from a selection process where feasible alternative options are compared on the basis of a given criterion.

Once a situation analysis has identified the target problem, the best course of action comes from assessing feasible alternatives following these logical steps (Smith and Larimer 2013): (1) Specify the criterion for evaluation; (2) Identify and describe *alternative* policy interventions; (3) Estimate the likely *consequences* of each alternative; (4) Determine the value of each alternative on the basis of its consequences and the selected criterion; (5) Rank alternatives on the basis of these values. The best alternative is the most desirable according to the chosen criterion.

Following implementation, those with a significant stake in the policy would like to know whether and the extent to which what was supposed to happen did, in fact, happen. Answering this question involves (i) monitoring what happened, (ii) comparing it with what was supposed to happen, and (iii) explaining significant discrepancies between the two (Frechtling 2007). The performance measurement system

¹ See Essama-Nssah (2013a) for framing evaluation within the logic of asking and answering a set of questions representing a request for information by decision-makers and other stakeholders about the performance or results of an intervention.

provides information on what is happening, while the intervention theory describes what was supposed to happen. Such a theory has two components: a *theory of change* and a *theory of action*: The former describes the causal processes through which a policy intervention is supposed to work to solve the target problem and the latter is a set of prescriptive assumptions about the requirements for those causal processes to occur.

Public policymaking is driven by a social goal. United Nations resolution 65/309 considers the pursuit of happiness to be a fundamental human goal and invites member countries to develop “additional measures to better capture the importance of the pursuit of happiness and wellbeing in development with a view to guiding their public policies.” In this perspective, economic growth has an instrumental value to the extent that it can provide resources to support the pursuit of happiness². Thus, good government requires information on happiness and wellbeing for the *formulation and evaluation* of policy interventions (Diener et al., 2009). In fact, the Kingdom of Bhutan measures social progress not in terms of gross domestic product but in terms of its citizens’ happiness.

Amartya Sen (1999) argues that any evaluation approach is characterized by its *informational basis*, or the information needed to pass evaluative judgments. The informational basis of an evaluation is therefore the information needed to answer key evaluation questions. Accounting for happiness in policy evaluation has implications for the informational basis. It is bound to influence the specification of the criterion for evaluation and hence the valuation of policy consequences and the ranking of policy options. There is an important distinction to be made between the immediate policy objective stating intended outcomes in terms of a resolution of the target problem

and used to estimate the consequences of an intervention and the ultimate social goal used to value such consequences from the societal point of view.

This essay discusses the implications for evaluation of adopting a *happiness-based approach* to policymaking. It reviews some conceptions of happiness and their implications for measurement and highlights the informational challenges associated with the approach.

Understanding Happiness

If the pursuit of happiness is to guide public policy, then it is important to have a clear understanding of the concept and how it can be measured. There is no consensus about the meaning of “happiness”. The colloquial term means *subjective wellbeing*. Frey (2010) proposes several basic connotations that suggest that wellbeing is a complex and multidimensional concept: (i) momentary feelings of joy and pleasure also known as a *positive affect*, (ii) overall contentment with life or *life satisfaction*, and (iii) the *good life* achieved by developing and fulfilling one’s potential. The potential constituents of individual wellbeing can be classified in a variety of ways. The most common identifies two basic aspects: *affective* and *cognitive*. The affective component reflects feelings or emotional responses to life events and experiences and includes both positive affect based on positive emotions such as joy or pride, and negative affect based on negative emotions such as pain, anxiety or anger (Stiglitz et al. 2010). The cognitive component relates to how one assesses one’s life. The very definition of wellbeing entails a value judgment about life dimensions that are worth promoting. In other words, defining wellbeing is essentially an evaluative exercise. Thus, Diener et al. (2009) argue that considering an individual’s wellbeing involves “an overall evaluation of an individual’s life in all its aspects.”

² Similarly, the 2008 Commission on Growth and Development reports that growth is not an end in itself, but creates opportunities for the achievement of other important things that individuals and society care about.

The perspective from which lives are being evaluated leads to an important distinction between *objective* and *subjective* definitions of wellbeing. Objective definitions are based on considerations that are totally independent of an individual's own subjective judgment. They focus on aspects of wellbeing that would be considered desirable (e.g. physical health and literacy) whether or not those who experience them agree with that assessment. Subjective definitions are based on the individual's own interests, needs, desires or preferences. In this perspective, a life is going well only if the person living it is satisfied with the way that it is going (Diener et al. 2009).

The capability approach to wellbeing emphasizes the role of *individual agency* in the pursuit of individual well-being. The “agency aspect” of a person focuses more on the individual as a doer whose actions bring about change and whose achievements are assessed in terms of her or his own values and objectives (Sen 1999). The use of individual values and objectives to assess a person's achievements means that wellbeing is taken to be subjective.

The approach also brings out the deep connection between individual agency and social arrangements. Indeed, what individuals can achieve is necessarily *qualified* and *constrained* by the social, political, and economic opportunities available to them. Five interconnected types of freedom considered to be instrumental in promoting free agency: (1) *political freedoms* (civil rights); (2) *economic opportunities*; (3) *social facilities* (provision of social services such as education and health care); (4) *transparency guarantees* (to promote trust), and (5) *protective security* (provision of social safety net).

Accordingly, the task of public policy is to remove major sources of unfreedom such as poverty, limited economic opportunities, neglect of public facilities, discrimination, and political repression. The identification of these problems calls for

an analytical approach integrating economic, social, and political considerations. The centrality of individual freedom in this approach implies that policy effectiveness depends critically on free human agency. Finally, *the effectiveness of interventions must be judged in terms of their contribution to the enhancement and guarantee of individual agency.*

Consistent with the capability approach, Stiglitz et al. (2010) note that subjective wellbeing (SWB) is determined by objective circumstances and capabilities and consider health, education, personal activities, social connections, political voice and insecurity relevant predictors of life satisfaction.

Indicators

The measurement of subjective wellbeing is of interest in policymaking because the measurements can be used to monitor progress, inform policy design, and assess the social value of policy outcomes. However, measuring subjective wellbeing makes sense only to the extent that it can change significantly as a result of changes in the socioeconomic environment induced by shocks or policy implementation. To what extent, then, is it responsive to such changes? There is evidence showing that genetic factors tend to make SWB stable over time, which has led to fears that subjective wellbeing may not be malleable. However, it is now known that SWB is not static and can be altered over long periods of time (Tay et al. 2015). In particular, research shows that SWB is sensitive to economic factors such as income, wealth and employment, and a good predictor of important outcomes such as health and longevity, social relationships, productivity at work and success in one's career.

If wellbeing is fundamentally considered to be a subjective construct, then objective indicators can only be measures of causes, effects, or simply correlates of wellbeing (Allin and Hand 2014)

rather than measures of wellbeing itself. Diener et al. (2009) discuss three basic approaches to measuring SWB involving both objective and subjective indicators. These approaches respectively base the measurement of SWB on: (1) Preference realization (2) Affective response, and (3) Evaluative judgments of life. The corresponding indicators of wellbeing are: *income*, *affect*, and *life satisfaction*.

While the definition of a concept determines its measurement, it is important to note that objective and subjective definitions of wellbeing are not the same as objective and subjective indicators of wellbeing. An objective indicator of wellbeing is external to the individual and believed to have an impact on his or her wellbeing or to be indicative of a state of wellbeing. Thus objective measures of a subjective phenomenon (or an internal attribute) are necessarily model-based to the extent that they rely on assumptions about the relationship between the measure and the target of measurement (Allin and Hand 2014).

Disposable income and wealth are objective indicators of SWB based on economic theory linking money with the subjective construct of wellbeing. Socioeconomic agents are supposed to behave optimally and interact through a set of competitive, complete markets subject to the prevailing ownership of resources (Dixit 1996). In this setting, available options are limited by individual purchasing power. Lindblom (2001) explains that each person's claims to available goods and services are limited by the amount of income obtainable by that person's successful sale of something of value on the market. He concludes that market interaction is governed by the rule of *quid pro quo*. The wealthier the individual the wider the range of options. Thus, income and wealth can be used as indicators of wellbeing.

The validity of an indicator depends on the strength of its link to the construct it purports

to reflect. The strength of the link between income and wellbeing depends on the validity of the underlying assumptions. Informational problems may limit the ability of consumers to make optimal choices. Market imperfections or failure means that not all preferences can be realized in the market place. Moreover, it is not easy to estimate household income validly and reliably, particularly for the self-employed and people engaged in the informal economy. These difficulties create some uncertainty around the relationship between income and wellbeing. It is therefore desirable to consider alternative indicators that might be used to validate this relationship. In fact, there is now abundant evidence that income is strongly related to and can have large effects on SWB (Tay et al. 2015). Interest in alternative indicators also stems from the possibility that they might provide valuable information beyond what income and wealth can provide.

Affect and life satisfaction are *subjective indicators* to the extent that they tap directly an individual's relevant internal characteristics. Affect represents good and bad feelings. It is believed that affective feelings can result from the match or mismatch between the actual and ideal conditions in an individual's life. A match would lead to positive emotions while a mismatch would induce negative feelings (Diener et al. 2009).

The second type of subjective indicators of wellbeing are known as *life satisfaction* or global happiness measures, which are based on questionnaires asking respondents to explicitly consider dimensions of their lives and evaluate them (Diener et al. 2009). The evaluation may indicate the extent to which a person is satisfied with life, or the extent to which the conditions match the ideals. There is a very tight link between these evaluative judgments and the concept of wellbeing, which is what makes the use of these indicators so appealing. However, biases can creep into the process of integrating

information about various aspects of an individual's life and determining whether they match ideals. Fortunately, there is mounting empirical evidence that subjective indicators are sufficiently valid. In other words, even though happiness scores measure SWB with some noise, available data show a sufficiently high signal-to-noise ratio (Di Tella and MacCulloch 2006).

Valuating Policy Outcomes

Public policy is made in the name of society as a whole. The consequences of alternative interventions must therefore be evaluated from this perspective. The logic of causal inference guides the determination of policy consequences. Basically, to infer the consequences of an intervention, an association must be established between the implementation of the policy and observed changes in the measure of the policy objective and all factors that could plausibly confound this association must be credibly ruled out (Essama-Nssah 2013b). Ranking policy alternatives on the basis of their social desirability hinges, critically, on how individuals value such alternatives. This creates a need for a way to combine or aggregate individual valuations of policy outcomes in a manner that reflects societal concerns. This section discusses these two inferential steps.

Individual Valuations

Policy outcomes are the effects on people of attaining the immediate objective of the intervention. The value of these effects is represented by the corresponding changes in SWB, which is not directly observable. *The whole point of valuation is therefore to estimate the worth of these effects to an individual on some numerical scale.* Ultimately, it's a matter of measuring changes in individual wellbeing induced by a policy intervention.

Most approaches to valuation take individual wellbeing as the source of both individual and social value. They differ mainly in the definition of the concept of wellbeing. The two most common definitions are *preference satisfaction* and *life satisfaction*. Robinson (2013) points out that the preference-based approach provides evidence to answer the question of how to best allocate scarce resources given an understanding of individuals' preferences, while the life satisfaction approach addresses how best to allocate these resources given an understanding of individuals' subjective wellbeing.

Basically, valuing policy outcomes at the individual level means *measuring* the value that each concerned person attaches to the policy outcome under consideration. Generally speaking, measurement is "the assignment of numerals to objects or events according to rules." (Stevens 1946) The preference satisfaction approach assigns numbers to policy outcomes according to the quid pro quo rule that governs market interaction. The life satisfaction approach assigns numbers to outcome according to how people think and feel about their lives.

The value of things resides within individuals and is therefore not directly observable. When policy outcomes involve goods and services that are traded in well-functioning markets, the preference satisfaction approach relies on revealed preference methods to infer willingness to pay (WTP) and willingness to accept (WTA) from observed market behavior. Most policy outcomes involve nonmarket goods. Both the preference and life satisfaction approaches therefore rely on sample surveys to infer individual valuations of policy outcomes. It is fundamentally a matter of asking questions of a sample of people drawn from a target population and using the answers to characterize that population (Fowler 2014). In particular, each respondent's answers to either

a preference satisfaction or a life satisfaction survey are used to infer the value attached to the outcome.

The value of a policy outcome inferred for an individual in the context of a preference satisfaction survey³ is a measure of her WTP if the effect is an advantage, or her WTA compensation if the policy effect is a burden. One key advantage of preference-based methods is that they value outcomes in terms of a common *money metric*. This allows the comparison of incommensurable things such as educational and health outcomes, purchasing power or community stability. This also makes it possible to compute net benefits as in the context of cost-benefit analysis (CBA). This computation entails the comparison of the sum of all monetized benefits to the sum of all monetized costs. Despite this advantage, it is increasingly believed that these money measures may not be good proxies for quality of life (Diener et al. 2009).

The life satisfaction approach seeks to tap subjective wellbeing directly by asking people how they think and feel about their lives as a whole. For instance, the World Value Survey asks (Adler 2013), “All things considered, how satisfied are you with your life as a whole these days?” Respondents are requested to answer on a scale ranging from 1 to 10 (1 = very dissatisfied; 10 = very satisfied.) The German Socio-Economic Panel asks a similar question but proposes a scale from 0 to 10 (0 = completely dissatisfied, and 10 = completely satisfied. In addition to life satisfaction scores, these types of surveys collect information on various characteristics including income and level of the outcome to be valued. These data are used to estimate a life satisfaction function linking the scores to individual characteristics. The ratio of the coefficient on the level of outcome in that function to the coefficient on

income is taken to measure the WTP or WTA (depending on the nature of the outcome).

Aggregation of Individual Values

The fact that public policy is made in the name of and for society as a whole means that evaluation must be conducted from the point of view of society as well. The final verdict in any policy evaluation must be based on a comparison of the social value of alternative options that stems from an aggregation of individual valuations. Any *aggregation* rule must deal with *meaningfulness* and *representation*. Meaningfulness has both statistical and social dimensions; an aggregate indicator of the value society places on policy outcomes is statistically meaningful if its computation involves statistical operations consistent with the measurement scale underlying the data. It is socially meaningful when it fully accounts for key societal concerns such as *efficiency*, *equity*, and *sustainability*.

The effects of an intervention on the target population depend on individual characteristics and circumstances. Some would therefore be better off and others worse off as a result of the intervention while others might conceivably have an unchanged situation. Aggregation entails weighing the gains of the winners against the losses of the losers. The social value of policy outcomes depends on the chosen weights. In the standard approach used in cost-benefit analysis, the aggregate WTP of the winners represents the social benefit of the intervention while the aggregate WTA of the losers is the social cost (Freeman III 2003). A situation where the aggregate WTP is greater than the aggregate WTA suggests that gainers could compensate the losers in such a way as to leave the former better off and the latter no worse off than in the counterfactual state. This would be considered a social improvement.

3 Such surveys are part of the so-called stated preference methods.

The problem is that, this criterion does not require actual compensation and focuses only on efficiency in the allocation of the resources involved in the intervention. It is therefore not *fully meaningful in the social sense* because it ignores equity, a fundamental consideration in policymaking. Individual wellbeing is certainly influenced by social interactions and external forces stemming from characteristics of the community or society at large. There is evidence that inequality in the distribution of resources across the population has a negative effect on perceived wellbeing. Furthermore, it is well known that the political feasibility of an intervention is determined by the way the political process balances the interests of the gainers against those of the losers. There could be a threshold at which a gain or a loss is so significant that an individual or a group feels compelled to organize and fight (Kanbur 1995).

If there is a social concern for equity (e.g. poverty reduction and shared prosperity) in addition to efficiency, then one should adopt a poverty-focused approach to policy evaluation. One possible way of accounting for distributional issues is to adopt a system of evaluative weights that respects the degree of inequality aversion of society (see Essama-Nssah 2004 and 2005). The basic idea is to combine individual values into an aggregate measure that fully respects the point of view of society. Such an aggregation rule is known as a social welfare function (or an indicator of collective wellbeing). This function defines a social evaluation criterion used to rank policy options. Thus, if *shared prosperity* is a social concern, then one may consider ranking policy options on the basis of a social welfare function that respects both efficiency and equity. Sustainability is another important social consideration that needs to be accounted for. However, the incorporation of this important concern into a social evaluation criterion presents significant challenges.

Validity Threats

Shadish et al. (2002) explain that validity is a characteristic of inferences. In particular, validity refers to the truth of an inference; a threat to validity is any factor that renders an inference either partially or totally false. Valuation is indeed a two-stage inference. The first stage infers individual values from the answers to the survey questions and the second stage uses the answers to infer social value. The credibility of this process hinges on three key considerations: (i) policy consequences must be accurately estimated; (ii) individual responses must accurately reflect the value of the outcome to respondents, and (iii) the subset of individuals participating in the survey must be representative of the target population. Credibility is therefore threatened by confounders and the inability to measure the right thing the right way, which ensures construct validity and reliability, hence *accuracy*.

A scale of measurement conveys information about the attribute under measurement. For an ordinal scale, it makes sense to state that one value is greater than another, but the magnitude of the differences is meaningless. Similarly, a cardinal or interval scale indicates that both levels and differences in levels of the attribute can be meaningfully compared. It follows that only ordinal statistics such as quantiles can be computed with data measured on an ordinal scale. In the case of a cardinal scale, in addition to statistics that are permissible under the ordinal scale, it is meaningful to compute averages. Thus, in order for aggregation of values across individuals to make sense statistically, the operations involved must be meaningful in terms of the underlying scale. In the context of subjective wellbeing surveys, the type of scale that respondents use to rate policy outcomes is not at all clear. Therefore, it is difficult to tell whether the values attached to individual responses have an ordinal or a cardinal interpretation (Kahneman and Krueger 2006). This, in turn, makes it hard to determine

the appropriate statistical operations to use in an aggregation rule.

There is evidence that factors such as *irrelevant cues* and *scope effects* might lead to biased responses (Dolan and Metcalfe 2008). Irrelevant cues arise when respondents are influenced by the elicitation procedure (i.e. the mechanism for requesting and collecting information from the respondents). Scope effects represent situations where responses are insensitive to the size of the outcomes being valued. In addition, factors such as framing issues, strategic behavior and non-response bias may threaten the credibility of the evidence produced by stated preference methods. It is therefore recommended that the validity and reliability of *stated preference results* be tested before they inform policy decisions.

One concern about the use of *measures of life satisfaction* stems from the fact that satisfaction judgments depend on standards that individuals set for themselves (Forgeard et al. 2011). Thus, individuals may find themselves in similar objective circumstances, and yet express different levels of satisfaction with their respective lives. Similarly, Adler (2013) argues that current practice is not based on clear normative concepts, making it difficult to determine the information content of life satisfaction scores. He further argues that the computation of the WTP and WTA on the basis of life satisfaction scores makes sense only if respondents fully understand the question as a request for information about their preference satisfaction and that they report their ratings on a cardinal scale. Under these circumstances, it would also make sense to use the mean of individual valuations as an indicator or an estimate of the social value of the outcome under consideration. *Miscommunication* occurs when the respondents misunderstand what is being asked or misrepresent the true value of her answer. This is a serious threat to the validity of the whole valuation process.

Conclusion

The pursuit of happiness is emerging as the ultimate goal of public policy. Policymakers should strive to identify and implement those interventions having the greatest potential to enhance the subjective wellbeing or happiness of citizens. As noted by Diener et al. (2009), measures of wellbeing are needed to *monitor* changes at individual and societal levels, to *compare* the relative situation of socioeconomic groups and to *evaluate* the impact of public policy on the population's overall quality of life.

The emphasis on measures of subjective wellbeing stems from the fact that they are meant to capture how people think and feel about their lives. This is important because two people can have the same levels of income and wealth and yet experience and evaluate their lives differently. This does not mean that observable indicators are not useful in policymaking. It is desirable that citizens “have both decent objective standards of living and feel subjectively satisfied with their lives” (Forgeard et al. 2011). Therefore, combining both *objective* and *subjective* indicators makes it possible for policymakers to capture a fuller picture of the quality of life of the target population. The validity of the conclusions and recommendations based on these indicators is seriously threatened by confounders, miscommunication and misrepresentation.

In the end, accounting for happiness in policy evaluation entails an assessment of individual advantages and social progress in terms of satisfaction judgments. It is a matter of making sure to measure the right thing the right way at the individual level and to aggregate individual values into an indicator of a social state that is statistically meaningful and respects prevailing social values and norms.

References

- Adler, Matthew D., 2013. "Happiness Surveys and Public Policy: What's the Use?" *Duke Law Journal* 62:8, pp. 1509-1601.
- Allin, Paul and David J. Hand, 2014. *The Wellbeing of Nations: Meaning, Motive and Measurement*. Chichester, UK: Wiley & Sons.
- Di Tella, Rafael and Robert MacCulloch, 2006. "Some Uses of Happiness Data in Economics," *Journal of Economic Perspectives*, 20: 1, pp. 25-46.
- Dixit, Avinash K., 1996. *The Making of Economic Policy: A Transaction Cost Politics Perspective*. Cambridge, MA: MIT Press.
- Dolan, Paul and Robert Metcalfe, 2008. "Comparing Willingness-To-Pay and Subjective Well-Being in the Context of Non-Market Goods," London School of Economics: Center for Economic Performance, Discussion Paper 890.
- Essama-Nssah, B., 2004. "Empowerment and Poverty-Focused Evaluation," *Development Southern Africa*, 21:3, pp. 509-530.
- Essama-Nssah, B., 2005. "A Unified Framework for Pro-Poor Growth Analysis," *Economics Letters* 89: 216-221.
- Essama-Nssah, B., 2013a. "What Questions Do Evaluations Answer?" PREM Note 25, Special Series on the Nuts and Bolts of M&E Systems, Washington, DC: The World Bank Group.
- Essama-Nssah B., 2013b. "Coping with the Attribution Problem in Program Evaluation," PREM Note 26, Special Series on the Nuts and Bolts of M&E Systems, Washington, DC: The World Bank Group.
- Forgeard, Marie J. C., Jayawickreme Eranda, Margaret L. Kern, and Martin E. P. Seligman, 2011. "Doing the Right Thing: Measuring Wellbeing for Public Policy," *International Journal of Wellbeing*, 1 (1): pp. 79-106.
- Fowler, Floyd J. Jr., 2014. *Survey Research Methods*. Thousand Oaks, CA.: Sage Publications.
- Frechtling, Joy A., 2007. *Logic Modeling Methods in Program Evaluation*. San Francisco: Jossey Bass.
- Freeman III, A. Myrick, 2003. "Economic Valuation: What and Why," in Patricia A. Champ, Kevin J. Boyle and Thomas C. Brown, Eds. *A Primer on Nonmarket Valuation*, New York: Springer Science and Business Media, LLC.
- Frey, Bruno S., 2010. *Happiness: A Revolution in Economics*. Cambridge, MA: MIT Press.
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings and Christel M.J. Vermeersch, 2011. *Impact Evaluation in Practice*. Washington, D.C.: The World Bank.
- Kahneman, Daniel and Alan B. Krueger, 2006. "Development in the Measurement of Subjective Well-Being," *Journal of Economic Perspectives*, 20:1, pp. 3-24.
- Kanbur, Ravi, 1995. "Welfare Economics, Political Economy and Political Reform in Ghana" in *African Development Review* 7 (1), pp. 35-49.
- Lindblom, C. E., 2001. *The Market System: What It Is, How It Works and What to Make of It*. New Haven, CT: Yale University Press.
- Quade, E.S., 1982. *Analysis for Public Decisions*. New York: North Holland.
- Robinson, Lisa A., 2013. "Cost-Benefit Analysis and Well-Being Analysis." *Duke Law Journal*, 62:8, pp. 1717-1734.
- Sen, Amartya Kumar, 1999. *Development as Freedom*. New York: Alfred A. Knopf.

- Shadish, William R., Thomas D. Cook and Donald T. Campbell, 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth.
- Smith, Kevin B. and Christopher W. Larimer, 2013. *The Public Policy Theory Primer*. Westview Press.
- Stevens, S. S., 1946. "On the Theory of Scales of Measurement," *Science* 103:2684, pp. 677-680.
- Stiglitz, Joseph E., Amartya Sen and Jean-Paul Fitoussi, 2010. *Mismeasuring Our lives: Why GDP doesn't Add Up*. New York: The New Press.
- Tay, Louis, Lauren Kuykendall and Ed Diener, 2015. "Satisfaction and Happiness—The Bright Side of Quality of Life," in Wolfgang Glatzer, Laura Camfield, Valerie Moller and Mariano Rojas Editors, *Global Handbook of Quality of Life*. Dordrecht: Springer Science and Business Media.

ABOUT THE AUTHOR

B. Essama-Nssah is a World Bank retiree currently working as an independent consultant on the analysis of the *social impact of public policy*. While at the World Bank, he performed economic analyses, prepared policy research and technical papers, and conducted an annual training course on policy impact analysis. Prior to joining the World Bank, he worked as a Senior Research Associate on the Food and Nutrition Program at Cornell University. Prior to that, he served as Vice Dean of the Faculty of Law and Economics and Head of the Economics Department at the University of Yaoundé, Cameroon. He also worked as Instructor, then Assistant Professor in the Department of Economics at Central Michigan University (USA), teaching Mathematics for economists, Statistics and Economics. He has published in the *Review of Income and Wealth*, *Economics Letters*, *Development Southern Africa*, *Journal of International Development*, *The Journal of African Economies*, *Research on Economic Inequality and the Handbook of Microsimulation Modelling*. He holds a Ph. D. in Economics from the University of Michigan, Ann Arbor.



Olivier Sossa

Ethics in Evaluation and a Global, Integrated Framework for Evaluating Development Policies

Since the year 2000, development aid has increased sharply and its efficacy has increasingly come under question from donors and beneficiaries alike. The many actors involved in development aid include international organizations, civil society, public officials, associations, labor unions, donors, and foundations. While it may be a good idea to have many actors engaged in development, their divergent and occasionally contradictory interests can be a hindrance. Moreover, the countries in greatest need of aid are most often politically unstable, lacking in infrastructure and experiencing high demographic growth. These challenges further complicate the approaches to building evaluation capacity and making the evaluations more useful. Innovative solutions are needed to meet these specific, special, and unique issues. The real challenge for evaluation in these countries is not only financial but also methodological. The challenge includes building the capacity to carry out evaluations focused on an analysis of change and complexity of the effects or impacts with a global and integrated evaluation framework.

This article addresses two issues linked to evaluation: making ethics a dimension of evaluative approaches (especially in developing countries) and the need for a global, integrated framework for evaluating development policies. Before addressing these issues, however, it is important to recall some of the characteristic imponderables of the context in order to better frame the pertinent development approaches.

The Context

In contrast with the rest of the world, demographic growth in several regions in Africa remains high. Some may consider this an advantage. However, with respect to the capacity of the countries to meet the basic needs of their populations, it raises significant issues for sustainable development. For example, will these governments be able to guarantee education for all and food and nutritional security for the population? How will they manage to provide decent employment for the 100 million youth who will be seeking work between 2013 and 2030?

It is clear that strong and unmanaged demographic growth in Africa, lack of infrastructure (in education, health, energy and transport) and political instability (notwithstanding the progress towards democracy achieved here and there) taken together, weigh heavily on the continent's development. In addition, the use of technologies is expanding rapidly in these regions and the populations are increasingly demanding – and legitimately so – more decent living conditions. This makes the context in which development policies are implemented and the quest for a coherent strategy for real development plans even more complex. Given this and the speed of change in certain sectors, this is an opportune moment to define innovative evaluation strategies and to conceive of a global, integrated approach that takes these issues and complexities into account. Evaluation practices should be oriented towards informing public policy and clearly identifying priority sectors in order to achieve veritable socio-economic development.

Ethics in Evaluation

The concept of ethics in policies that support evaluation approaches draws on social mediation in which ethics serve as a link between citizens and public authorities so as to encourage informed political decision making that is coherent with shared social priorities and values. The goal of taking ethics into account in evaluating development policies is to identify the ethical implications of political decisions and to help give voice to citizens' values and concerns regarding political choices. To this end, it is important to be transparent and provide information about policy orientations and their consequences and to organize public consultations about the population's priorities, concerns, and values.

Burls et al. (2011) point out that no single method fits all evaluation situations. Different methods make it possible to contextualize an ethical analysis to identify underlying values, reflect on implicit normative standards, carry out a social and historical analysis, make a formal identification and analysis of the ethos of the involved players, and various forms of popular civil participation. In addition, since the ethical stakes of development policies are intimately connected to the context in which they were developed and used, the approaches seeking to integrate ethics into evaluation

must make it possible to take that context into account.

Methodological approaches to the ethics of policy (including development policies) must take the context into consideration. They must also be able to capture the interactions among the elements of the complex systems, as well as their development over time and space, where ethical dilemmas emerge.

Identifying the difficulties facing people living in vulnerable situations can illuminate the ethical issues facing an entire population given that people find themselves in a vulnerable situation as soon as they have to interact with public services. Typically, those ethical issues that affect people living in vulnerable situations raise questions about such fundamental values such as fraternity, solidarity, justice, and so on.

For a Global and Integrated Approach

Several organizational analysis models direct and structure the analysis of change in program evaluation: the goal-oriented model, the human relations model, adaptation, internal processes, and so on. Each of these illustrates a different organizational perspective.

- The goal-oriented model considers that organizations exist to meet certain goals and objectives. It supposes a rational, instrumental organizational vision based on goals and objectives that are generally translated into a volume and quality of goods and services. In this perspective, the degree to which the organization manages to produce these goods and services is the emphasis. Guisset et al. (2002) points out that this view of an organization is criticized primarily because of the difficulty in making organizational objectives clear and in defining the most important goal among various possible goals.

- In the human relations model, an organization is considered to perform well if it manages to create and maintain a healthy work environment and harmonious relationships among the people working in it.
- The adaptation model considers an organization to be successful if it manages to get the resources it needs to keep afloat and grow.
- The internal process model considers an organization to be performing well if it runs smoothly and puts emphasis on the production process.

In the development sector, where program evaluation should do more than provide a partial view and integrate each of these viewpoints, the theory of social action makes sense and provides a basis for a general framework for evaluation. To succeed in infusing reality into development, the paradox of development approaches has to be acknowledged. For example, when emphasizing national capacity building to mobilize resources needed to meet the needs of a population, other functions must also be kept at a satisfactory level at the same time so that the resources can be appropriately transformed and used. If this does not happen, resources notwithstanding, the lack of synergy can be more damaging. This paradox leads to a more global and integrated approach to begin real development that is also essential as a strategy for evaluating development policies.

Concretely, we see that current evaluation approaches to the MDGs, an extremely interesting initiative, have no framework for a global and integrated evaluation of the various MDGs and their interrelationships. The absence of such a framework suggests that the MDGs contribute to development on a one-by-one basis and their synergy and interdependence are not taken into account. It is possible to make a heuristic reading to analyze the efforts and initiatives so as to enrich the development process in participating countries. It is important to have an overall, integrated vision for evaluating the MDGs, which

must be seen as a whole set of interdependent goals and modeled as such, emphasizing the conjunction of implementation strategies and the objectives that were achieved. Implementation implies programming activities, mobilizing actors and country interests as well as coordination efforts (structuring, orchestrating, planning) to reach the collective goals. Approaches to evaluating the MDGs must integrate these elements.

With Parsons' theory of social action, human action is set into its context and can be analyzed more holistically. Parsons' perspective complements the theoretical approaches to explaining social action of Weber, Durkheim and Pareto that encompass all sorts of individual and group behavior motivated and directed by the meanings that an individual discovers in the world. More precisely, action – behavior to achieve certain ends, anticipated goals or states that unfold in situations – is regulated by norms, and implies expending energy or motivation (Parsons, 1951) and action exists only in the interaction among individuals (Parsons, 1973). Parsons' work enables us to analyze social action as systems with structures, processes, and functions. His work presents the concerns linked to integrating culture and values that become more important for an analysis of development policies to explain social action, the functioning and structuring of systems.

The four functions of the action system consist in adapting to a physical and social environment; setting goals; integrating members into the social system, and maintaining shared values.

- *Adaptation* is a society's capacity to interact with its environment. Among other things, this includes collecting resources and producing consumer goods for redistribution. As its name implies, this function involves activities designed to ensure the system's adaptation to its environment, constraints, exigencies and limits, as well as those through which the

system adapts the environment to its needs, modifying, controlling and developing it.

- *Goal attainment* covers the set of actions that serve to define the system's goals of mobilizing and managing resources and the strategies designed to reach them – the ability to define goals for the future and to make decisions accordingly.
- *Integration* refers to the mechanisms that regulate exchanges among social units, in other words, harmonizing the society in its totality, maintaining coordination among the various parties for the necessary coherence and smooth operation.
- *Maintaining shared values* serves to preserve stakeholders' motivation for their actions. A "pool" of motivation exists that the entire action system must use. The action system needs to preserve a defined minimal level of motivation. This function therefore resembles something like a system for accumulating and sending energy as motivation. (Rocher, 1988) It also serves to preserve the system's stability by maintaining its institutionalized cultural models. (Thibault, 2013)

A framework is needed to evaluate the performance of a social action so as to understand its complexity. Sicotte, Champagne and Contandriopoulos (1998) use Parsons' (1951) theory of social action as a basis for developing a model to analyze organizational performance. This approach considers that a system must be conceived and analyzed with respect to four interacting functions: adaptation, production, goal attainment, and maintaining values.

- *Adaptation* involves acquiring and structuring resources to meet the population's needs. It translates the ability to adapt to external forces, to mobilize partners, innovate and provide goods and services to the population. It depends on the capacity of the actors to anticipate political, social, economic and technological trends in a context of sustainable development.

- *Production* covers not only the amount of goods and services for the population but also their optimal use as a function of invested resources. The quality of the goods and services and their adaptation to the real needs of the public are another aspect of production.
- *Maintaining and developing values* reflects the social and cultural values governing a society's value system.
- *Goal attainment* translates a State's ability to guarantee its population's wellbeing.

To conclude, it is possible to imagine the evaluation of development policies in a global and integrated manner. Theory of social action, although not the sole model, is a model that offers a solution worthy of consideration. The examples of its use are in the health area, but the model's underlying logic shows that it can be used for any organized system of action, that is, any system made of structure, actors with their logic and practices, action processes, and objectives, all in a given context. Applying this theory can lend greater synergy and coherence to development interventions.

References

- Burls, Amanda, Lorraine Caron, Ghislaine Cleret de Langavant, Wybo Dondorp, Christa Harstall, Ela Pathak-Senand and Bjorn Hofmann, 2011. «Tackling ethical issues in health technology assessment: A proposed Framework», *International Journal of Technology Assessment in Health Care* 27 (3): 230-37.
- Guisset Ann-Lise, Claude Sicotte, Pol Leclercq, William d'Hoore, 2002. «Définition de la performance hospitalière : une enquête auprès des divers acteurs stratégiques au sein des hôpitaux», *Sciences sociales et santé* 20 (2): 65-104.
- Parsons T., 1971. *Le système des sociétés modernes*, transl. by Guy Melleray, Paris: Dunod, 1973.
- Parsons T., 1951. *The Social System*, London: Routledge.
- Rocher G., 1988. «Talcott Parsons et la sociologie américaine,» *Les classiques des Sciences Sociales*, University of Montreal
- Sicotte, Claude, François Champagne, and André-Pierre Contandriopoulos, 1999. «La performance organisationnelle des organismes publics de santé» in *Rupture, Revue transdisciplinaire en santé* 6 (1): 34-46.
- Thiebaut Georges-Charles, 2013. *Les dimensions négligées de l'évaluation de la performance des systèmes de santé : les valeurs et la qualité de vie au travail*. Doctoral dissertation, University of Montreal.

ABOUT THE AUTHOR

Olivier Sossa holds a Ph.D in public health with a specialization in health care organization from the University of Montreal. Currently, he is the president of the *Société québécoise d'évaluation de programme*. His professional interests include program evaluation, analysis of public policy and change management, on which he has published several articles, particularly on establishing a culture of evaluation, managing and institutionalizing change and population-based responsibility.



David Slattery

Whither Multilateral Aid? A Bilateral Perspective on Non-core Funding to the Asian Development Bank and the World Bank*

* This paper draws from a recent ODE evaluation of Australian aid through the Asian Development Bank and the World Bank, which I led (see references). Although the views in it are very much my own, the paper has drawn and benefited from the work and ideas of the evaluation team, and others that contributed to the evaluation.

More Complex Aid Architecture

Non-core funding, or funding earmarked for specific purposes, was almost unheard of twenty years ago. Today it amounts to close to USD 20 billion annually. This equates to roughly 20% of bilateral aid and over half the volume of multilateral aid (Figure 1).

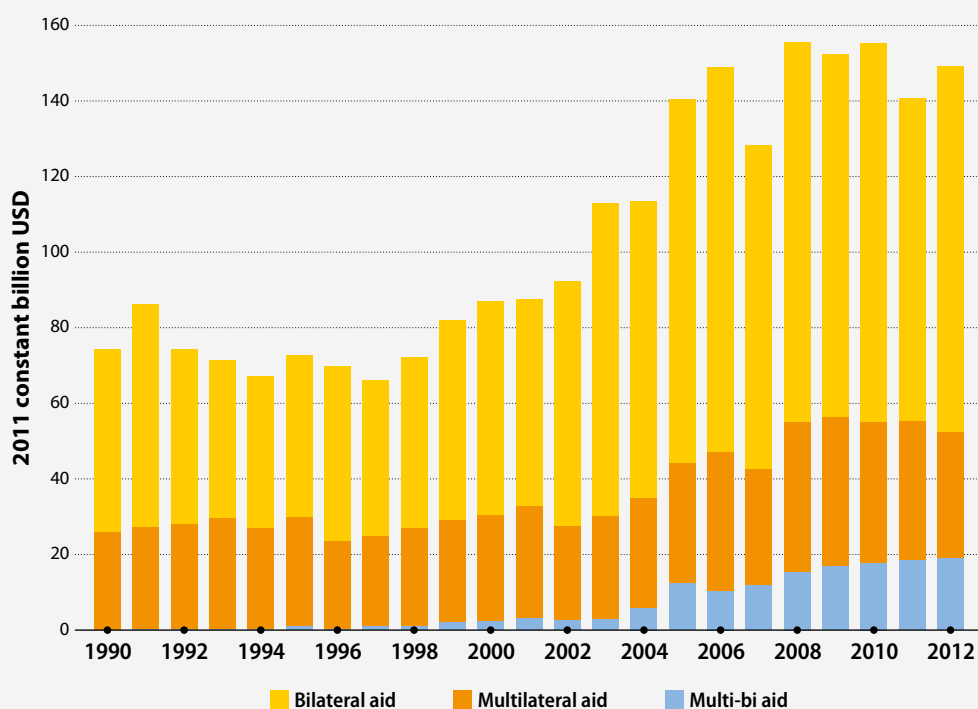
The growing importance of non-core funding in multilateral financing is noteworthy for the current and future effectiveness of the multilateral system. It is a prominent symptom of the challenge the system faces to maintain its relevance in a rapidly changing world.

Non-core funding has grown in part because of its many advantages over traditional aid. Its primary advantage is that it allows multilateral organizations to address challenges that are not covered by their core resources or mandates,

which were largely established in the post-war period. It has provided a basis for addressing global, cross-border challenges, which are arguably underfunded by the development community (see Birdsall, N. and Diofasi, A. 2015); it has enabled collective responses to countries recovering from conflict or natural disasters, and has increased aid flows to fragile states. For development banks, non-core funding has been much-needed for technical assistance, for support to preparing projects, for analytical and advisory services, and for testing innovative approaches.

Non-core funding also brings with it a number of risks (DAC 2015). These include fragmenting the multilateral system, creating a management headache for multilaterals and recipients, and reducing rather than enhancing the effectiveness of aid and the legitimacy of multilateral actors. Non-core funding has clearly contributed to a more complex global aid architecture: driving

Figure 1: Comparative Aid Amounts 1990-2012



Source: Eichenauer and Reinsberg 2015

a proliferation of trust funds within established multilaterals, many of which have outgrown their hosts to become new organisations. Moreover, the introduction of sustainable development goals will give further impetus to a more issue-driven approach to development assistance through new vertical funds and pooling mechanisms addressing specific goals.

How can we ensure that these trends do not result in less effective and less accountable aid? This question guided a recent evaluation by the independent Office of Development Effectiveness (ODE) of the Australian Department of Foreign Affairs and Trade (DFAT) of non-core funding to the Asian Development Bank (ADB) and the World Bank, Australia's largest multilateral partners and the biggest beneficiaries of its enthusiasm for non-core funding (ODE 2015). This paper describes some of the key insights from the evaluation as well as implications for evaluators to consider about the benefits and costs of non-core funding.

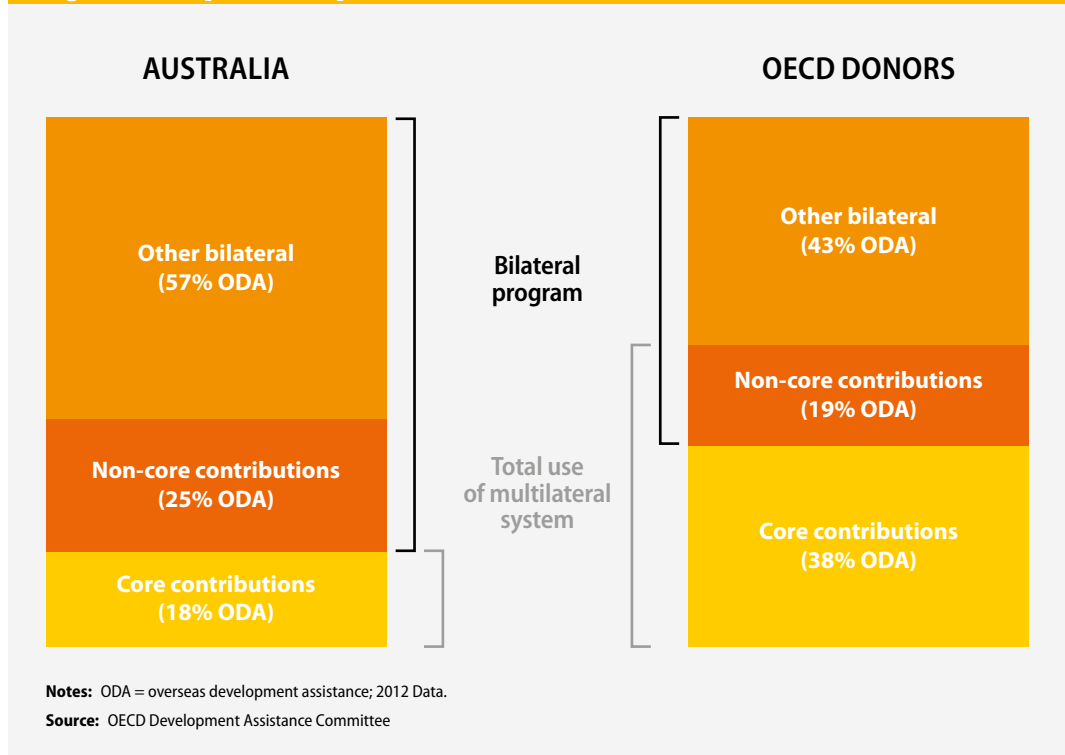
Australia's Use of the Multilateral System

OECD donors collectively deliver over half of their aid through multilateral organisations, through a mixture of core and non-core contributions. Australia uses multilateral organizations less than the rest of the OECD donors and provides much more of its funding as non-core rather than core contributions (Figure 2).

Australia's preference for providing aid through bilateral programs and its more selective participation in the multilateral system has been endorsed by most major reviews of the Australian aid program. The basis for this is the view that multilaterals do not place sufficient emphasis on countries in Australia's geographical region. This is nicely illustrated by the following quote from a 2011 study:

"Many multilaterals have a primary focus on Africa and tend to have less engagement in

Figure 2: Comparative Proportions of ODA as Non-core Contributions



Asia and very little engagement in the Pacific. Those that do operate in these regions, and particularly in the Pacific, do not necessarily prioritise these areas either for core funding or for their best staff. This has driven a tendency for AusAID to use non-core funding as leverage, helping to influence the focus and performance of multilaterals in those areas to which it attaches greatest importance.

(Dinham 2011, p. 10)

Within bilateral programs, non-core contributions have enabled Australia to rapidly expand its use of the multilateral system, growing from a relatively minor part of the Australian aid program in 2005-2006 to overtaking core contributions in 2008-2009. Most of the growth in the number and value of bilateral partnerships came in contributions to the World Bank, many of which went for global and regional partnership programs initiated by policy or thematic areas that reflected a 'sectoral' approach. Partly because of its smaller geographic footprint, the increase in funding to the ADB was smaller (Figure 3).

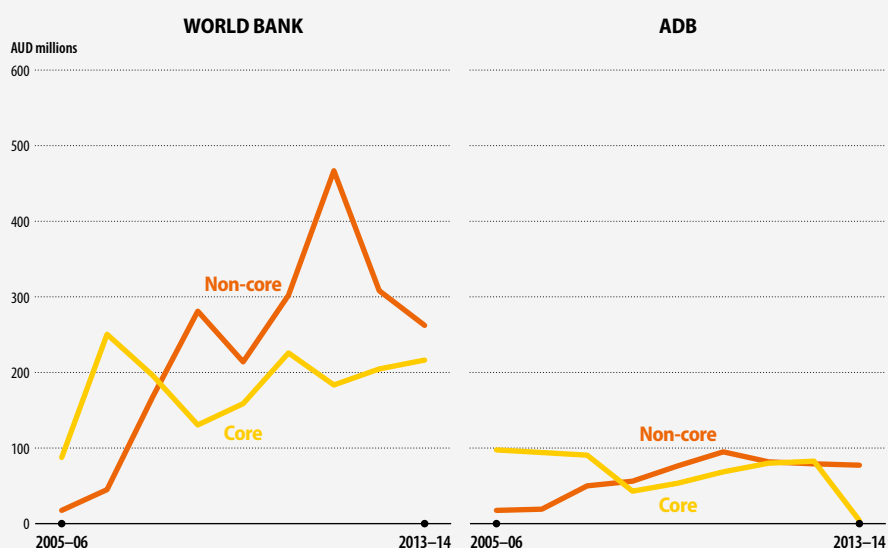
What Motivates Non-core Contributions?

Until the recent ODE evaluation, no thorough research had examined this part of Australia's aid spending. The reasons behind the growth were not well understood as a result. To address this gap, ODE surveyed DFAT staff about the motivations behind projects that they had managed over this period and asked them to weight the different factors (Figure 4).

The various motivations for funding MDBs identified by staff can be divided into four categories.

1. To facilitate a more active role in *shaping and influencing policy*. A good way to do this was to gain access to the banks' specialist expertise and roles as providers of policy advice for recipient governments.
2. To help *overcome the bank's limitations, for example, in the rules governing their budget allocations*. This is visible in the desire to support or strengthen bank engagement

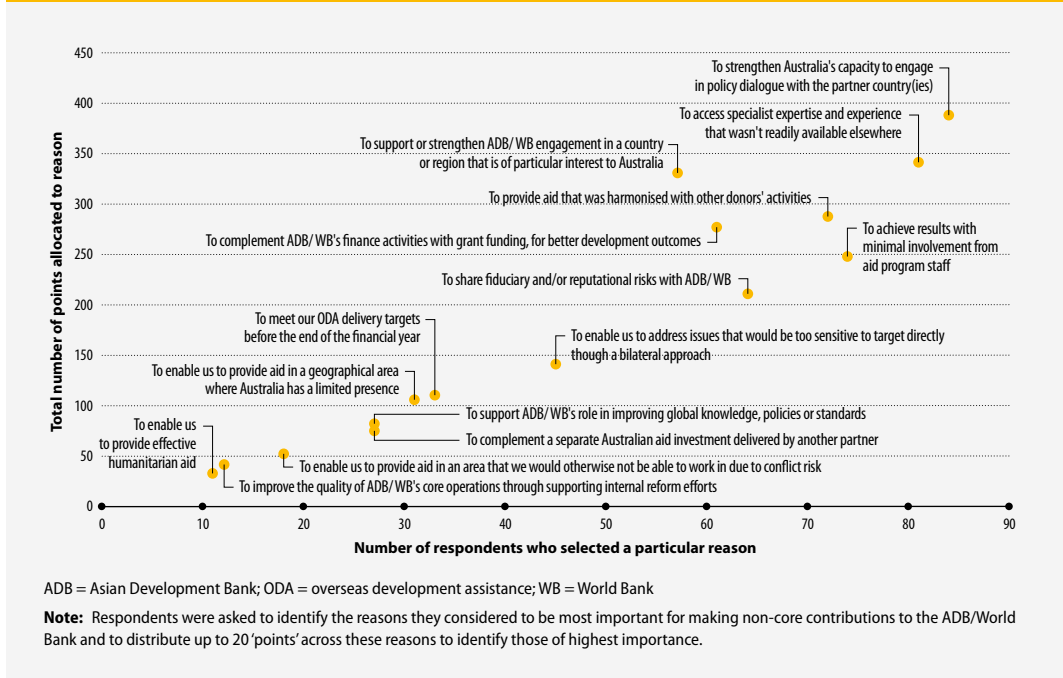
Figure 3: Growth in Non-core and Core Contributions to MDBs



Notes: Data does not include financial intermediary funds.

Source: DFAT

Figure 4: Motivations for Non-core Contributions



in countries or regions of special interest to Australia. In particular, non-core contributions have played an important role in encouraging the banks to do more in the Pacific and in Timor-Leste than their internal rules would allow. It is also visible in the motivation to complement bank loans, thereby offsetting shortages of grant funding.

3. *To harmonize aid with other donors.*
4. *To achieve results with minimal administrative overhead or staff involvement.* It was used to provide aid in countries where the aid program had few people on the ground. (This is where the emphasis on harmonization was important in providing a strong justification for a partner-led approach to programs where there was insufficient staff to manage projects directly.) A small but still significant motivation for providing non-core funding was that it was a good way to meet budget targets. Contributions to trust funds were particularly useful as balancing items in budgets: they could be paid out in

a given financial year or held over to the next depending on how units were tracking against their spending targets. While the flexibility of non-core funding is one of its main advantages, it is also undoubtedly one of the primary reasons for its volatility as a funding source (DAC 2015).

Most of these motivations for the growing use of non-core funding to the ADB and the World Bank make perfectly good sense in terms of aid effectiveness. However, some motivations have little to do with aid effectiveness.

Why Donor Motivations are Important

There are a number of reasons to be concerned about donors' more pragmatic motivations, including the risk of fragmentation in the multilateral system. Non-core funding is prone to creating new 'financially insignificant aid relationships', which have grown by one-third over the last decade. The 2012 DAC Report on Multilateral Aid noted, "... it is obvious that

the increased number of aid relations [in recent years] has led to increased fragmentation.” The evaluation found that the impacts of non-core funding on fragmentation and complexity are difficult to assess because they are very context dependent. In some contexts, there is evidence that working through partners has actually exacerbated the problems that the collaboration was designed to address. However, there are also many good examples where non-core funding has been a very effective strategy for minimizing the aid management and coordination burden on recipient governments and providing excellent transparency and strong accountability over results.

Another major risk is that pragmatic motivations will result in aid driven by donor priorities that are disconnected from recipients’ needs. This was mostly a concern with global and regional funds that MDBs sometimes created to exploit donors’ most current priorities, but which occasionally turned out to be spent with less accountability in ways very similar to core funding. This view is supported by the World Bank’s Independent Evaluation Group (IEG) reviews of global and regional partnership programs that have identified a range of legitimacy, accountability, efficiency, transparency and fairness issues (World Bank 2010). More broadly, it is significant that the IEG has found less internal accountability for non-core funding than for regular core resources.¹ The closer direct involvement of recipient governments made these factors far less problematic with recipient-executed projects, especially those involving loans.

A final reason for donors to be concerned about motivations is more self-interested. Where pragmatic motivations have dominated, the quality of DFAT’s engagement with the banks

¹ For example, in a survey of sector managers, the IEG asked, ‘Is the quality assurance of trust-funded activities routinely exercised within your unit with the same degree of rigor as Bank-financed activities?’ One half saw the quality assurance as equal and another half rated it as inferior. See World Bank (2011), *Trust Fund Support for Development: An Evaluation of the World Bank’s Trust Fund Portfolio*.

has been limited. This has tended, in turn, to reduce satisfaction with results,² which was evident in a range of ways. Without the ability to engage substantively, non core contributions have sometimes been more successful in helping the MDBs develop their own lending activities rather than strengthen Australia’s engagement in policy discussions with partners. Without substantive engagement, it was difficult to realize many stated bilateral objectives such as learning about innovative aid approaches or making other links between MDB-managed and bilateral programs. Moreover, it was difficult to gain recognition for contributions, which is one of the main potential advantages of non-core as an alternative to core funding in partnering with multilateral organizations.

Why Engagement Matters

The evaluation claims no direct relationship between the weak engagement of any given donor and bank accountability. Other donors, aid recipients, and the banks themselves, play a big role in this. However, to the extent that individual donors do have an influence, a problem with their weak engagement is that it can undermine accountability. In a general sense, weakly engaged shareholders and stakeholders, especially developing countries, present a real risk, particularly in global and regional funds. More specifically, lack of donor engagement can increase the risk that the banks will not invest sufficient effort in rigorous monitoring and accounting for results.

To find out where weak engagement was most likely to be a risk, we analysed our internal

² Survey and case analysis show that DFAT’s engagement has been tenuous in a significant minority of projects, and that the problem was particularly acute for multi-donor bank executed programs. In ODE’s view, this reflects a tendency to invest minimal staff time where there is a perception that other donors can be relied upon to engage well with the project, or that low engagement will not increase reputational risk. In part, reputational risks are considered low because they are shared with other donors. In part, this is because a significant proportion of multi-donor funds are global or regional, which means that donors are one step removed from the contexts in which they are implemented. This lowers the risk that any particular problem will be credited to one donor.

performance data about the quality of M&E of projects that had been funded. M&E is one of six criteria that DFAT program managers give a rating to for every aid project, once a year. An analysis of M&E ratings reveals significant differences depending on how projects are executed. M&E ratings for World Bank projects were identical for bank and recipient-executed projects during implementation.³ However, by project completion, M&E ratings for recipient-executed projects had improved significantly. By contrast, unsatisfactory M&E ratings for Bank-executed projects increased significantly by completion. Similar to the trend for Bank-executed projects, M&E ratings for projects implemented by managing contractors deteriorated from implementation to completion, as indicated by the proportion of projects given unsatisfactory compared to highly satisfactory M&E ratings. (Figure 5)

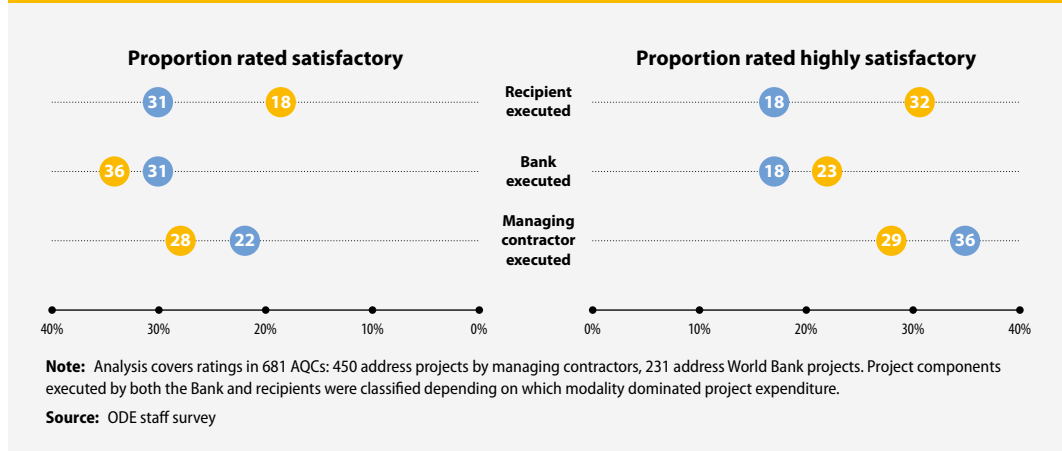
A number of things can explain the superior results of recipient-executed projects: their scope is limited to a single country and, usually, to a single sector; they are owned by recipient governments, which adds another layer of accountability, and they are subject to standard Bank M&E policies and requirements contrary to Bank-executed projects.

3 This analysis covered only World Bank projects because there were insufficient numbers of ADB projects in different categories to produce robust findings.

Some features of Bank-executed projects make them very difficult to evaluate: These projects often include large numbers of sub-projects across multiple countries and sectors and they often complement or support larger investments, making it difficult to assess their value over and above those investments. The difficulty of accounting for the results of these types of projects is most acute in multi-country arrangements where there is the added difficulty of presenting aggregate information about results obtained in heterogeneous contexts. As a result, over 36% of DFAT M&E ratings of World Bank projects and 62% of the ratings for ADB projects in this category were unsatisfactory. Similarly, the evaluation found that umbrella funds (mostly World Bank), which have many of the characteristics described above, have not typically had strong M&E arrangements.

The IEG has levied similar criticisms against World Bank global and regional partnership programs, which have been a major focus of its evaluations. One recent paper observed, *Partnership programs are not required to have results frameworks at inception and as a result often take many years to set them up. Moreover, partnership programs by their nature represent compromises between multiple partners. Partners agree on broad directions but often differ on nuances and priorities. As a result, programs are*

Figure 5: M&E Ratings for Recipient-executed World Bank Projects Improved on Completion v. Implementation While Ratings for Bank and Managing Contractor-executed Projects Dropped



settled with vague or excessively broad objectives, resulting in weak links between programs' activities and their larger objectives. IEG found that partnership programs have started to develop a culture of evaluation but that many programs regard periodic evaluation as a substitute for putting adequate monitoring systems in place. Periodic evaluations often result more from donor pressure than from program managers' desire to learn lessons to improve their effectiveness. (World Bank Group, p. 13)

In short, monitoring Bank-executed projects needs to improve to get the best results from Bank non-core funding as a major component of multilateral aid.

Conclusion

The emergence of non-core funding as a significant feature of the aid landscape reflects its advantages over more traditional forms of aid. Many of these advantages figured prominently in Australia's motivations for providing non-core funding to the ADB and the World Bank. It is however also important to recognize that some of the advantages of non-core funding have little to do with its merits in terms of development. The international community must therefore be alert to the risk that this funding might undermine rather than enhance aid effectiveness.

ODE's evaluation of non-core funding to the ADB and World Bank suggests that a number of things need to happen to obtain the best possible results while minimising the risks to aid effectiveness.

1. Pay close attention to understanding donor motivations in providing non-core funding. In particular, there is a need to counteract some of the more pragmatic drivers for contributions, which may contribute to an unnecessary fragmentation and proliferation of channels, and reduce accountability.

2. Donors and developing countries must be strongly and consistently engaged in the governance of trust funds. Evidence from the evaluation and from the IEG suggests that weaknesses on both fronts are more of a concern with bank-executed projects than with recipient-executed projects, at least in the case of the World Bank.
3. Improve M&E frameworks for non-core funding, especially for global and regional programs where long, complex causal chains have stymied efforts to account for results and value for money. Intermittent evaluations are no substitute for rigorous monitoring underpinned by strong demand by shareholders and aid recipients.
4. Continue to draw attention to these issues so that they remain on domestic and international agendas.

References

- Birdsall, Nancy and Anna Diofasi, (2015). *Global Public Goods for Development: How Much and What For?* <http://www.cgdev.org/publication/global-public-goods-development-how-much-and-what>
- Dinham, M (2011). *Study of AusAID's Approach to Assessing Multilateral Effectiveness*, www.aidreview.gov.au/publications/study-multilateral.pdf
- Eichenauer, V. and B. Reinsberg, (2015). "What determines earmarked funding to international development organizations? Evidence from the new multi-bi aid dataset" <http://www.ipz.uzh.ch/de/forschung/lehrstuehle/ep/research/internationaldevelopment/multi-bi-aid.html#5>
- OECD DAC, (2015). *Multilateral Aid 2015: Better Partnerships for a Post-2015 World* at http://www.keepeek.com/Digital-Asset-Management/ocd/development/multilateral-aid-2015_9789264235212-en#page23

World Bank IEG, (2010). *The World Bank's involvement in global and regional partnership programs: an independent assessment.*

--- (2011). *Trust Fund Support for Development: An Evaluation of the World Bank's Trust Fund Portfolio.*

--- (2015). *Opportunities and Challenges from Working in Partnership: Findings from IEG's Work on Partnership Programs and Trust Funds.* <http://documents.worldbank.org/>

[curated/en/2015/01/24321675/opportunities-challenges-working-partnership-findings-iegs-work-partnership-programs-trust-funds](http://documents.worldbank.org/curated/en/2015/01/24321675/opportunities-challenges-working-partnership-findings-iegs-work-partnership-programs-trust-funds)

ODE (2015), *Banking our aid: Australia's non-core funding to the Asian Development Bank and the World Bank*, Commonwealth of Australia, <http://dfat.gov.au/aid/how-we-measure-performance/ode/odepublications/Pages/evaluation-australias-non-core-contributions-adb-wb.aspx>

ABOUT THE AUTHOR

David Slattery is the Director of the Office of Development Effectiveness (ODE), the independent evaluation arm of Australia's Department of Foreign Affairs and Trade. David joined ODE from a career with the Australian Government's "watchdog," the Commonwealth Auditor-General, where he led high-stakes reviews of large Government programs. He has conducted over 14 in-depth evaluations of international aid and domestic programs. David has a Masters in Program Evaluation from the University of Melbourne.



Foster N. Ofori

Building Africa's Innovative Capacity: the Role of the AfDB

To reach a reasonable level of sustainable development, every society needs to integrate technologies in a carefully planned manner. As Africa pursues an agenda of transformation, there is an urgent need to bridge the knowledge and capacity gaps in many countries and sectors. The Millennium Development Goals (MDGs) called for a new global partnership for development using science and technology to address the problems facing the poor. The governments of many African countries recognized the need to reorient their national science, technology and innovation policies in order to serve developmental needs more effectively and coherently. A recent report jointly published by the United Nations Commission for Africa, the African Development Bank, the African Union and the United Nations Development Programme points to noteworthy progress on technology indicators (UNECA, 2015) largely attributable to great advances made in the diffusion of Information and Communication Technologies (ICT) in Africa.

ICTs are credited with a capacity to transform national economies, organizations, and the global economy itself. Assumed to offer significant potential benefits for national socioeconomic development, they hold out the promise to developing countries to develop more rapidly. There is a general belief that ICT dissemination leads to economic growth given the enormous potential to increase the productivity of almost all economic sectors and to expand and improve the quality of services. But ICTs can only be integrated into other well-developed technologies that increase productivity in such sectors as agriculture, energy, manufacturing and trade and services.

This is reflected in the newly-adopted Sustainable Development Goals (SDGs). SDG Goal 9 specifically focuses on the capacity to build a resilient infrastructure, to promote inclusive and sustainable industrialization, and foster innovation. It calls for supporting domestic technology development, research and innovation, and upgrading the technological capabilities of industrial sectors, and

significantly increasing access to ICT. Beyond science and technology, realizing Goal 9 would also require the capacity to understand the local, social, environmental and institutional contexts of innovation and technology development. Innovation requires skill development and technological development.

To achieve this degree of technological advancement, Africa needs to create innovative processes that acquire or develop new technologies and also disseminate them to determine their social and economic applications system-wide (Ofosu, 2005).

Exploring the Concept of Innovative Capacity

If economic development requires innovation, then innovative capacity is essential. The concept was first introduced by Prof. Suarez-Villa (1990) to measure the potential for innovation in any nation, geographical area, or economic activity by assessing national ability to produce and commercialize a flow of innovative technology over the long term.

National innovative capacity depends on the strength of a nation's common innovation infrastructure (cross-cutting factors that contribute broadly to innovativeness throughout the economy), the environment for innovation in a nation's industrial clusters and the strength of their linkages (Furman, Porter, Stern, 2002). The nature of the linkages between government policy, universities and research institutions, and private actors must be clearly defined to develop creative and innovative competencies. Moreover, these must be combined with complementary business capabilities. A strong organizational and institutional support for knowledge absorption and generation and for supporting the development of associated innovative capabilities in firms is also required.

Building Innovative Capacity for Africa

If innovative capacity holds the key to unlocking Africa's potential and achieving inclusive growth and sustainable development, Africa must explore new avenues to enhance technological innovation and create an environment that facilitates local, national and international initiatives to develop new technologies for tackling some of the many socio-economic challenges. Sustainable development also requires economically viable, socially needed, relevant and accepted sustainable technologies.

The innovation agenda requires resolute, coordinated action at the highest levels of development cooperation. Development Finance Institutions (DFI) have a responsibility to support countries to build linkages in national and regional innovative systems. They can be both financiers and policy advisors in promoting innovation. Beyond funding, the AfDB can provide technical assistance to state organizations that are relevant to creating an environment for innovation.

Capacity to Mainstream Innovation into Development Project Management

The African Development Bank's (AfDB) Ten-Year Strategy 2013-2022 includes five operational priorities, one of which is to develop skills and technology to transform Africa. This drive is designed to equip young people with the appropriate skills to make them employable or able to create small businesses, important drivers for poverty eradication and sustainable development in Africa. As AfDB president, Dr. Adesina states, "My vision is to expand opportunities and unlock the potential of African countries and African people AfDB's "High Five" Goals for unlocking Africa's potential through sustainable development. 1) Light up Africa 2) Feed Africa, 3) Industrialize Africa 4) Integrate Africa and 5) Improve Life in Africa. These broad goals

are set within three areas of special emphasis: Gender, Climate Change and Fragility (Figure 1).

Achieving these goals and special areas of emphasis can be best done by reorienting approaches and processes and integrating relevant technological innovations. This requires more than ICT or technological innovation, however; it includes bringing new ideas to make better and more efficient processes (Process Innovation) and creating new organizational forms to manage projects (Organizational through the Innovation) or a new and improved service delivery system.

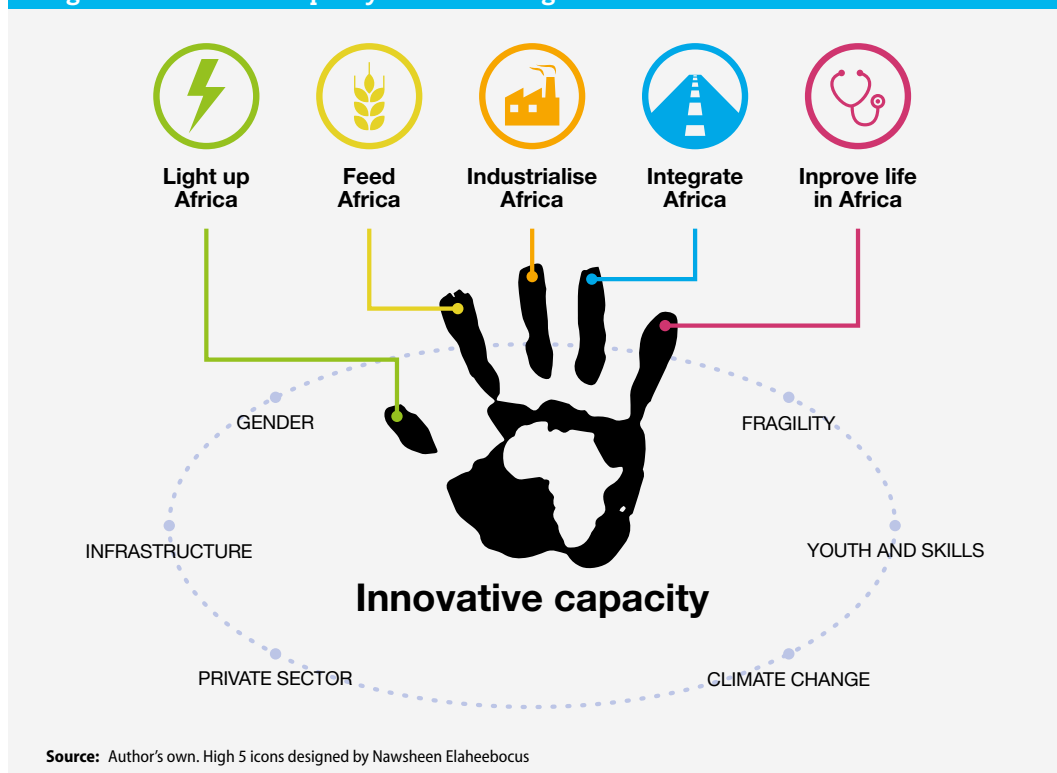
While AfDB operations reflect its mandate to eradicate poverty and promote sustainable development, there is growing evidence that delays in AfDB-funded projects can be attributed, at least in part, to cumbersome processes or lack of appropriate expertise to understand and implement them.

Therefore, to achieve greater efficiency and more effective impact from development projects, the AfDB needs to build its innovative capacity to improve project cycle management by shortening and smoothing cycle activities by developing and using technological interventions. For example, well functioning technological platforms could enhance data collection in country and regional programming activities. Similar technologies could simplify procurement, disbursement and financial management and speed up project implementation.

Conclusion and Recommendations

Recognizing the significance of innovation in driving the sustainable development agenda and the pivotal role it plays in both the Ten-Year Strategy and the High 5 Goals of AfDB, it is even more critical to mainstream innovation

Figure 1: Innovative Capacity and AfDB's High 5 Vision



into Bank operations and processes, both for Bank Staff and RMCs.

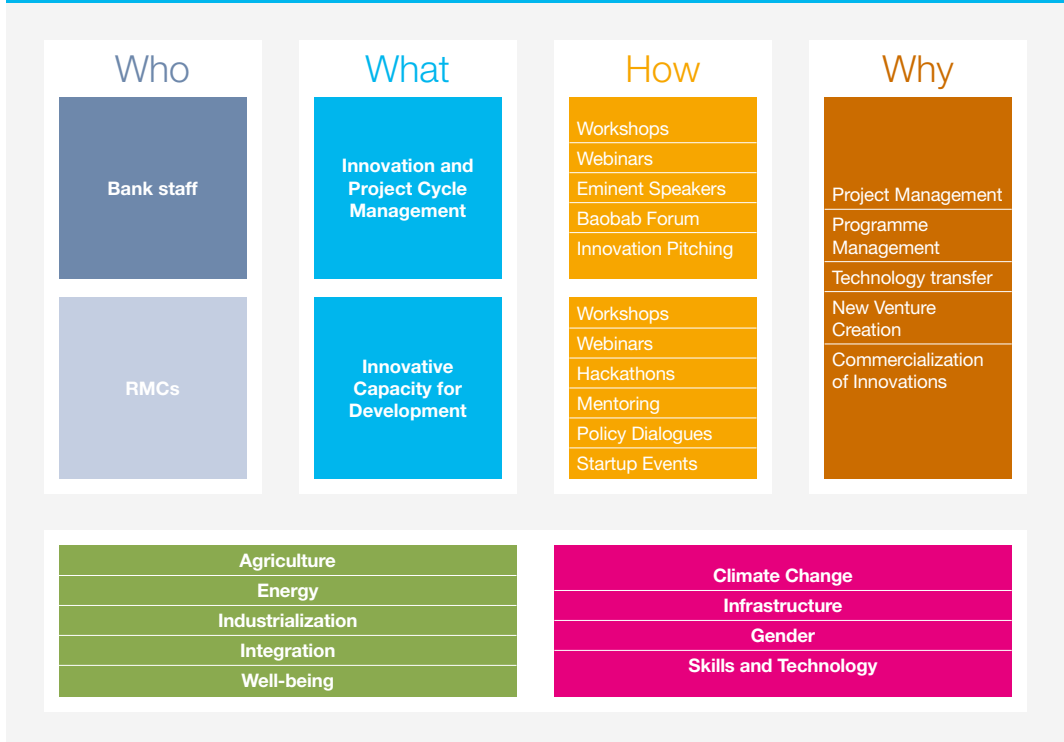
To operationalize the proposed framework consideration should be given to two key areas of focus: Innovation and Project Cycle Management and Innovative Capacity for Development. The former should focus on mainstreaming innovation in managing bank-funded projects to enable staff to plan, implement and evaluate project cycle activities. The latter should involve a series of activities aimed at building capacity of RMCs to develop homegrown technological, process and organizational innovations for value addition.

Figure 2 summarizes the key components, key target audience and rationale for mainstreaming innovation into bank business and processes.

References

- Furman, Jeffrey L., Michael E. Porter and Scott Stern, (2002). "The Determinants of National Innovative Capacity," *Research Policy* 31:899-933.
- Ofori, Foster N., (2005). "Towards a Global Agenda for Technological Innovations in Africa: A Systems Approach," in Stenberg and Sutinen (eds). *Perspectives on Corporate Social Responsibility in International Business*, University of Turku. https://books.google.com/books/about/Perspectives_on_Corporate_Social_Respons.html?id=cBhGAAAACAAJ
- Suarez-Villa, L., (1990). "Invention, Inventive Learning and Innovative Capacity." *Behavioral Science*, 35:4, pp. 290-310.

Figure 2: Mainstreaming Innovation: Key components, target audience and rationale



ABOUT THE AUTHOR

Foster N. Ofofu is a Capacity Development Specialist at the African Development Bank Group (AfDB). He created and manages the institution's Distance and eLearning Initiative. Prior to joining the AfDB, Foster was a Senior Lecturer in International Business and Innovation Management at both the Turku School of Economics and Turku University of Applied Sciences, Finland, and at the Ghana Institute of Management (GIMPA). Foster has also consulted for European and African organizations in internationalization and commercialization of innovations in the ICT and Biotechnology industries. He has published several articles on International Business as well as Innovation Management. Foster has a background in International Business Management with special focus on Innovation and International Technology Transfer.



Shikoh Gitau

Innovation: Africa's Forte

The rhetoric of and about Africa is that of hunger, poverty, desolation and war. A whole industry has been built around this premise. For decades, development economics has been developing models to eradicate poverty from the continent. The somewhat successful Millennium Development Goals were optimistic attempts to improve the lives of Africans. The newly ratified and adopted Sustainable Development Goals are an even more ambitious attempt to improve the lives of the poor, most of whom reside in Africa.

As the world makes another attempt to eradicate poverty, with different goals, there is a lot of optimism. But when something that could be greatly improved has not changed for a long time, the easy thing to do is get frustrated and presume it can not be changed, a stance that many SDG commentators have taken. However, the smart question to ask is how one can change it.

The goals are smart. The approach to achieving them is not as smart. What can be done differently in the approach towards achieving the SDG's that was not considered in the MDGs?

We believe it is important to take a step back and, perhaps, take advantage of potential, encourage and support homegrown solutions, identify, amplify and **celebrate** Africa's innovators, especially young men and women.

A Series of Firsts

African Development Bank President Akinwumi Adesina laid out what the Bank was seeking to accomplish through innovation: “We want to generate big ideas to generate big wins, and that means taking big risks.... For women and youth, technology is a fantastic opportunity to bypass traditional structural challenges.” This was in his address to the participants of inaugural innovation weekend that took place October 9 to 11, 2015.

During the weekend, the Bank opened its doors to West Africa’s aspiring entrepreneurs. This first-ever Technology Innovation Weekend was also a first for Abidjan. It was hosted by the Office of the Special Envoy on Gender (SEOG), Geraldine Fraser-Moleketi, who challenged participants to draw inspiration from the long tradition of innovation in West Africa. “As with Timbuktu, a formerly great center of learning and innovation in the region, we need to reclaim our heritage of innovation that is indigenous to West Africa.”

Participating government and business leaders had the ambitious objective of kick-starting a hub of innovation. Freddy Tchala, CEO of the Côte d’Ivoire telecommunications giant MTN talked about the need to create a tech hub in Francophone Africa, something like an Abidjan Valley. He emphasized the importance of engaging young people: “What’s important for the youth is that they have new creations and can reduce the costs of participation and collaboration.... In Côte d’Ivoire, we are trying to create ways for youth to express themselves, and that expression will be innovation, perhaps offering solutions to several problems.”

The weekend gave a special focus to creating technology innovations to improve the livelihoods of youth and women in Africa. Workshops and speakers, including executives from Google, Facebook, Orange, and MTN, focused on two main themes for technology innovation in West Africa: financial inclusion and skills development, to address the region’s need to spread modern finance and create jobs. About 60% of the unemployed in Sub-Saharan Africa are 15 to 24 year olds, and many -- most of them young women -- live on less than \$2 day. GSMA, which represents mobile operators, reports that mobile technology is growing rapidly in the region and could create 6.6 million jobs in the next five years.

Participants were enthusiastic. Michel Aka, a young, aspiring technopreneur commented, “During this event which literally broke the ice between youth and decision-makers, I saw a responsible youth, capable of deliberate sleep deprivation during an entire weekend to work on projects, being motivated by the burning desire to change things and give a better standard of living to a sister, a mother, a father, a family, a city, a country, or even a continent. Participating in this weekend gave me the opportunity to see the Bank’s strong will, as well as its faith in youth who want to mature as actors in their own development. Such an event brings back confidence in our continent, and shows the existence of a deep sense of sacrifice from managers to change things.” Pikiz co-founder Ines Affo remarked, “What really motivated me to participate is that the young face many difficulties, especially women, and the emphasis on women of this event really motivated me.” Morissa Djaha, a student, said, “ I learned to work with people I do not know, with foreigners in whom I have never had an interest. It was really new for me and very instructive.”

Bank staff Winifred Greywoode observed, “I volunteered to assist with the planning and ended up offering guidance to the teams. The groups consisted of young people who very often had little work experience but plenty of life experiences. The ideas they came up with were to solve problems for a mother whose tomatoes rotted because she could only carry one bag to the market, for a father (the sole provider of a family) who died due to lack of medical facilities or expertise and for Aisha who never got to go to school because she had nine little brothers to look after. Listening to these ideas, offering advice, provoking thought, challenging the viability of the projects was what I put in. What I got out was a whole lot more: Watching some of the participants evolve into confident individuals who could pitch a business idea (and interest an investor); Teaching them with the help of some fabulous facilitators how to begin with the problem and how to wear the shoes of your client; that smile and that person who came up to me and said I do not know if we will win, but thank you.”



Photo ©: AfDB flickr

The Bank drew lessons from the event, not only about how it can further instill a culture of innovation, but also how it should nurture technology-led innovations in Africa, that it will ultimately fund while fulfilling its strategic mission to boost jobs and improve livelihoods, especially for women and youth. Director Simon Mizrahi pointed out the link between opportunities and innovation, which “fundamentally means creating new opportunities for people. The Bank supports this idea by creating an environment to generate new ideas from different countries across Africa.”

Background and Aims

The idea for the event was born during the AfDB 2015 Annual Meetings, during which the Office of the Special Envoy brought together a panel of technology companies to discuss the role of technology and innovation in Africa’s transformation. The companies challenged the Bank and the Ivorian Ministry of ICT and Transport to find ways for young people to become an integral part of the solution. The SEOG committed to organizing and supporting an innovation event involving youth this year. The event became a Bank-wide endeavor, with staff who reached out to partners and secured over 15 mentors, six speakers and six private sector partners as well as internal Bank funding.

The weekend was designed to engage all Bank professional and support staff and to prompt a shift towards innovative thinking, leading to new products, programs, projects and processes. It was also designed to engage technology design experts, entrepreneurs, young men and women eager to tackle their daily challenges by finding “technology-based solutions to improve the quality of life of the people of Africa” with an emphasis on financial inclusion and job skills. It also sought to find

new ways to collaborate with policymakers, the private sector, entrepreneurs, academia, and civil society to develop, deploy and disseminate technology in Africa. Google, the International Telecommunications Union, Facebook, and Women Techmakers, among many others, partnered with the Bank.

The response

The interest in the Tech Weekend far exceeded expectations: for 120 spaces, over 1200 people from across West Africa and within the Bank tried to register, including young men and women entrepreneurs from technology and development communities, ranging from app developers to students and academics from Niger, Burkina Faso, Mali, Benin, Cote d'Ivoire, Guinea, Togo and Senegal wanted to take part.

A Design Thinking workshop led by Dahlberg Design Impact Group kicked off the event to help contestants shape and refine their business ideas. The Dahlberg facilitator expressed his hope “that we can get the innovators to design something they understand fully that they can implement in competition. And whether they win or not, doesn't matter for them. Every participant is a winner and what is important is that they've figured out the problems, that they've seen the potential value that the product may have, and proceed to implement it later.”

Winning innovations

The 80+ participants were divided into teams that brainstormed over the weekend to develop innovations. Each team had access to private sector mentors and to Bank staff. In the end, 11 ideas were entered into the innovation challenge and four projects won a Bank-sponsored innovation fellowship in partnership with AMPION and Orange Fab, the Orange start-up acceleration program in West Africa: (i) PayFree, a multiplex platform for payments; (ii) La Ruche, a marketplace for artisans to sell their wares; (iii) Coliba, a mobile platform

for managing urban waste, and (iv) BioPro, a mobile app to help rural people get access to energy and electricity. All 11 teams received a free six-month membership to Facebook's FBStart, an online resource and mentorship platform for start-ups that also gives social impact services, such as fundraising advice.

Social media broadened the impact of these selected innovations: An organization looking to finance the La Ruche winning project, contacted the Bank following the event.



Photo ©: AfDB flickr

A highly tweeted event

A dedicated facebook page, and a twitter hashtag were created to reach out to those who had tried to register but could not attend. The online response was as overwhelming as the event: it generated the second-most twitter impressions of any AfDB event (second to the Bank's annual meetings), with over 460,000 unique twitter accounts reached and 2.7 million unique interactions with the Twitter handle #AfDBIWE15

The Big Take Aways

The Bank can take away three main lessons from the first Innovation Weekend.

1. Awareness of innovation within the Bank is very low

Many discussions with Bank staff during the organization and execution of this project suggest the need for deeper engagement and more opportunities to develop innovative ideas in their daily work.

2. Design thinking capacity

What aspiring entrepreneurs need most is training in design thinking to succinctly identify a problem and work towards a solution on turning innovative ideas into companies that the Bank can fund. However, it cannot do it alone.



Photo ©: AfDB flickr

3. Private sector engagement is crucial

Given the speed of the private sector, the Bank must reduce red tape and fast-track the process of partnering.

Minding the GAP: AfDB Fellowships

The AfDB fellowship addresses the main reason that many of these ideas never make it to fundable companies – entrepreneurs need support to move from an initial concept to investment. To make good on the potential of technology-led entrepreneurship for women and youth is one of the most powerful yet challenging solutions available for the need to create millions of jobs and boost African economies. The AfDB faces this challenge in creating an environment to build a pipeline of bankable projects for its Africa Innovation Fund set up through OITC3. With private sector partners, a Bank fellowship is testing a model for the whole value chain from idea to incubation, from acceleration to investment.

The four teams comprising 24 fellows selected from the Tech Innovation Weekend will benefit from mentors, professional consultants, potential investors and access to incubators, international conferences, and other support from AMPION, a pan-African ICT entrepreneurship initiative, and other partner networks. Start-ups will be kept engaged throughout the process and build their capacity to think through their invention from the user's perspective. Over the 18 months of the fellowship, high potential start-up ideas will be taken through refinement into funding in two phases that provide live and virtual mentoring and a program for growth.

- **Phase I:** a six-month incubation period during which teams will be guided on refining their ideas and taking them to market with a tailor-made program.
- **Phase II:** An acceleration phase during which Orange's Fab program helps entrepreneurs take their start-ups to investment readiness.

The Way Forward for the AfDB

The experience of launching the first AfDB Innovation Weekend has highlighted four areas that the Bank must urgently address:

1. A Bank-wide agenda to promote the importance of innovation and to raise awareness of its strategic role.
2. Meet the thirst for tech innovation instruction from Africa's youth, especially in underserved West Africa, and particularly in francophone countries. To that end, two events are planned in 2016, one in Abidjan

and one in Zambia to coincide with the AfDB Annual Meetings.

3. Multiply and expand similar bottom-up approaches to build tech-led entrepreneurship for youth.
4. Form partnerships with private initiatives such as the Orange accelerator offering participants free boot camps and other networks of support and services, which involve modest investments for the Bank.

More lessons will undoubtedly emerge when the first group of 24 fellows completes the program.

ABOUT THE AUTHOR

Shikoh Gitau is an ICTD technology design and implementation expert who currently leads the AfDB Technology Innovations for Inclusive Growth Initiative. She has worked in innovation in technology with private, non-governmental and governmental institutions including Microsoft, Google and Mercy Corps. She is the founder of Ummeli, a mobile-based social network that connects unemployed youths to employment and skills development opportunities.

A member of the GSMA Women's Steering Group, UN-Women, Women in ICTD group, and the Clinton Global Initiative on Women Leading Women in ICT, she was recognized as a Top 40 under 40 woman to watch in Kenya 2015 and was an ASPEN New Voices Fellow for 2015. In 2013, she received the Anita Borg Change Agent Award, and was on the AfroElle 2013 Power List of African women who affect change. The Huffington Post named her a technical woman to watch. Dr. Gitau earned a Ph.D. and a M.S. in Computer Science from the University of Cape Town.

COMM
ideas for a
write

evaluationmat

ments?

an article?

to us

ters@afdb.org

“Recent and coming changes to the development co-operation agenda will require evaluators and the evaluation profession more generally to evolve. Evaluators will need to be ready to evaluate whole of government approaches, increasingly integrated policies, global public goods, a more diverse range of partnerships and increasingly complex forms of development assistance. The key questions to be asked are: Are we fit for the future? What will it take? How will we get there?”

Karen Jorgensen, Development Co-operation Directorate, OECD DAC, p.8



AFRICAN DEVELOPMENT BANK GROUP



IDEV

Independent Development Evaluation
African Development Bank

idev.afdb.org