1 BETTER CRITERIA FOR BETTER EVALUATION: REFLECTIONS ON THE ADAPTED EVALUATION CRITERIA AND NATIONAL EVALUATION SYSTEMS

MEGAN G. KENNEDY-CHOUANE
HEAD OF EVALUATION, DEVELOPMENT CO-OPERATION DIRECTORATE,
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Drawing on the outcomes of the recent adaptation of the widely-used "DAC evaluation criteria", this article provides reflections for those working to strengthen national evaluation systems. Key messages are that the evaluation criteria can be a useful starting point for asking the right questions; that the criteria should be used within a broader normative framework with attention to evaluation purpose and careful selection of topics to meet national needs; and that questions of ownership in evaluation are shifting in the global development landscape. The article draws on the public consultation and survey about the criteria carried out in 2017-2019, discussions at the 2019 National Evaluation Capacities (NEC) Conference and other international conferences, input received during the keynote plenary session in Hurghada, and material from the pre-conference training on using the criteria in national evaluation systems conducted by Ms. Kennedy-Chouane and Mr. Velayuthan Sivagnanasothy.

INTRODUCTION

A GLOBAL EVALUATION CONVERSATION

On 10 December 2019, the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) formally adopted a new version of the widely-used evaluation criteria commonly referred to as the "DAC criteria".[121] The official outcome of the two-year process that led to the adoption was a new document laying out the purpose of the criteria, principles for their intended use and definitions for relevance, coherence, effectiveness, efficiency, impact and sustainability (see Figure 1), as well as a new guidance (currently being developed).

Beyond these formal outcomes, the process of modernizing the criteria was itself an important moment for the global evaluation community. Over the course of two years, discussions on the criteria and their use in today's development context kicked off far-reaching reflections – with lessons that can be useful for those working to strengthen the role of national evaluation systems.

---

121 OECD, Summary Record of the 1070th meeting of the DAC held on 10 December 2019, Development Assistance Committee OECD, Paris, France, 2020.

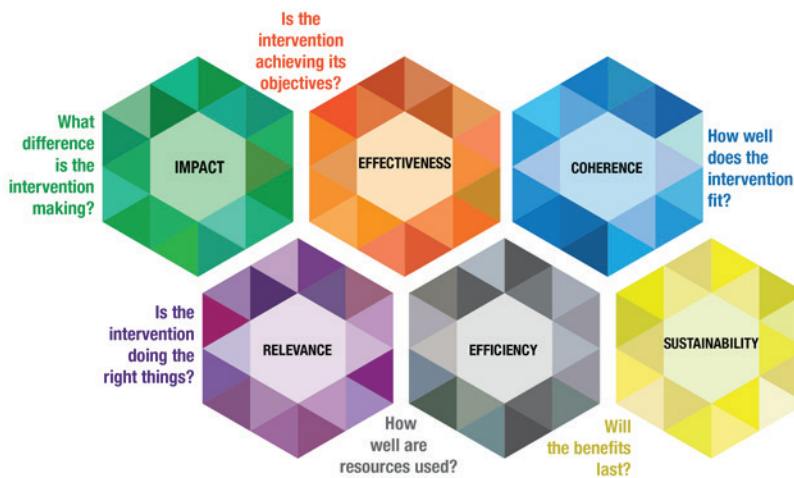**Figure 1.** The new evaluation criteria

Image: Stephanie Coic, OECD

THE ORIGINS OF THE CRITERIA AND THE ADAPTATION PROCESS

The criteria of relevance, effectiveness, efficiency, impact and sustainability were first laid out by the OECD DAC Network on Development Evaluation (EvalNet) within a set of principles for evaluating development assistance,[122] and later codified in a Glossary.[123] The criteria are normative, describing the desired attributes of interventions, and providing a guide to the questions evaluations aim to answer in determining the merit, worth or significance of an intervention.[124] Each criterion provides a different perspective on the intervention implementation and its results. DAC members brought the criteria to life through evaluation policies, guidelines and manuals, and these criteria eventually became a cornerstone of evaluation practice.

In 2017, EvalNet started a process to adapt the criteria in light of the 2030 Agenda for Sustainable Development, the Paris Agreement on Climate Change and other changes

---

122 OECD, Principles for Evaluation of Development Assistance, Development Assistance Committee OECD, Paris, France, 1991, page five. https://www.oecd.org/dac/evaluation/2755284.pdf

123 OECD DAC, Glossary of Key Terms in Evaluation and Results Based Management. OECD, Paris, France, 2002.

124 The term "intervention" is used in the new criteria document to refer to the subject of the evaluation. Intervention encompasses all the different types of efforts that may be evaluated using the criteria, such as a project, programme, policy, strategy, thematic area, technical assistance, policy advice, an institution, financing mechanism, instrument or other activity. The criteria can be used to evaluate international cooperation activities, as well as the interventions of private sector actors, non-government actors, and national or local governments in domestic policy contexts. (Source: OECD, 2019)

in the development landscape. The current context for evaluating sustainable development looks very different from the context in 1991, when the criteria were first laid out. Today, the development evaluation landscape reflects new political priorities – including addressing migration and climate change, the rise of new development actors, the different characteristics of poverty and fragile States – and an increasing recognition that development cooperation has multiple objectives, including poverty reduction, the pursuit of national interest and supporting private sector investments.[125]

The desire to revisit the criteria also stemmed from experience with implementation and a request of the OECD DAC.[126] The adaptation process was managed by the OECD (EvalNet Secretariat) and involved a public survey; document reviews; stakeholder and expert interviews; major events in Africa, Europe and Asia; blog posts and online discussions; and consultations with evaluation practitioners in the United Nations Evaluation Group and other evaluation networks.[127]

THE CRITERIA BEYOND THE DAC: THE CONTEXT OF NATIONAL EVALUATION CAPACITIES

Though originally developed by and for donor agencies evaluating international development cooperation activities – with a focus on projects – the criteria have since been much more widely applied, both beyond projects and beyond development cooperation. They are mentioned in the national evaluation mandate of the Government of the Philippines, for example. The language has also begun to shift, with the new criteria definitions dropping terms like "donor" and "recipient".

This shift to a more global, shared practice of development evaluation was demonstrated by the fact that the criteria adaptation process was bookended by two NEC conferences. The global consultation was launched at a plenary session during the NEC in Turkey in October 2017, and the results were first unveiled at the NEC in Hurghada, Egypt in October 2019, before being approved by the DAC EvalNet in November. Members of the NEC community were both important contributors to the adaptation process and will be key actors in the implementation of the new definitions and principles. The DAC EvalNet – made up of the central, independent evaluation units of bilateral development agencies/ministries and key multilateral partners – continues to play a valuable

125 Mckee, Blampied, Mitchell, Rogerson 2019. "Revisiting Aid Effectiveness: A New Framework and Set of Measures for Assessing Aid "Quality." CGD Working Paper 524. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/revisiting-aid-effectiveness-new-framework-and-set-measures-assessing-aid

126 OECD DAC, A New DAC: Innovations for the 2030 Agenda. DAC High Level Communique, 31st October 2017. OECD, Paris, France, 2017.

127 OECD/DAC Network on Development Evaluation, OECD DAC Evaluation Criteria: Summary of consultation responses. OECD, Paris, France, 2018. https://ieg.worldbankgroup.org/sites/default/files/Data/DAC-Criteria/ConsultationReport_EvaluationCriteria.pdf

normative role (see box on page 140) and its members are increasingly working with peers in national evaluation systems around the world. This theme of evaluation as a global enterprise ran throughout the criteria consultation process, highlighting how evaluation practice has changed in the last 30 years.

Three other themes stood out as relevant for those working on national evaluation capacities.

KEY MESSAGES FROM THE CRITERIA DISCUSSIONS

THE CRITERIA ARE A GOOD STARTING POINT

The consultation on the criteria revealed that they are widely appreciated for their simplicity and conceptual clarity. The criteria are viewed as a good starting point for asking the right evaluation questions. Drawing on the experiences of Nepal, the Philippines, Sri Lanka and other countries, Mr. Velayuthan – who has championed the criteria in his native Sri Lanka – points out that use of the criteria beyond DAC members is not imposed, but demand-driven. Different actors have picked them up because they are useful in their local contexts. Specifically, the criteria have helped to shift attention beyond inputs and activities, to focus more on real changes in the well-being of people and the planet.

At the same time, the criteria are just a starting point; they say nothing about how evaluations should go about answering these questions. Take, for example, the analysis of unintended effects. The word "unintended" and the coverage of both positive and negative effects have been in the definition of the impact criterion since the beginning. Their inclusion serves as a reminder that despite the best intentions, development interventions can – and do – cause harm. Implementation in diverse contexts can also result in innovations and local adaptations that create unexpected positive effects. Using the impact criterion to ask good questions about these unexpected outcomes can thus help generate valuable insights.

However, a meta-evaluation of evaluations by the United States Agency for International Development shows that only 15 percent covered unintended effects. One of three evaluations by the Norwegian Agency for Development Cooperation does not mention unintended effects, even if it was explicitly mentioned in the terms of reference.[128] The criteria are broad. Even when they are included systematically in evaluations, not all dimensions will be covered. The criteria definitions themselves are therefore not sufficient to ensure all important perspectives are captured. Specific guidance based on institutional ways of working and priorities needs to accompany the

---

128   Ministry of Foreign Affairs, the Netherlands. Preliminary take-aways of the Unintended Effects of International Cooperation Conference of January 16th & 17th 2017, The Hague. IOB Evaluatie, The Hague, the Netherlands, 2017.

DEVELOPMENT EVALUATION NORMS AND STANDARDS

The criteria are one component of a broader set of norms and standards developed by the OECD DAC Network on Development Evaluation (EvalNet) with the aim of supporting improved evaluation policy and practice for accountability and learning. Shared standards contribute to harmonised approaches in line with the commitments of the Paris Declaration on Aid Effectiveness. The body of norms and standards is based on experience and evolves over time.

- **Principles for evaluation of development assistance** are at the heart of the EvalNet approach to evaluation. The principles focus on the management and institutional arrangements of the evaluation system.

- **Quality standards for development evaluation** provide guidance on evaluation process and product.

- **Glossary of Key Terms in Evaluation and Results-Based Management (second edition forthcoming),** available in 17 languages, defines key terms including results, outcome and different types of evaluation, as well as the evaluation criteria.

criteria. Evaluation systems must create the right incentives to cover those dimensions of each criterion that are of most importance and question common assumptions.

EVALUATION PROGRAMMING AND PURPOSE REQUIRE MORE THOUGHT

Another aspect of ensuring proper coverage relates to evaluation programming. During the plenary discussion of the criteria in Hurghada, Dr. Bagele Chilisa raised the important question of who has a say in what is evaluated. Local communities are too often excluded from these processes. Others point out that the criteria are not used sufficiently to evaluate concerns of equity and those "left furthest behind". These gaps relate to the selection of evaluation topics, an aspect of evaluation systems that does not always receive sufficient attention.

The process of elaborating an evaluation programme – what topics, strategies and interventions will be evaluated individually or collectively – is a critical element of an effective evaluation system. National evaluation systems must tackle this issue by setting out their own evaluation agendas linked to national sustainable development priorities. This requires being selective about how scarce evaluation resources are used to address learning and accountability needs across the range of interventions, potentially contributing to achieving the Sustainable Development Goals. Good

evaluation programming goes hand in hand with being clear about the overarching purpose of the evaluation function and of individual evaluations. As described in the new criteria principles for use, the evaluation purpose is the primary driver of the evaluation criteria, questions, methods and processes. National policies can be used to strengthen local involvement in identifying priority evaluation needs and translating these into a strategic learning and accountability programme.

WHO OWNS EVALUATION? A CHANGING LANDSCAPE

Perhaps the most widely heard message in the criteria consultation was that, while the criteria are key, what really matters is how they are used and by whom. The Paris Declaration on Aid Effectiveness already made it clear in 2005 that monitoring and evaluation systems should shift from donor accountability to mutual accountability between development partners, and be owned and led by national stakeholders. Over recent decades, and thanks to the work of many, including the NEC conferences, practice continued to shift, from the involvement of partners in donor evaluations to joint evaluations, country-led evaluations and finally to the universal 2030 Agenda.

The 2030 Agenda rightly positions development cooperation within the national, regional and global drivers of sustainable development, sharpening our collective focus on the need to evaluate much more effectively across the public sector in all countries. National systems critically look beyond aid, capturing the impacts of domestic spending, private financial flows and other government policies. The universal nature of the 2030 Agenda reminds us that no country has fully achieved equitable, climate-safe, sustainable development. Likewise, there is no country that cannot improve the use of evidence and evaluation in public policymaking to speed its progress.

The experience of adapting the criteria suggests however that the outdated donor/recipient paradigm of aid-centric evaluation is still not entirely behind us, and more effort is needed to continue the shift to nationally-led evaluation. Supporters of national evaluation systems can help by focusing on adapting and translating the new definitions and principles into their national contexts and institutions. Evaluation partners in OECD countries, including members of the DAC EvalNet, can continue to play a supporting role, for instance through capacity development efforts and engaging in joint evaluations.

CONCLUSION

The banner of the criteria adaptation process was "better criteria for better evaluations", drawing our attention to the overarching purpose of this exercise, which was to improve evaluation's contribution to achieving the 2030 Agenda and addressing the climate crisis. As described here, more work is needed to support thoughtful use and focus on the most important dimensions of each of the criteria in the local context. Scarce evaluation

Reflections on the revised DAC criteria

BAGELE CHILISA
PROFESSOR, POST-GRADUATE RESEARCH AND EVALUATION PROGRAM,
UNIVERSITY OF BOTSWANA AND MEMBER, UNDP INDEPENDENT EVALUATION
OFFICE EVALUATION ADVISORY PANEL

Plenary 3, "Transforming evaluation for transformative development – new Evaluation Criteria", discussed the current thinking on the DAC criteria, including emerging definitions and principles for use, with reflections and critiques from different perspectives. Panel speaker Bagele Chilisa reflected on the DAC criteria from the unique perspective of local communities.

"I think Megan did an excellent job of explaining the criteria, and I think they are good criteria. My question is, what worldview informs these criteria? How different are they from the old ones? From where I'm standing, it looks like they are the same thing. For me, the criteria look more like criteria stemming from the methods branch, concerned with the accuracy or precision of methods in social science research. This seems to be the trend especially in international organizations, where there seems to be an underlying belief that transformation in evaluation can come through the use of big data, artificial algorithms, Bayesian statistics and new technologies. This is made worse by the tendency to assume political neutrality of methodologies.

"Sometimes I think language can mislead people. A transformative tool needs a transformative language. I looked, for example, at the criterion of relevance and I said to myself, relevance is just a common word, it's not radical. We say the criteria have changed, that they are 'better'. However, when I look at them, they are still the same.

"I tried to look at the way relevance is used. I had the privilege of reviewing a few of the UNDP IEO evaluation reports, and what is interesting is that the word relevance is part of the narrative for almost every evaluator. Every evaluator writes that the intervention is relevant. I am of the view that relevance is redundant, because almost everybody is going to write that the intervention is relevant.

"How do we, for example, address the African culture and values in the new criteria? Where would they fit? Where would community resilience – part of what is going on in our communities, e.g., interventions on peace, poverty eradication and so on – fit? Where would one account for a community's time? If I'm an evaluator and I wanted to show the cultural strength that the community brings to the intervention, where do I factor that in?

"Often, we talk of unintended outcomes. The malaria interventions in Africa, for example, include the distribution of mosquito nets, which have unintended outcomes. In some interventions, husbands take the mosquito nets and use them for fishing, and wives fight over them because they want the mosquito nets to protect their kids. Yet, these are narratives that you rarely find in evaluation reports.

"How do we intend to get these narratives out, so that we can also begin to accommodate other knowledge systems? Where do African values, cultures, innovativeness, harmony, connectedness, relationality, togetherness become integrated? What part do they play, and in which criteria within the five that were presented do they fit? Evaluation is political; we cannot run away from it. Where do the criteria interrogate historical power? For example, colonization and the tendency to impose the use of only tried measures. Where does political power fit in the criteria? What does relevance mean, when partners fight over what is relevant? Whose priorities do interventions address? Whose priorities and values matter?

"Societal power come in to play when the beneficiaries say, 'we cannot use this clinic because it is built where our ancestors were buried.' When are we as evaluators going to interrogate these dimensions of context? As I said, most of the time when we say context, we are talking of anthropological descriptions of who the people are. When do we address relational power, institutional dynamics and power structures?

"In some of the reports I have read, the evaluator would state that the intervention does not work, or it did not work as intended. The evaluator glosses over these power dynamics. It is as if they are not supposed to be interrogated and yet these are the things that matter. Certain powers force those that are at the bottom to remain at the bottom. We look too much into objectives, while contexts and community priorities are not addressed in detail.

"To make the criteria more exciting, I have one suggestion. The criteria that you call relevance, is a very important one, but I would say that we use a different term. Borrowing from evaluators who proposed an evaluation branch called context,[129] I propose that a new evaluation branch called context/aspiration/needs form the umbrella for relevance. See the illustration below.

"And if we do that, it will put pressure on evaluators to conduct culturally-responsive evaluation. Although evaluators often state they are using mixed methods, I find that evaluations are invariably quantitative. There are no qualitative data. In other words, the beneficiaries cannot speak. They cannot speak in their own voices. There is a lot about missing data in most of the evaluation reports I have read. My take is, okay, there's missing data, so yes, and that is a problem. Data collection in some of the developing countries is a problem. It's not going to improve very soon, but the people are there to tell their stories. Why not take advantage of the people and let them tell their stories?

A five-branch tree of evaluation approaches
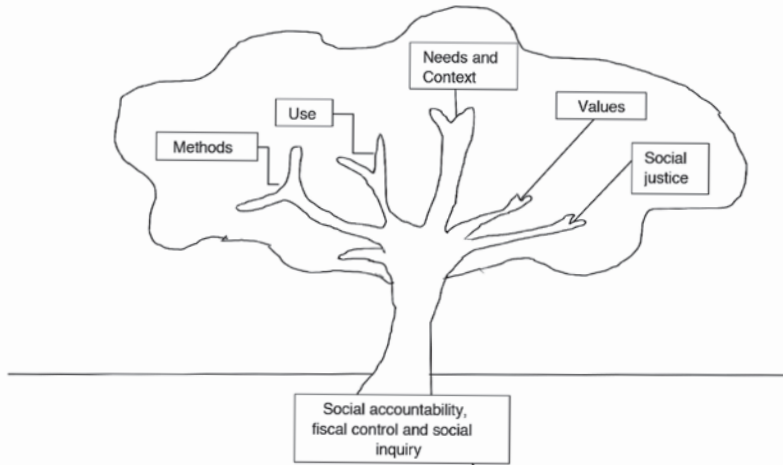
Source: Chilisa, B. (2020). Indigenous Research Methodologies, Thousand Oaks, CA: Sage

"In conclusion, I would like to say that the criteria are a good beginning, but we still need to bring in untried measures. We still need to bring in marginalized local indigenous knowledge. We still need to put more emphasis on the beneficiaries as people who can solve their own problems. A lot of times the interventions are just, but the criteria do very little to compel the evaluation regime to respond to the needs and priorities of the people in low- and middle-income countries. We need to think outside the four-branch evaluation tree and include a fifth branch that addresses context, needs and priorities of those whose knowledge systems have been excluded from the evaluation discourse. That is relevance in context."

ADDITIONAL REFERENCES

Chilisa, B. (2020). Indigenous Research Methodologies, Thousand Oaks, CA: Sage.

Ofir, Z., & Shiva, K. S. (2012). Evaluation in developing countries: What makes it different? In S. I, Donaldson, T. Azzam, & Conner, R. F. (Eds.), Emerging practices in international development evaluation (pp. 11-24). Charlotte, NC: Information Age.

resources should be used to support national evaluation priorities and meet local learning and accountability needs.

The process of adapting the evaluation criteria provides hope for the evaluation field. It showed how the growing international evaluation community could make progress when evaluators with diverse perspectives work across institutional, national and political differences to strengthen the core evaluation principles and concepts shared by all. Future NEC conferences can carry this work forward and help evaluation realize its full potential.

ADDITIONAL REFERENCES

Evans, Alison, "A Pivotal Year for Monitoring and Evaluation - Building and improving monitoring and evaluation capacity worldwide is key to achieving the Sustainable Development Goals within ten years.", World Bank IEG, Washington DC, January 7, 2020.

OECD/DAC Network on Development Evaluation, Quality Standards for Development Evaluation. OECD, Paris, France, 2010.

OECD/DAC Network on Development Evaluation, Better Criteria for Better Evaluation Revised Evaluation Criteria definitions and Principles for Use. OECD, Paris, France, 2019, page five.

United Nations, Transforming Our World: the 2030 Agenda for Sustainable Development (A/RES/70/01). United Nations, 2015.

Reflections on the revised DAC criteria, Bagele Chilisa, Professor, Post-Graduate Research and Evaluation Program, University of Botswana and member, UNDP Independent Evaluation Office Evaluation Advisory Panel