

2. Rethinking the DAC Criteria for the SDG Era

CAROLINE HEIDER

Director General
Independent Evaluation Group, World Bank

INDRAN NAIDOO

Director, Independent Evaluation Office, UNDP
and Vice Chair, United Nations Evaluation Group

RIITTA OKSANEN

Senior Advisor, Development Evaluation, Ministry for Foreign Affairs, Finland
and President (2016-2017), European Evaluation Society

SUSANNE FRUEH

Director, Internal Oversight Service, UNESCO
and Chair, United Nations Evaluation Group

PER ØYVIND BASTØE

Director, Evaluation Department
Norwegian Agency for Development Cooperation and Chair, OECD DAC



In 1991, the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) Evaluation Group set out evaluation criteria, five of which—relevance, effectiveness, efficiency, impact and sustainability—continue to be widely used and applied in evaluations today. The United Nations Evaluation Group (UNEG) incorporated the OECD/DAC criteria, which had been in use for years, in the norms and standards that it adopted in 2005. In 2015, the world adopted the Sustainable Development Goals (SDGs). This new agenda for “transforming our world” is of “unprecedented scope and significance”,¹⁶⁰ and the plans for follow-up and review processes include an emphasis on evaluations and data which are of high quality. Development practitioners and evaluators are discussing whether the five traditional evaluation criteria are sufficient for evaluations in the SDG era.

Caroline Heider, Director General of the Independent Evaluation Group of the World Bank, took these questions forward in the blog series, “What Works”, asking, among others, if in the light of the SDGs, shifts in norms and values, increasing complexity and the pace at which technology is developing, “have we had enough of R/E/E/I/S?” and “is relevance still

160 A/RES/70/1, para. 5.

relevant?”¹⁶¹ At the National Evaluation Capacities Conference (NEC) 2017, she led a new discussion, asking a panel of evaluation experts if we need to look at the DAC evaluation criteria and update them, rethink them or implement them differently. This paper presents excerpts from this conversation.

Ms. Caroline Heider: Today's event is not intended to put forth proposals, but rather to engage in a conversation. Are we asking and answering the right evaluation questions? Are these questions strategic? Is evaluation informing a strategic conversation around development and development results, around choices that policymakers need to reflect on? Are we helping with transformational change or are we providing routine evaluations that reflect primarily on things that were done as planned? Are we asking the question: 'Were the right things done or were they done in the right way? Does the local context, does the local value system matter in evaluation and how far are our criteria actually helping to address these things?'

To discuss these questions, we have Riitta Oksanen, the president of the European Evaluation Society (EES). She is from Finland and can speak from a bilateral donor perspective as well as that of the EES.

Next to her is Per Bastøe, the chair of the DAC Evaluation Network, which was at the centre of the discussion and articulation of the criteria in the beginning. He will represent the network in today's discussion, but he is also the Evaluation Director for the Norwegian Agency for Development Cooperation. He has also worked with various development banks. Susanne Frueh is the chair of the UNEG, and has experience with many United Nations organizations, the Inter-American Development Bank and other development and humanitarian agencies. And last but not least, we have Indran Naidoo, the host of the NEC series of conferences and the head of the United Nations Development Programme (UNDP) Independent Evaluation Office.

To begin, reflecting on the past years where we have used the DAC evaluation criteria, what is the one that thing stands out? What did these criteria achieve? In other words, why were the criteria important and what matters about them?

Mr. Indran Naidoo: The main benefit of the DAC criteria has been the resulting consistency of approach in conducting evaluations. An evaluator offers a professional judgement based on evidence. And one can only do so if one has benchmarks. The value added of the criteria is the reference points they provide, which has seen a great benefit in the consistent shaping of evaluations.

Ms. Susanne Frueh: For the United Nations system, the DAC criteria have raised the game in evaluation. The United Nations now uses the criteria in its definition of evaluation. The DAC criteria created a common language, and increased coherence and credibility of evaluations. All in all, the criteria have been a very positive contribution to the field.

161 The complete blog series is published in World Bank Independent Evaluation Group, 'ReThinking Evaluation', Washington, D.C. 2017 (<http://ieg.worldbankgroup.org/sites/default/files/Data/RethinkingEvaluation.pdf>).

Mr. Per Øyvind Bastøe: It is important to note that evaluation is not commonly understood the same way everywhere, and the criteria therefore have real importance. The criteria ensure that we are talking about the same things, not only in evaluations in the development field, but in other fields where evaluations are undertaken.

Ms. Riitta Oksanen: It always has been important to focus not only on the DAC criteria, but to remember that they are an element within a broader set of norms and standards. The criteria guide us on *what* to evaluate, but in addition, there are principles that guide us on *how* to evaluate. These standards guide us on what constitute a high-quality evaluation process and high-quality evaluation products. It is important for me to see the criteria in this context. For Finland, these norms and standards have had tremendous value. We have been able to benchmark our evaluation function against these norms and standards, and we have been able to develop a relatively sound development evaluation function. For a small donor, it has been a huge benefit to have that support from the international community. Without the DAC standards, we would not be where we are today.

Ms. Caroline Heider: Following on from these fairly uniform views—that the DAC criteria provide across-the-board standardization, harmonization, and opportunities to up the game—what are the current challenges and future challenges? Do the SDGs pose a new reality that requires evaluators to rethink how we work or how we value what the development community is trying to achieve and to deliver? The SDGs are a prominent feature in the United Nations now. In the UNEG, there must have been a lot of discussions on the consequences of the SDGs for the United Nations system. In those discussions and for the evaluation thereof, have the criteria been an important topic or are there other things that are more important?

Ms. Susanne Frueh: There has been a somewhat simplistic adoption of the DAC criteria in the United Nations. They are clear and they resonate, but there is more to evaluation than five DAC criteria. For example, in the humanitarian context, following the tsunami response in 2005, we talked about additional criteria such as connectedness, as in connecting emergency response to recovery to development. We also talked about coherence. Thus, we had already started discussions about adapting the criteria. When the SDGs were adopted, the United Nations took the opportunity to reflect on its own norms and standards. The 2005 UNEG norms and standards set out the way United Nations evaluation should be conducted. With the evolution in context and the adoption of the SDGs, we took it upon ourselves to update these norms and standards. It was a painful process, as more than 50 different institutional actors had to agree on what was most important. The norms and standards already were fairly strong on human rights and gender, and we further strengthened these aspects in the revised norms and standards so that they are in line with the SDGs. We had a very healthy debate on the need to address environmental sustainability as a cross-cutting norm or standard. We didn't quite get there, but we recognize that this issue must be addressed. I do believe we still have some other issues to grapple with as well. For instance, the importance of culture. Culture is a driver of change and also an impediment to change. If we miss the cultural elements which effect change, we miss the big picture or miss key information

on how to influence desirable change. The norms and standards should be a living document. We need to improve as we go along and we need to make sure we continue to ask the right questions.

UNEG also commissioned a study in 2016 looking at the SDGs and lessons, challenges and opportunities for evaluation. The report proposed seven new criteria for new evaluations in the SDG era: equity, gender, human rights, inclusiveness, participation and partnership. There have also been calls to include additional criteria for humanitarian assistance and for the environment. This however is probably too broad, and we run the risk of a smorgasbord approach where we try to do everything under the sun and then we don't do anything at the end of the day, or don't do anything right.

But I do think it worth asking the question again, what can we do better, based on the SDGs, and how can we make sure that the principle of "no one left behind" is fully incorporated in all the questions we ask.

Ms. Caroline Heider: The DAC Evaluation Network is the forum where the conversation started. Per Bastøe, do you see a similar conversation happening within the DAC network now? Is the DAC going to step up to the challenge to update the criteria, should that be necessary?

Mr. Per Øyvind Bastøe: The short answer is yes. The OECD DAC Evaluation Network has about 40 different member organizations and it is still an important forum for discussing evaluation standards and criteria. However, it has been 15 years or more since the last round of discussions took place around these basic criteria. In addition, there are many misunderstandings surrounding the criteria. Some see them as a straitjacket, whereby you need to use for all criteria for everything. Others do not fully understand that other standards have also been developed, as mentioned in the discussion of humanitarian evaluation. The criteria are not meant to be a guide everywhere, all the time. We continue to have this debate in the Evaluation Network. However, it is important that EvalNet not be alone in this game. Twenty years ago, it may have been the case, but now all of you need to be involved: United Nations organizations and evaluation societies, the UNEG, the multi-development banks. Evaluators in 2017 are a different group than in the late 1990s. We need to own these criteria as an evaluation community.

Ms. Caroline Heider: It is therefore important to avoid becoming mechanical. Rather, there is a need to keep this conversation alive and adapt. Riitta Oksanen, could you elaborate on the system in Finland? How does Finland's experience inform the discussion on revising the DAC criteria?

Ms. Riitta Oksanen: My belief is that if we had followed the whole set of DAC norms and standard as faithfully as we have the criteria, we would not have as many problems as we do. *How* evaluations are done is important, not just *what* is evaluated. Finland has been able to integrate the 2030 Agenda in its national planning system, building on a long tradition of working for sustainable development. The pre-2015 foundation has been important. This integration has been possible because there is high-level political commitment and a genuine commitment to work together with stakeholders, including citizens and parliament.

However, evaluation is lagging behind. Based on the 2030 Agenda, evaluation has been flagged and understood to be part of the management system. Yet, we are still in the process of trying to understand what it will mean in practical terms. In Finland, we have a decentralized evaluation system where the ministries have their own mechanisms for evaluation. It is only during recent years that public-sector evaluators have started to come together and discuss what a government policy on evaluation, as a whole, would mean. We don't have a central evaluation policy.

To develop a public policy evaluation system in Finland, we don't have the same kind of framework that we had for development evaluation. There are no norms and standards for public policy evaluation anywhere. It would be almost unthinkable for my Government to take the OECD DAC norms and standards, meant for development cooperation, into public policy evaluations.

Ms. Caroline Heider: This is an excellent point at which to turn to Indran Naidoo. Riitta Oksanen has pointed to the idea that the DAC norms and standards apply to Finland's support for development cooperation programmes, but not to their national policies. These NEC conferences are designed to look not only at evaluation of development programmes, but capacities to evaluate all that happens in partner countries. Over the course of the past five conferences, has there been a shift in how we think of criteria, how we talk about criteria and how they might need to be revised or updated?

Mr. Indran Naidoo: From the first National Evaluation Capacities Conference in Morocco until now, the idea has been to create space for conversations with both evaluators and those governments that want to strengthen evaluation. Initially, the focus was on understanding what kind of enabling environment is required for evaluation to happen. Thus, the early conversation was about the utility of evaluation for decision-making. In South Africa, we shifted the discussion to public policy. We then looked at principles with respect to development and at the present event in terms of the SDGs. Of the five conferences, the conversations in Brazil on independence, credibility and use touched the most on criteria. The issue of criteria hasn't been a consistent stream over the last 10 years, but one we would like to address more directly.

With respect to work within my own agency, UNDP, the Independent Evaluation Office uses the DAC criteria in its evaluations of UNDP country programmes, and where necessary, adds in additional criteria, for example in our thematic evaluations. We make explicit reference to the criteria and frame our evaluations around them. UNDP programme units also conduct evaluations, which we refer to as decentralized evaluations. Our office assesses the quality of these evaluations, including with reference to the DAC criteria. Through this process, we see that there is great variation of understanding from the evaluators conducting these evaluations. For example, with relevance, the understanding of relevance by an evaluator in one country may be quite different than another evaluator in another country. This means there is still work to be done, and we are starting a conversation to strengthen a common understanding of the criteria. Now that we are moving towards the SDGs, we shouldn't throw the baby out with the bathwater. The criteria are still important. They still give us a

reference point from which to work. We need to be careful that when we have a deeper conversation, we don't confuse the criteria with principles, norms and standards.

There is also a danger of creeping incrementalism. We add more and more criteria, and reports become longer. Sometimes evaluation reports are difficult to read, because they have a chapter on each criterion, and within that, sub-criteria, so it is almost impossible to read. The principle of criteria remains; what is more important is how we apply them. We want to see this remain on the agenda, including at NEC conferences, to move to total commonality, a similar understanding of criteria, and their use and their methodological application.

Ms. Caroline Heider: I would like to come back to the question of coherence. Riitta Oksanen mentioned the difference between domestic and international development assistance policies, and Susanne Frueh spoke of policy coherence as an issue arising in the evaluation of the tsunami response in 2005. Twelve years on, are we getting closer to policy coherence and evaluating it, or is something standing in our way?

Ms. Riitta Oksanen: The simple answer is that we must. Reflecting on the Finnish experience in development, we have seen the OECD DAC criteria as providing a framework on which we have based our development evaluation work. However, this has always been just a base. Many years ago, we expanded on the criteria in two important areas that we felt were not adequately reflected in the DAC criteria: evaluation ethics; and gender-responsive and human rights-based evaluation processes. We borrowed these from the UNEG norms and standards. Where the DAC criteria did not meet our needs, we looked for other resources.

Finland is a member of the European Union. In the European Union, evaluation of development cooperation work has long incorporated the "3 Cs": coordination, cooperation, and coherence. Finland has integrated coherence into its set of criteria. Note that once again, this applies to international development evaluation. On the national front, we are not there yet.

Ms. Caroline Heider: Turning to the SDGs again, many questions arise. There are synergies between many goals, but there are others that may compete for resources. Labeling the goals "sustainable" may not actually result in sustainability if trade-offs that are necessary to achieve the goals are not considered. People are also questioning how to assess impact and asking if the impact criterion needs to be updated. Indran Naidoo, what are your thoughts? What must we—as an evaluation community—do to address the dual challenge of impact and end results on the one hand, and on the other, synergies either creating or detracting from greater impact?

Mr. Indran Naidoo: First of all, clearly the work of evaluators has become more complex. If we go back 20 years, a lot of the monitoring and evaluation work was done at the project level. The scope was easily defined, the timeline was defined and evaluators could articulate clear findings, conclusions and recommendations. The move towards a globe where there is ever greater interconnectedness and ever more complex movement goes to the question of impact. We use that term very carefully in the United Nations system and within the UNEG, we have had many conversations on impact methodologies. Yet, it is difficult to discern an impact within the time horizon of most interventions in United Nations programmes.

Secondly, with respect to the SDGs, I think we're only now beginning to understand how difficult it is going to be to ensure that evaluation is meaningful for the SDGs. The SDGs are illustrated in 17 boxes that appear to be the same size, but within each of these boxes, there are different constructs and different sources and types of data. In addition, different SDGs interact with different goals. Evaluation criteria—which tend to be static—need to account for dynamic interactions. We as evaluators are not used to doing this. We still tend to think in linear terms, where if you have a good input and good management, you will have a good output, and if you have enough outputs at the project level, they will eventually lead to a programme and policy, and ultimately change the world. However, what do we see? Inequalities remain. Intra-regional inequality has increased. Human rights are not taken as seriously as they should be. The United Nations pushes a normative agenda, advocating for intervention wherever progress towards a better world can be achieved; measuring this is not easy. Formulating the UNEG SDG road map took over a year, as it is so complex.

Ms. Caroline Heider: Another dimension of the SDGs that often gets lost is sustainability. Environmental, economic and social sustainability are embedded in the SDGs, but there are unanswered questions. How do we provide more services but consume less? Take the fairly straightforward dimension of energy: we want everyone to have access to energy, but we want that energy to be more efficient. Are we equipped to evaluate something that doesn't occur? Do our criteria help us do that?

Ms. Susanne Frueh: Indeed, the criteria tend to lead us in terms of what we look for and expect to see. There is a need to move beyond the criteria and to look for what is not visible and also what has not yet occurred. If we cannot look more deeply and understand the dynamics and context of what drives change, we will not be able to provide our professional judgement. I think it is our role as evaluators to help the process. If you ask, for example, what success looks like, success for me as an evaluation community is that we are at the table, that we get to inform the 2030 Agenda and that it is being achieved. We have 13 years to do this. Unless we ask the right questions, unless we unpack and demystify the concepts, unless we look for the interlinkages, we will not be able to do that. We need to step up our game. We need to build on the criteria but we also need to ask the right questions.

Ms. Caroline Heider: To sum up, I would like to ask each of the panel members to take a stance. We are evaluators. We make judgments and we come to conclusions. Given this whole conversation including questions from the audience, do we need to completely reform the DAC criteria? Do we just need to update them, tweak them a little bit? Or do we simply need to apply them better?

Ms. Riitta Oksanen: When it comes to development evaluation, we need to apply the criteria better. We need to use them in a flexible manner given the context. However, a whole different issue is, what should be the criteria whereby we evaluate our public policies. This is an extremely important issue because this is also an issue about power. In this case, what is the right forum for agreeing on those criteria? My feeling is that it is not the OECD DAC.

Mr. Per Øyvind Bastøe: We need to clarify what we mean by these criteria and what we see as evaluation standards. For instance, when we talk about independence, credibility and utility, we also need greater clarity. I do think we need to examine and probably revise some of the elements in the five criteria. For instance, impact is not a clear term. Perhaps it was clear at the end of the 1990s, but I struggle to fully understand what impact is, and what the term implies. We need to revisit the criteria and we need to clarify them. We need to apply them, and we need to understand them. Finally, we need to be pragmatic with respect to this effort.

Ms. Susanne Frueh: Coming back to the baby and the bathwater, I think we should check the temperature and maybe raise the water level or reduce it, but I do believe that we have a good foundation. As Per Bastøe has said, we need to ask some fundamental questions. I think relevance is one of the most misunderstood criteria. We can bring in coherence, we can bring in other issues, we can also look at some of the paradigms such as “no one left behind,” equitable development and sustainability, and as we go through the criteria, weave in everything we need to ask.

Finally, I think there’s an element that many evaluations miss, that is, design. Design is very often the origin of all bad that happens afterwards. Perhaps that sounds a bit technocratic, but I do think that we need to reflect more on why we are doing certain things. Are we doing the right things from the get-go or just coming in at the end of the evaluation to say the design was poor? Can we unpack this to identify additional criteria? For me, the solution is to keep the criteria, refine them, and explain them better. And make sure that the new paradigms are reflected throughout.

Mr. Indran Naidoo: As evaluators, we need criteria. These are important because we need a frame of reference. We need to add as we go along, depending on what the subject and topic is. We definitely need to apply criteria better, unpacking them across space and taking scale into account. An evaluation at the project level is quite different than one at the programme or policy level. Across all levels, we first need to look into the principle of independence, because when we ask the essential question of what an organization is doing, independence makes the message even stronger.

Secondly, we need to take into account mandate. If the organization’s mandate includes a normative element, as in UNDP, which aims for a better life for all, reducing poverty and improving governance, this requires examining principles and norms and identifying the right questions. There are exciting times ahead as we continue these conversations.

Ms. Caroline Heider: From my perspective, we have more work to do. There are incredible opportunities where we can sharpen our evaluation instruments including the criteria. We can use criteria to incentivize different behaviors: when we signal what we evaluate, people pay attention and think about the importance of those criteria. Is this throwing the baby out with the bathwater? I certainly don’t think so. Rather, it is building on the strong foundation that the DAC criteria prepared, while keeping our evaluation practice dynamic and growing with the times, so that we are prepared to serve a strategic purpose all the way to 2030 and beyond.