The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement

Thomaz Chianca
Independent Consultant
Brazil

The five evaluation criteria from the Development Assistance Committee of the Economic Cooperation and Development (OECD/DAC) have been a strong foundation for international development evaluation since 1991. They have been the most prominent and widely adopted criteria used for aid evaluation by most bilateral and multilateral donor agencies, as well as international non-governmental organizations (INGOs). However, critiques of the quality of development aid evaluation are still abundant. Thus, it is reasonable to question how those criteria can be improved. This paper provides a critical look at the five DAC criteria and proposes major recommendations for changes, including revisions of definitions, addition of key missing criteria, and discussions about the level of importance of the criteria. If followed, these changes could contribute for increasing the quality of evaluations for the purposes of (re)funding, program changes, and other decision options.

Efforts in the direction of establishing guidelines, standards and/or criteria for improving evaluation practice within the development sector are longstanding trends.[1] The World Bank's Operations Evaluation Department was certainly one of the pioneers in this area. Specifically, in 1976, this department issued the "Standards and Procedures for Operations Evaluation", which provided specific guidance for the evaluation processes conducted at the end of a project, the Project Completion Reports, and after a few years of project completion, the Project Performance Audit Reports (Willoughby, 2003, p. 11). However, to date, the evaluation standards for development aid, established in 1991 by the Development Assistance Committee (DAC) from the Organization for Economic Cooperation and Development (OECD), have been by far the most influential work in the field of development evaluation.

DAC was established by the OECD to improve cooperation between the governments of its 30 members (the most affluent nations in the world such as the U.S., Japan, and Germany) and governments of developing or transitional countries. In late 1992, the OECD/DAC released a document (OEDC, 1992) devising key principles for aid management. Monitoring and evaluation functions formed a substantial part of those principles.

Since their inception, the OECD/DAC evaluation guidelines have shaped the way most donor agencies and their clients/grantees commission or design and conduct program evaluations. These guidelines are based in six general principles:

[1] Throughout this paper, *international development evaluation* and *aid evaluation* are used interchangeably.

ISSN 1556-8180
March 2008

1. All aid agencies should have an evaluation policy.
2. Evaluations should be impartial and independent.
3. Evaluation results should be widely disseminated.
4. Evaluation should be used—feedback to decision-makers is essential.
5. Donor and recipient agencies should be partners/cooperate with the evaluation—strengthen recipient agencies and reduce administrative burden.
6. Evaluation should be part of the aid planning from the start—clear objectives are essential for an objective evaluation (p. 132).

The five criteria to evaluate development interventions (relevance, effectiveness, efficiency, impact, and sustainability) are undoubtedly the most known and adopted features that emerged from the OEDC/DAC evaluation guidelines.

The great acceptance and influence of the DAC criteria can be partially explained by the powerful and influential composition of its Committee. More than 30 heads of evaluation units from virtually all bilateral[2] and multilateral[3] agencies have a seat in the Committee. The agencies represented by these professionals have adopted the five criteria. Even though some of those agencies have introduced small adaptations, interpretations, or expansions, the underlying core ideas of the criteria have been maintained.

INGOs have also been affected by the DAC criteria partially because several of them operate grants from bilateral and multilateral donors and these funders request the integration of the five criteria into the INGO evaluations. There are signs, however, that some INGOs have also integrated the ideas of the DAC criteria independently from official requirements from donors. INGOs that traditionally do not operate with large direct support from donor agencies, such as Heifer Project International, have also adopted the five criteria as part of some of their requests for proposals (RFPs) for evaluations.

The establishment of the DAC criteria can be considered, at the time of its inception, a great step forward in the direction of improving the quality of development evaluations. These criteria shifted the focus of development evaluations away from solely assessing program outputs or use of funds according to what was proposed, or from the adoption of the economic rate of return (ERR)[4] estimation as the single criterion to assess an aid intervention. Instead, these criteria proposed considering a broader set of key elements.

The five criteria have been in use now for more than 15 years without going through any major revisions. Given their importance and level of influence in the field, it is pertinent that independent professionals take a critical look at them, especially since many scholars and practitioners consider that the quality of evaluations in development aid has been quite disappointing (ALNAP, 2006; Chianca, 2007; Clements, 2005; Goldenberg, 2001; Kruse et al., 1997; Leading Edge Group, 2007; Russon, 2005; Savedoff et al., 2006).

[2] Agencies representing a donor country and responsible for establishing individual cooperation efforts with low- or middle-income countries (e.g., US Agency for International Development—USAID, Swedish International Development Cooperation Agency—SIDA, UK Department for International Development—DFID)

[3] International agencies supported by several nations and responsible for coordinating cooperation among more than two states (e.g., the World Bank, the United Nations Development Program—UNDP, the African Development Bank)

[4] The interest rate at which the cost and benefits of a project, discounted over its life, are equal. (Business dictionary 2007) Generally speaking, the higher a project's internal rate of return, the more desirable it is to undertake the project. (Investopedia, 2007)

Defining the DAC Evaluation Criteria

The five DAC evaluation criteria are based on the conception that evaluation is an assessment "to determine the relevance and fulfillment of objectives, developmental efficiency, effectiveness, impact and sustainability" of efforts supported by aid agencies (OECD, 1992, p. 132). The OECD/DAC members view these criteria as essential in guiding development aid evaluation. The following are the current definitions of the criteria provided at the OECD/DAC (2006) Website:

> **Relevance:** The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor. In evaluating the relevance of a program or a project, it is useful to consider the following questions: To what extent are the objectives of the program still valid? Are the activities and outputs of the program consistent with the overall goal and the attainment of its objectives? Are the activities and outputs of the program consistent with the intended impacts and effects?
>
> **Effectiveness:** A measure of the extent to which an aid activity attains its objectives. In evaluating the effectiveness of a program or a project, it is useful to consider the following questions: To what extent were the objectives achieved or are likely to be achieved? What were the major factors influencing the achievement or non-achievement of the objectives?
>
> **Efficiency:** Efficiency measures the outputs—qualitative and quantitative—in relation to the inputs. It is an economic term which signifies that the aid uses the least costly resources possible in order to achieve the desired results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted. When evaluating the efficiency of a program or a project, it is useful to consider the following questions: Were activities cost-efficient? Were objectives achieved on time? Was the program or project implemented in the most efficient way compared to alternatives?
>
> **Impact:** The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators. The examination should be concerned with both intended and unintended results and must also include the positive and negative impact of external factors, such as changes in terms of trade and financial conditions. When evaluating the impact of a program or a project, it is useful to consider the following questions: What has happened as a result of the program or project? What real difference has the activity made to the beneficiaries? How many people have been affected?
>
> **Sustainability:** Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable. When evaluating the sustainability of a program or a project, it is useful to consider the following questions: To what extent did the benefits of a program or project

continue after donor funding ceased? What were the major factors which influenced the achievement or non-achievement of sustainability of the program or project? (pp. 1-2).

The five criteria tackle very important aspects of an evaluation. They have the relevant feature of being applicable to the ample range of aid interventions from single projects or groups of projects (programs), to large scale sector interventions (e.g., investment in a country/state health system) or the whole portfolio of interventions supported by a donor agency in a country or state. Also, these criteria are clearly more comprehensive than the set that was commonly used (and still is quite preponderant) to assess the work of international development agencies which comprise measuring outputs, monitoring resources' application, and, where more sophisticated, estimating a project's economic rate of return.[5]

Since its implementation, the DAC criteria have remained relatively unchanged. In 1998, a report was released by the OECD (1998) that included the results of a comprehensive study commissioned by the DAC Working Party on Aid Evaluation focusing on members' experiences with the application of the 1991 "Principles for Evaluation of Development Assistance."[6] The report concluded that the principles were still valid and sound. However, because of changes in the general aid context in many donor countries, the report suggested the need to rethink some of the interpretations and applications of the principles (p. 7).

Assessing the OECD/DAC Evaluation Criteria

Given the importance and level of influence of the DAC criteria in the development world, it is appropriate to submit them to independent scrutiny. Three sensible questions to orient a reflection on the five criteria include: (i) Are they sufficient to provide a sound assessment of the quality, value, and significance of an aid intervention? (ii) Are they necessary? and (iii) Are they equally important?

To address the first question is to consider whether key elements related to determining merit, worth or significance of an aid intervention were left out of the criteria definitions. To do so, the first step included a careful comparison between the DAC criteria and one of the most comprehensive and current set of program evaluation criteria proposed by Scriven (2007)—the Key Evaluation Checklist (KEC). The results from this initial exercise were critically reviewed and expanded by a group of 10 professional evaluators with broad experience in international development programs and diverse background (public health, community socio-economic development, management, engineering, public administration, political sciences, and education). These 10 professionals, currently pursuing doctoral degrees in evaluation through the Interdisciplinary Ph.D. in Evaluation program at Western Michigan University, created a taskforce on international development evaluation and conducted eight meetings over a 4-month period to specifically discuss improvements to the OECD/DAC evaluation criteria.

The overall conclusions were that:

- The definition of *relevance* currently focuses primarily on the *goals and priorities* of donors or country/local governments, instead of focusing on *meeting the needs* of the targeted population. This criterion should be

[5] ERR estimations are especially common in evaluations of interventions supported by the World Bank.

[6] Those are the six overall evaluation principles mentioned earlier in this paper (p. 2) under which the five OECD/DAC criteria were developed.

refocused to address the needs of the intervention's impactees.

- Similarly to relevance, the definition of *effectiveness* focuses on determining the extent to which the intervention met its *goals*, and not the *needs* of aid recipients. This criterion should be refocused or possibly subsumed under the *impact* criterion, since goals cover only the expected positive results from an intervention.
- The current definition of *sustainability* is limited to prospective (likelihood of) sustainability and do not make any reference to retrospective sustainability (how sustainable it has been). Furthermore, it only mentions the need to consider environmental and financial aspects of sustainability, leaving out other essential elements to the sustainability of interventions such as political support, cultural appropriateness, adequacy of technology, and institutional capacity.
- *Efficiency* even though tackling some of the right issues, falls short on the coverage of *costs* (e.g., non-monetary costs) and *comparisons* (e.g., creative alternatives). Furthermore, the term efficiency often gets defined as *least costly approach*, but it is a limited definition given the way evaluations are structured. Cost-effectiveness seems a better term to define this criterion.
- Two key criteria are missing: *quality of process* (e.g., ethicality, environmental responsibility) and *exportability* of whole or part of the aid intervention, meaning the extent to which it could produce important contributions to other aid interventions (e.g., via use of its innovative design, approach, or product, and cost savings).

We will now address these points in some detail. The main issues emerging from the analyses related to relevance and effectiveness have the same conceptual root. The DAC criteria seem to assume that the evaluation should be conducted to determine whether the program met the aid intervention goals in order to determine its success. As discussed in the literature (e.g., Davidson, 2005; Scriven, 1991), using goals as the primary guide to evaluations can be quite misleading because measuring program goals may not necessarily determine the value of the program to the recipients.

There can be no doubt that program goals are important for planning and monitoring functions. They provide the necessary orientation to managers regarding how the intervention should be implemented and the specific indicators that should be tracked over time in order to measure important aspects of the project outcomes, and to determine how well the intervention is evolving. However, measuring the level of goal achievement cannot be considered a sound basis for an evaluation of an intervention because goals, if not grounded in a sound needs-assessment, reflect only the expectations of program designers, managers, and other stakeholders. As such, goals are not necessarily connected to the real needs of the targeted populations. Of course, there are cases where goals are defined based on well-designed needs-assessment, thus making them soundly aligned with the main existing needs. Nevertheless, what is at stake in an evaluation, and should make up the primary aim for an evaluator, is the search for what is really happening as a result of the aid intervention, regardless of what was initially intended by the program managers or other stakeholders Also, often times, depending on the context, goals can be set too low or too high, and thus not provide a good parameter for evaluating an intervention.

In the definition provided by OECD/DAC for assessing relevance of an aid intervention, the evaluator is challenged to consider whether the program design, activities, and outputs are aligned with the policies and priorities of a target population, fund recipients, and donor agencies. In practice, this discussion usually explains how an aid intervention relates to the donors' and governments' strategies. It certainly helps to establish the context and significance of the intervention for the donors and governments, but it is not necessarily evaluative. While the call for considering priorities of the target group may lead evaluators to take into account people's needs, the other components of the definition are directly connected with the established goals either by the recipient countries or by the donor agencies. This approach can blur the perspective of the evaluators and divert their attention from the core function of the criterion—which should be to determine whether the intervention's design, activities, and initial results are adequate to respond to existing needs. It seems reasonable to make adjustments in the definition of this criterion by focusing the definition on program recipients' needs.

A similar argument applies to effectiveness. In this case, the OECD/DAC definition indicates that the level of goal achievement (or the likelihood of their achievement) should be used as one of the main criteria to determine the merit of an aid intervention. As explained above, program goals can be misleading and a focus on them can sidetrack evaluators from what is really essential, i.e., determining if an evaluand[7] is producing meaningful outcomes that are addressing existing needs instead of fulfilling pre-established goals. Again, if the goals are perfectly aligned with people's needs, then measuring the achievement of the goals will certainly point evaluators to the right direction. However, a good evaluator should never take for granted that the program goals adequately reflect the needs of the target population. Revising the definition of effectiveness to encompass this perspective is another option for improving the DAC criteria.

A more radical possibility could involve the dissolution of this criterion, assuming that it could be subsumed under impact. The logic for the latter option is that impact requires a careful and comprehensive assessment of the results produced by an intervention including expected and unexpected, positive and negative impacts. One could argue that the search for the positive and expected impacts would correspond to the revised version of the definition of effectiveness and, consequently, eliminating the necessity for a stand-alone criterion.

It is also relevant to recognize that the concept of need overlaps substantially with impact. It does not seem possible for a project to have highly cost-effective impacts and not address a real need of a population or group. Furthermore, a project should not be negatively assessed for not addressing all the needs of the beneficiary population/group. Implementing agencies are not necessarily competent to address needs outside their area of expertise. Only in some cases, e.g., emergencies, can a project be properly criticized for not addressing the population's most pressing needs (however these may be identified).

The definition offered by the OECD/DAC for sustainability has missed important elements. First, it seems to ignore evaluative studies conducted several years after the original funding has been withdrawn—retrospective studies. Second, while it clearly addresses economic and environmental aspects of sustainability, it falls short in discussing several other essential elements of sustainability such as political support, socio-cultural adequacy, technological appropriateness, and institutional capacity. For instance, if an intervention does not take into consideration the specific culture of a given region or community, even if initial results are positive, the likelihood of maintaining a program intervention will sharply decrease when the initial funding is withdrawn. This is especially relevant to programs that require direct participation of program recipients to achieve success—e.g., in a water

[7] Whatever is being evaluated (e.g., programs, projects, policies, etc)

and sanitation intervention, community groups are responsible to organize and pay for maintenance of water pumps and pipes. This aspect is also relevant to the possibly new quality of process criterion, since delivering culturally inappropriate activities or services can considerably decrease an evaluator's assessment of the quality of an aid intervention. Making those dimensions explicit in the definition of sustainability will certainly strengthen it. It is interesting to note that one of the OECD/DAC members, the Danish International Development Agency (DANIDA), has already included aspects beyond financial and ecologic issues. They identified seven determinant factors for sustainability of aid interventions including: policy support measures, choice of technology, environmental matters, socio-cultural aspects, institutional aspects, economic and financial aspects, and external factors (DANIDA, 2006, p. 57).

Efficiency has been defined by OECD/DAC as the determination of whether aid interventions use "the least costly resources possible in order to achieve the desired results" (OECD, 1992, p. 1). The definition clearly states that in order to arrive at good conclusions about efficiency, it is necessary to conduct a cost analysis and compare the intervention with possible alternatives.

There are many important components in a cost-analysis besides direct money cost that are, unfortunately, quite often overlooked in development evaluations. It seems appropriate to urge evaluators to take into consideration non-monetary costs (e.g., participants' time or stress), as well as other important types of cost including indirect, start-up, close-down, maintenance, and opportunity costs (Scriven, 2007).

In terms of assessing alternatives to an aid intervention, it could also be valuable to call the evaluators' attention to think broadly, and not restrict themselves to the most obvious comparisons. Evaluators should be challenged to consider possibilities that are both less expensive than the current intervention, and more expensive, as long as these alternatives produce reasonably similar results. Thinking about existing alternatives, including options that could be logically predicted for the future, would also expand the evaluator's perspective in determining the value of the intervention under consideration.

Complementing the current version of the criterion with some specific guidance on what to look for on cost and comparisons could make the criterion even stronger. Furthermore, the term efficiency has been associated more with *least costly approach* which is a limited definition given the broaden meaning of the criterion. Cost-effectiveness is a more comprehensive term and seems to better define the many concepts embedded under this criterion.

Finally, quality of process and exportability are key criteria that are missing in the OECD/DAC list. It can be argued that some components of the aid intervention's process are already contemplated under efficiency (e.g., how the intervention is performing in terms of using resources to produce results)[8], and, to some extent, under relevance (e.g., how important the activities and outputs are in terms of addressing people's needs). However, there are a number of very important process elements left out from the five criteria that can be determinant in assessing the quality of an intervention. Those aspects include (i) ethicality (e.g., are any ethical norms not observed in the delivery of services to recipients or in treating staff?), (ii) environmental responsibility (e.g., are the activities completed by the intervention producing current or future damage to the environment?), (iii) scientific soundness (e.g., does the program follow sound scientific

[8] Indeed, some could make the argument that efficiency should be subsumed under the new quality of process criterion. However, this would make the new criterion overweighed, with too many and too important aspects embedded in it. Keeping them separate might be a better solution to avoid the risk of overshadowing some important aspects.

knowledge or accepted best practice guidance of the relevant sector, based on research and evaluations of similar interventions?), (iv) adoption of alleged specifications (e.g., is the intervention delivering what was promised?), (v) coverage (e.g., are the targeted people being covered, do men and women, boys and girls have equal access to benefits, and is the intervention covering an appropriate number of recipients?), (vi) responsiveness (e.g., is the intervention adequately responding to the changing environment?), and (vii) stakeholder participation (e.g., do men and women, and/or boys and girls or relevant sub-groups in the society have equal opportunities to participate in program decisions and activities?), and (viii) cultural appropriateness (e.g., are the services and activities being delivered in accordance to local cultural norms?). Failing to provide credible answers to these and other similar questions will certainly affect the quality of the evaluation of any aid intervention.

Exportability is the other important aspect missing from the five criteria. It determines the extent to which an aid intervention as a whole or some of its elements (e.g., innovative design, approach, or product) is transferable (e.g., could be potentially worth or produce a key contribution) to another setting (Davidson 2005, p. 6). A positive response to the previous question will clearly affect the way an evaluator will determine the importance or significance of an aid intervention, and also the way he or she will assess the intervention's sustainability[9]. It is important to note, however, that the meaningful application of this criterion will require from evaluators broaden knowledge outside the intervention being evaluated, e.g., other similar (or not) aid interventions, and a certain doses of creativity for considering possible applications of successful ideas to other settings. Another caution is the need to avoid confounding exportability with replicability—a criterion loosely and, sometimes, harmfully used in development aid. More often than desired, managers push for the full transferability of a successful aid intervention to other settings, without careful consideration of the specific socio, economic, and cultural specificities with disastrous results.

The addition of quality of process and exportability to the existing list of DAC criteria will make them much stronger.

It is interesting to note that some of the changes we recommend in this paper are also, to some extent, reflected in the Active Learning Network for Accountability and Performance in Humanitarian Action's (ALNAP) recent reinterpretation of the five criteria for application in humanitarian actions. For instance, ALNAP expanded the relevance criterion to umbrella appropriateness and restated it to emphasize assessing the project's alignment more with local needs than with donor policies. ALNAP also included the idea of quality of process by incorporating cross-cutting themes such as environmental responsibility and gender equity in all criteria (ALNAP, 2006).

The Relative Importance of the OECD/DAC Evaluation Criteria

The current definition of the five criteria implies that they all have the same level of importance. A reasonable question to ask is whether the criteria should have different weights in determining the overall assessment of an intervention. For instance, should the impact produced by a project receive higher weight in comparison to the other criteria in the overall summative assessment about that project?

Even though the establishment of weights for the criteria seems to present some relevant benefits, the accomplishment of such a task is not easy, if at all feasible. It is possible to defend that producing substantial positive impact, in many situations, is a more crucial criterion to determine merit and worth of a project than

[9] Considering sustainability in a broader perspective than only the continuation of program activities beyond donor initial funding.

other criteria. For instance, if a project eradicated hunger in a poor region, even if it did not present very good performance in terms of efficiency, sustainability, effectiveness, or relevance, it might still be considered a good project. However, this statement can only be taken seriously if the project's performance in the other criteria was not at a level considered unacceptable. With this in mind, the answer to the appropriateness of weighing the criteria will have first to address the issue of *bars*.

A bar, according to Scriven (1991), is the minimum acceptable level of performance on a criterion below which an intervention will be considered fully unacceptable regardless of its performance on other evaluation criteria. In considering the five DAC criteria, impact, efficiency and sustainability criteria should have minimum acceptable levels of performance (bars) associated with them. If quality of process is included in the DAC criteria, it should also be considered a good candidate for setting bars.

As for the impact criterion, a bar should be established at the dimension *negative side-effects*—i.e., if an aid intervention is affecting the people or the environment in any serious detrimental way, then the aid intervention should be considered unacceptable regardless of how well it performs in other criteria (e.g., being efficient, having high quality of implementation, producing positive impacts). Bars should be established for efficiency at the level of waste of scarce resources or high costs (monetary and non-monetary costs). For instance, if an aid intervention is producing good results in meeting people's needs, but, in order to do so, it is requiring much greater resources than what would be acceptable, or, to access benefits, participants need to spend too much time or encounter serious distress (all at unacceptable levels) then the intervention cannot be deemed acceptable.

Sustainability is also an important dimension that can require 'bars'. An aid intervention will likely be seen as an unwise investment of scarce resources if the positive outcomes produced by the intervention disappear (or are likely to disappear) right after the original funding is withdrawn and the situation of project participants returns to its original, or even less desirable condition. Of course one may argue that the benefits produced during the intervention's lifetime were so significant that they might have overshadowed the lack of sustainability in the future (e.g., several lives were saved). Also, sustainability will only be essential to the extent to which meaningful outcomes are produced by the project for a reasonable cost, with no, or minimal and acceptable, waste of resources without incurring any ethical negative impact. There is benefit in placing a bar on sustainability, but only after the evaluand clears the bar in the other four criteria.

At least two components of the quality of process criterion—ethicality and environmental responsibility—constitute particularly important features of any evaluand and should have bars associated with them. Discrimination of participants or staff based on gender, religion, ethnicity, sexual orientation, etc, is a serious ethical issue and could justify failing a given intervention even if it performs well in other criteria. Similarly, if an aid intervention is producing important immediate benefits to participants (e.g., increase in people's income) but placing environmental conditions into jeopardy, its acceptability becomes questionable. If there are unavoidable damages to the environment due to extreme reasons (e.g., survival), then the program must consider a strong plan for implementing effective measures that will progressively recover the damages.

Quality of process also has other components that even though not as crucial as ethics or environmental responsibility, can certainly influence the performance assessment of any evaluand. The main examples include provision of alleged services (if these services address a local need), following acceptable standards of practice in the field, and adoption of most current scientific knowledge.

It is possible to identify a bar for relevance, but only in pre-formative or formative evaluation processes. Those are evaluations conducted during the design and implementation phases of an intervention; they provide an opportunity for organizations to use their findings to introduce changes (improvements) to the aid intervention early in the design phase or as its ongoing during the implementation phase. If the evaluand is found not to address existing needs then it is reasonable to conclude that it is not performing at a minimum acceptable level and, therefore, should be immediately revised.

It is hard to defend that effectiveness should lend itself to the establishment of bars. Even if the project's goals and objectives are connected to the needs of the participants, not achieving some of the goals (in part of in full) might not provide grounds to determine that the intervention was unacceptable. This is the case because the intervention might still have provided some important, unexpected benefits to the participants which were not thought out as objectives or goals of the intervention.

Returning to the issue of weighing, relative to the other criteria, whether the aid intervention is producing meaningful changes in people's lives certainly carries much weight and places the impact criterion on a possible superior position in terms of importance. If an intervention is producing significant impact, even if it is not very efficient or the original objectives are not being achieved as planned, as long as it clears the bars for the other criteria, it will probably be considered a good intervention; while the reverse will not be true—if an intervention is very efficient, but is not really producing relevant impact then it will probably not be considered as good. However, providing a correct numeric weight to impact is tricky, since there are no clear grounds to establish that value—should it be weighted 50%, 100% or another percentage more that the other criteria? One way to display a higher level of importance for impact in relation to the others would be to set a higher bar for that criterion.

Final Comments

The five OECD/DAC evaluation criteria have been an important step forward to make the evaluation of aid interventions more comprehensive. However, there are some key issues related to focus (the need to refocus relevance and effectiveness on needs of potential beneficiaries and not on funders' and/or governments' priorities), omissions (need to include quality of process and exportability as part of the criteria) and importance determination (need to establish bars for some key criteria) that should be addressed so the DAC criteria can, once again, lead the international aid evaluation field to a more advanced position.

There are some scholars and practitioners who might rightly argue that one thing is to have good evaluation criteria for international development interventions; another thing is to properly apply them. It would be naïve to think that just because we have an improved set of criteria, evaluations in the field of international aid will improve accordingly. However, the common say of first things first is applicable to this discussion; without a good set of evaluation criteria, chances are that evaluators will not look for the right things when conducting assessments of international development interventions. Therefore, getting those criteria right, is the first key step to help push the field forward.

Author Acknowledgments

I am very thankful to several colleagues who provided important suggestions for improving earlier versions of this paper, who include: Todd Harcek, Mina Zadeh, Krystin Martens, Tererai Trent, Paul Lamphear, Ron Visscher, Daniela Schroeter, Otto Gustafson, Ryoh

Sasaki, Claudia Ceccon, Paul Clements, Michael Scriven, and Niels Dabelstein.

References

Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP). (2006). *Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies*. London: Overseas Development Institute.

Chianca. T. C. (2007). *International development Evaluation: An analysis and policy proposals*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.

Clements, P. (2005). *Inventory of evaluation quality assurance systems*. Unpublished manuscript prepared for the United Nations Development Program.

DANIDA (2006). *Evaluation guidelines*. Copenhagen, Denmark: Evaluation Department, Ministry of Foreign Affairs of Denmark.

Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Goldenberg, D. A. (2001). *Meta-evaluation of goal achievement in CARE projects: A review of findings and methodological lessons from CARE final evaluations, 1994-2000*. Retrieved January 31, 2007 from http://www.care.ca/libraries/dme/CARE%20Documents%20PDF/CARE%20MEGA%20Evaluation%20Synthesis%20Report.pdf

Goldenberg, D. A. (2003). *Meta-evaluation of goal achievement in CARE projects: A review of findings and methodological lessons from CARE final evaluations, 2001-2002*. Retrieved January 31, 2007 from http://www.kcenter.com/phls/MEGA%202002.pdf

Kruse, et al. (1997). *Searching For impact and methods: NGO evaluation synthesis study. A report prepared for the OECD/DAC expert group on evaluation*. Retrieved January 28, 2007 from http://www.valt.helsinki.fi/ids/ngo/

Leading Edge Group (2007). *Evaluation gap update April 2007*. Retrieved on August 9, 2007 from http://www.cgdev.org/section/initiatives/_active/evalgap/eupdate.

Organization for Economic Cooperation and Development (OECD). (1992). *Development assistance manual: DAC principles for effective aid*. Paris: Organization for Economic Cooperation and Development.

Organization for Economic Cooperation and Development (OECD). (1998). *Review of the DAC principles of development assistance*. Paris: DAC Working Party on Aid Evaluation.

Russon, C. (2005). *Meta-evaluation of goal achievement in CARE projects: A review of findings and methodological lessons from CARE final evaluations, 2003-2004*. Retrieved on January 31, 2007 from http://pqdl.care.org/pv_obj_cache/pv_obj_id_3F0964E46D34E15DD78EB2D03DF1DFEFE1FC0200.

Scriven, M. (1991). *Evaluation thesaurus*. (4th ed.) Newbury Park, CA: Sage.

Scriven, M. (2007). *The key evaluation checklist*. Retrieved on September 7, 2007 from http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf.

Savedoff, W. D., et al. (2006). *When will we ever learn? Improving lives through impact evaluation*. Retrieved on January 31, 2007 from http://www.cgdev.org/content/publications/detail/7973