

## Comments and Proposals Concerning Chianca's "The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement"

Hellmut W. Eggers

Having been involved in the discussions preceding the establishment of the OECD/DAC criteria as member of the DAC Working Party on International Development Cooperation Evaluation (IDCE) representing the European Commission, it is understandable that I have read Chianca's (2008) article with exceptional interest. There is hardly anything in this article that I would disagree with, except a few minor aspects that I will mention in due course. I also fully support the idea of taking a critical look at the DAC criteria and at IDCE approaches in general, a critical review that should always continue as an ongoing process. "If followed," Chianca says the changes he proposes "could contribute for increasing the quality of evaluation for the purposes of . . . decision options" (p. 1). I will try to formulate some proposals with a view to bringing that "If followed" premise a bit closer to reality. My comments have been written with the contents of my article, "Planning and Evaluation, Two Sides of the Same Coin," in mind and should be understood in that context. (Eggers, 2006).

The author lists the five DAC criteria of Relevance, Effectiveness, Efficiency, Impact, and Sustainability together with their DAC definitions and underlines that they represent

remarkable progress compared with the time preceding their establishment. That is certainly true considering the set of criteria that was then in use (and is, according to the author, "still quite preponderant" today!); to wit, "measuring outputs, monitoring resources' application and, where more sophisticated, estimating a project's economic rate of return" (Chianca, 2008, p. 44). Well, that reminds me of the difference between the dark ages and Enlightenment and, after so many years, I am still proud of the contribution the DAC Working Party made in that respect. But can the result of our work be further improved? Of course it can, and it should!

A task force of ten professionals, all pursuing doctoral degrees in evaluation at Western Michigan University, reviewed the DAC criteria and reach these conclusions:

- (a) Meeting the needs of the targeted population should be the top priority.
- (b) The *effectiveness* criterion should be refocused on "impact" instead of "goals."
- (c) The criterion of *sustainability* should make reference not only to future but also to "retrospective" sustainability and include, besides environmental and financial aspects, political support, cultural appropriateness,

adequacy of technology, and institutional capacity.

- (d) *Efficiency* should cover not only monetary but also nonmonetary costs, include comparisons with alternative solutions and not be limited to defining the least costly approach (“cost-effectiveness” appears to be a better term). (Chianca, 2008)

I agree unreservedly on point (a) above, both as to the substance and as to the wording (terminology). I also agree on the substance of points (b), (c), and (d), but I think that the terminology has to be reviewed. It is important to keep substance and terminology strictly apart. What matters first and foremost, of course, is substance. How to convey the substance with adequate terms is strictly a matter of convention and in no way alters the substance. But conventions, once established, should be adhered to if there are no convincing reasons to change. The bottom-up chain of effects in classical “logical framework analysis” has always been this: inputs, outputs (or results), outcome (or “project/program/policy (PPP) purpose” or “specific objective”), goal (or wider objective). Acquiring lowest cost inputs is called “economy”; transforming inputs into outputs in the most rational manner is called “efficiency”; transforming outputs into outcome that meets the needs of the target group is called “effectiveness”; and the contribution of the intervention to the next higher PPP level, i.e. to the goal or “wider objective,” is called “impact.” Maintaining benefits during a prolonged period without international assistance is called “sustainability.” Readers who have the traditional terminology in mind will thus find it very difficult to understand, for instance, the statement that “the definition of effectiveness focuses on determining the extent to which the intervention met its goals” (Chianca, 2008, p. 45). The confused reader will also balk at the idea of replacing “efficiency” with “cost-effectiveness,” because this last expression reaches in one go from inputs to outcome.

What advantage does that have, and why should the much clearer and logical linking of inputs to outputs and of outputs to outcomes be abandoned in favour of an expression that advances confusion rather than clarity? That confusion would not exist if we could all agree on a common terminology, and as long as it is not demonstrated that there are convincing reasons in favour of changing the traditional terminology that terminology should be maintained. I know what your immediate reaction to the above statements will be, dear task force! It will be that I have not understood that the substance you want to cover by each of these terms does not fit the traditional meaning and that this is important because otherwise these areas of substance are left uncovered. You will point out, for instance, that “efficiency” should cover not only monetary but also nonmonetary costs, that “impact” should not only cover intended but also unintended consequences of the intervention, that sustainability should not only be considered for the future but also for the past, etc. You are perfectly right. But rest assured, I will address those aspects later in these comments. But let us clear up the terminological jungle first and assume common conventions on terminology!

As far as the main substance of the task force’s conclusions goes, I perfectly agree. But this substance is not new and was developed in the mid-eighties by the European Commission’s Evaluation Division of Directorates General VIII and I, headed by me at the time. The results of that work are contained in the “Manual of Project Cycle Management (PCM), Integrated Approach and Logical Framework, February 1993.” This manual has been translated into all of the official languages then in use within the European Union and has been distributed in many thousands of copies all over the world, including through the World Bank and UNDP. The approach, and more particularly its “Basic Format,” (BF) were conceived on the basis of the relevant documentation of *all of the members of the DAC*.

This means that it is based on thousands of years of experience, counted in “expert-years.” So, one would expect that there is really not much of real importance that has been left out. And indeed, all of the important issues raised by the WMU task force, ranging from needs assessment to government policies, etc., are included in the BF. I invite the task force to look at the catalogue of the additions to the five fundamental DAC criteria they propose and then cross them out, one by one, as they find the same criteria already included in the PCM BF. To their surprise, they will find that there are virtually none left on their list. Here is the updated version that entirely preserves the 1993 substance of the Basic Format, containing all of the criteria considered important by the DAC Evaluation Working Party. As can be seen, the Basic Format contains, almost verbatim, all of the suggestions formulated by the task force. The importance of nonmonetary cost, however, as underlined by the task force with convincing reasons, should be included in the BF, as I have done, inserting the parenthesis under BF point 5.4. The taskforce is also right, I think, in stressing the need to consider alternatives to the solution chosen by an intervention, and I have therefore inserted a new point 3.5.: “Alternative solutions.” Further improvements, following arguments by the task force, concern BF points 5.2. and 6.6 (see the last two paragraphs on the following page).

## The PCM Basic Format

### 1. Summary

### 2. Background

- 2.1 Government/sectoral policy - donor policy coherence/complementarity - democracy - human rights - good governance
- 2.2. Features of sector
- 2.3. Beneficiaries and other stakeholders
- 2.4. Problems and opportunities to be addressed (Relevance)

2.5. Other interventions, cooperation with other donors, past best practice

2.6 Documents available

2.7 Project/program history, including past, present and future application of PCM

### 3. Intervention (intended and unintended results)

3.1. Overall objectives (Impact)

3.2. Outcome: Project/program purpose (Effectiveness)

3.3. Outputs (Efficiency)

3.4. Inputs and activities (Economy)

3.5 Alternative solutions

### 4. Assumptions

4.1. Assumptions at different intervention levels

4.2. Risks and flexibility

### 5. Implementation

5.1. Physical and nonphysical means

5.2. Organization - procedures - transparency - ethicality

5.3. Timetable

5.4. Cost estimate (including nonmonetary costs) and financing plan

5.5. Special conditions: accompanying measures taken by government

### 6. Quality Factors ensuring Sustainability

6.1. Policy support

6.2. Appropriate technology

6.3. Environmental protection - climate change

6.4. Sociocultural aspects: Gender issues - participation - empowerment - ownership

6.5. Institutional and management capacity, public and private; decentralization of responsibilities; subsidiarity

6.6 Innovations

### 7. Economic and financial viability

### 8. Monitoring and evaluation

8.1. Monitoring plan and indicators

8.2. Reviews/evaluations

### 9. Conclusions and proposals (including overall sustainability assessment)

But there remain other important points raised under the conclusions of the task force,

alluded previously: Goals (better: “outcomes”) as compared with needs and future as compared to past sustainability, etc. The solution of these problems becomes self evident the minute one understands the following:

- (a) Evaluation and planning should be amalgamated, following the same structure: that of the Basic Format above.
- (b) Evaluation considers exclusively the past and planning exclusively the future;
- (c) Planning and evaluation pursue a common objective: the creation of sustainable benefits for the target groups of international development cooperation (IDC).

When planning, i.e., trying to predict the future, one should, according to point 2.4 of the Basic Format, analyze the problems and the opportunities of the target group and then address these under the “Intervention” points listed under point 3 of the BF. Of course, planners will only have one outcome (unfortunately called “impact” by the task force) in mind to be worked for: the one covering the problems and opportunities, i.e., the “needs,” of the target group as recognized by planners. Planners, having long ago been warned by evaluators, will be familiar with the danger of producing undesired outcomes and impacts and will try to avoid these. So, they will try to match exactly the needs to be covered by the outcome anticipated. Similarly, planners will anticipate the most probable sustainability scenario and plan the intervention accordingly. So, it is perfectly normal as well as desirable that planners should have one, and only one, scenario in mind when planning for each of the points contained in the BF. They simply cannot, and should not, do otherwise. Not so evaluators: Evaluators have the tremendous advantage of looking at facts, not just desires, because they analyze the past using, however, the same guide planners do: the BF. So, if needs

are not covered by the intervention as they should have been or, worse, if important needs have been overlooked by planners, evaluators will draw planners’ attention to this fact. If planned outcomes and impacts seem unrealistic or exaggerated in light of actual experience, evaluators will draw planners’ attention to that probability. Planners may then be in a position to redress, even if only partly, the situation concerning the current intervention and learn what to do and what to avoid in future similar interventions. Corresponding considerations hold concerning sustainability, past and future. In other words, evaluators will never judge the success of an intervention just by comparing planned outcomes with outcomes realized. They will never judge sustainability just by analyzing planners’ intentions. They will be looking at the facts; have past needs, past sustainability, past intended and unintended impacts in mind; and formulate their recommendations for future interventions on that basis.

The taskforce further proposes that what they call two missing key criteria should be added: (a) “quality of process” and (b) “exportability” of innovative ideas. Both are important aspects, but I don’t think that they should be integrated into the planning and evaluation systems on the same level as the other five classical DAC criteria. As to (a), I fully agree that ethicality and environmental responsibility are important. However, environmental responsibility is already contained in the BF. As to ethicality, there is no doubt as to its significance, and so it does deserve serious consideration. But, again, it should not be a stand-alone criterion, but should be fitted into a logical system of planning and evaluation and can thus find its place within the BF under its point 5.2., which would become “5.2. Organization – procedures – transparency – ethicality.” Similarly, “best practice” experience as pioneered elsewhere in the development cooperation community deserves a special mention under point 2.5 of the BF, which would thus become “Other

interventions, cooperation with other donors, best practice.” All other aspects mentioned by Chianca are already contained in the BF.

Turning to “exportability,” it would be rather confusing to maintain this term as such. Planners would, quite understandably, have no use for this term as they cannot be sure which parts of interventions planned by them will eventually be recognized, by evaluators, as useful for other planners in different circumstances, i.e., as “exportable.” It may, however, be useful that planners indicate those products, measures, and procedures that they think will be innovative and then leave it to evaluators to arrive at a final judgment. If so, one could think of a new category in the BF, under point 6.6.: “Innovations.” In any case, I would strongly advise against a term like “exportability,” since it is confusing and will incite people to think of foreign trade or other nonsense issues. All other aspects mentioned under “exportability” in the article, including those applied by ALNAP, have already been part of the BF for many years.

I have thus made sure that the entire additional substance as contributed by the taskforce to the DAC evaluation (and, let me add, planning) criteria are incorporated into the BF. These additions are useful and important, and it is only normal that the BF, notwithstanding its being backed up by solid international experience during several decades, should come constantly under critical review and be improved accordingly. So, thank you, task force, for your positive contribution and keep questioning and improving DAC and other criteria (as well as your own)! Turning to this last mentioned issue (in parentheses), may I point out that the DAC criteria as reinterpreted and completed by the task force still fall far short of the optimum solution:

- (a) There are numerous important aspects that are indispensable for good planning and evaluation but that the task force fails to include. I leave it to the task

force to work out, on the basis of the BF, the list of these missing factors.

- (b) The task force does not realize that it is vital to establish these criteria not only for evaluation but also for planning; if, indeed, such amalgamation of planning and evaluation is neglected, feedback of lessons learned will remain the thorny problem it has always been, forgetting proceeding at the same pace as learning, with the result that the same deplorable planning and evaluation quality in international development cooperation persists. If such, “built-in” forgetting is to be avoided, learning must be systematically integrated into the planning and evaluation structure as contained in the BF and, feathering out into further and further refinements, in Basic Formats per sector, per cross-cutting theme, per instrument, thus leading eventually (and after further refinements through “regionalization”), to the final level: the case-by-case “Terms of Reference” for planning and evaluation. All of this material would be integrated into and permanently upgraded within the “PCM Database.”

Placing a “bar,” i.e., a “minimum acceptable level of performance” (p. 49) on each criterion, as proposed by Chianca, seems an excellent idea to me and a real improvement of the PCM approach. Anchoring such bars within the PCM Database appears to be a straightforward, perfectly feasible procedure. Including “relevance” in the “barring” procedure seems to me a particularly pertinent idea. Indeed, the backward-looking evaluation function begins the moment after the start of planning, i.e., without waiting for damage to be done by planners on the basis of, say, an erroneous needs assessment.

I also fully agree with Chianca when he pleads in favour of “weighing” the criteria, i.e., establishing a hierarchy among them. We

further agree on the most important among the criteria, the bedrock of the entire system. He writes, "Whether the aid intervention is producing meaningful changes in people's lives . . . places the impact criterion on a possible superior position in terms of importance" (p. 50). I only think that there is one superfluous and confusing word in this sentence, and that's "possible." I also have my doubts as to the term "impact" in the above context (see paragraph concerning terminology below). But these points are of minor significance only.

Indeed, the entire international development cooperation enterprise has one single objective: the creation of sustainable benefits for the target group. This is what planning and evaluation are all about. Let us not be sidetracked by confusing statements of stakeholders sitting on the sidelines and seeking their own advantage. Of course, corruption, politicians' and civil servants' hidden agendas, power plays along the aid delivery line and so on, will always be there. But nobody will ever say that they don't care for the intended beneficiaries, the "target group," the final delivery station of all aid and all cooperation effort! They can't without losing face! And that is what the international development community should nail them down to! It will make cheating a really hard business for them!

Before pursuing to formulate further considerations on substantive issues, I should like to make a few observations on terminology. Terminology is a matter of convention and practice. It does not have any "value" whatsoever as such. We can say people, folk, pueblo, popolo or people, always meaning the same subject. None of these expressions is better than the other. What's important, however, is that those who use these expressions and those who listen to them know what they are talking about. So, if everyone understands the same thing when talking about "impact," that would be OK. Unfortunately, that is not the case. Traditionally, "impact" has always covered, in classical Logical Framework

Analysis (LFA), the contribution of a project or program towards the realization of a higher-level policy goal. Why change? But if we do, let's do it all together, while not forgetting that it is of substantial, not just formal, importance to cover the entire substance between inputs and policy goal achievement. If we define, as does Chianca, "impact" as covering the creation of sustainable benefits for the target group called "outcome" in LFA, then we would have not only two but, together with "purpose" and "specific objective," four expressions for "outcome" and none for covering the contribution of a project/program towards the realization of wider policy objectives. That's no advantage, is it? It's ridiculous! We might just as well decide to eliminate the expression "fire" (as sounding too dangerous) and replace it by the expression "water-minus."

Last but not least, no, last but most important of all, there is the question of the practical realization of Chianca's proposals: He says that "if followed," the changes proposed "could contribute for increasing the quality of evaluations for the purpose of (re)funding, program changes, and other decision options" (p. 41). I take this to mean *more success in international development cooperation in terms of creating sustainable benefits for target groups of IDC*. I agree. So, let us have a hard look at that fateful: IF FOLLOWED!

Chianca pleads in favour of the old saying "first things first," and who would not agree? Publishing this article is a first step. But we all know that this publication won't be sufficient to cause even so much as a tiny start in improving the quality of IDC. How many of the people who should will read this article? A tiny minority. And how many of those will want to have a serious go at improving IDC? Again, a tiny minority of that tiny minority. So, we may safely replace the "IF" by "NOT" if things don't go beyond this publication. Chianca will certainly agree with this statement as he does not hesitate to affirm that "it would be naive to think that just because we have an improved set

of criteria, evaluations in the field of international aid will improve accordingly” (p. 50). I agree. Not only would it be naive, it would be downright stupid. But let us assume that, by a magic trick performed by the development fairy, evaluations would, indeed, become better because of more pertinent criteria. Would that improve, in itself, the success of IDC? Not one bit! Again, we would need the magic trick by the development fairy so as to induce planners, implementers and intended beneficiaries to apply the lessons learned through those improved evaluations. And as forgetting proceeds at the same pace as leaning, planners, implementers and beneficiaries, by and large, don’t apply those lessons, and thus do no better jobs than they used to, and IDC remains of the same shaky quality today as it has always been. So, there’s no light at the end of the tunnel, and that is indeed what many critics of IDC affirm. If they are right, we might just as well give up. But that is precisely what we are NOT going to do.

Paul Clements (2008) goes considerably beyond those “first-things-first” steps when proposing measures with a view to improving IDC evaluations. He suggests (a) concentration on two evaluation criteria: cost effectiveness and impact; and (b) an organizational approach to ensure an appropriate back-up to such improvement via the creation of what he calls “Evaluation International” (EI). Thus, he and Chianca underline the need to ensure a privileged position for “impact,” understood as the positive change in living standards of the target group to be brought about by the development intervention considered. I fully agree, being certain that one can easily come to an understanding concerning the terminological issues discussed above. But I do not believe that these suggestions, if realized, will bring about the necessary change in the quality of IDC (see previous paragraph). I believe that we should go beyond these proposals in order to arrive at operational results as measured in terms of sustainable benefits for the target groups of

IDC. So, what seems necessary to construct that all-important bridge between theory and practice?

First of all, evaluators should give up talking of evaluation only and considering themselves “bande à part” (as the French would put it). Evaluators have always had that nasty tendency to keep themselves to themselves. Time and again, they have invented, modified, reinvented, and modified again, the structure of evaluations. In doing this, they have made it very difficult indeed to do what would have been, and what continues to be, the only justification of their efforts: contribute, together with planners, implementers/monitors and users (i.e., all the other stakeholders), to the creation of sustainable benefits for IDC target groups. It is time for evaluators to join the team and adhere, together with all the others, to a common fundamental structure, based on solid experience, for evaluation and planning alike, to be kept under review and to be continually improved by all. The PCM Basic Format is such a structure. Planners and implementers should apply that structure prospectively, evaluators and monitors retrospectively.

Next, all involved should agree on the “centre of gravity” of the entire system. Chianca proposes “impact,” Clements “impact” and “cost-effectiveness,” I agree with both as far as the concentration on “impact” is concerned (I would, let me repeat, call it “outcome,” “project/program purpose” or “specific objective”). Indeed, the “centre of gravity” should be one and one only. It seems easy enough to define “impact” (meaning what I have called “outcome”) in such a way that all of the other criteria will contribute to its realization, including “cost-effectiveness” which, in the hierarchy of criteria (Chianca) must range below “impact” (= “outcome”). Such concentration on the one objective that encompasses all the others and that, moreover, everyone can understand is, in my view, absolutely essential if we want to cross that all-important bridge between theory and practice.

And when I say “we,” I mean the entire development cooperation community, including not only politicians, administrators, technicians, planners, and evaluators, but also taxpayers and farmers in developing countries, no matter whether or not they know how to read and to write (there are even some taxpayers around who don’t). A farmer in Mali will scratch his head when asked what the ERR of an intervention might be, but he will know immediately what he would consider a good well: one that will provide enough drinking water for his family, himself and his livestock for many years (and that’s SBTG!).

The deplorable quality of IDC evaluations is not only the result of haphazard structuring, but also their failure to produce operational feedback, i.e., their lack of integration into the planning cycle. As shown above, this weakness can and should be overcome by adopting a single structure for planning and evaluation, represented by the Basic Format. The BF should be accompanied by a more detailed commentary covering each of its different points. For each major development sector (agriculture, industry, education, public health etc.), different sector BFs should be established, and these should be adapted to the specific conditions prevailing in each country/region. Furthermore, the BF should be adapted to specific development instruments, like structural adjustments, direct budget support, etc., again subsequently to be regionalized. It should also be adapted to crosscutting themes, like gender issues, environmental questions, organizational problems, etc. Finally, the specific terms of reference for each intervention should be established ad hoc following, however, the BF structure. This entire, voluminous documentation, the PCM Database, should constantly be kept under review, and new insights gained by evaluations can thus immediately be integrated into the system. Learning, instead of being systematically forgotten, will never be lost and will be made

available automatically to the planning and implementation of new interventions.

Comparisons, as everyone knows, never cover the entire substance of the reality they are referring to. They cannot be “scientific.” They can, however, fire the imagination and stimulate creative thought. Here, then, is a comparison referring to the system as outlined in the previous paragraph: The BF is like the DNA of the model “organism” we are dealing with when talking of projects, programs, policies, sectors, crosscutting themes, instruments, etc. It represents the fundamental structure of all these different organisms: they all have heads, spines, blood vessels, and so on. They are released into the world and interact with it and with each other. They are challenged by their environment and “evaluated” over and over again. Changes in their genomes will allow them to react and adapt to these challenges. They will be “improved” and “redesigned,” ever so gently and progressively, thus increasing their survival chances. Planning and evaluation, constantly renewed along their life cycles, will mould these organisms into shapes and forms that improve their survival chances and make them “sustainable.” Adaptations to local circumstances and pressures will proceed relatively quickly, but changes in the master-mould, the Basic Format, will be rare, and if they occur, they will have been thoroughly tested by real-life circumstances. Information contained in DNA’s vast database, gathered and constantly renewed during the life cycles of many generations, will thus be kept up-to-date and will constantly guide the emergence of new and better adapted forms of life.

I will not overstretch that comparison so as not to overstep the limits of the ridiculous, but I will point out, nevertheless, that the “evolution” as traced above has to be kept moving by a source of steadily renewed energy, and that is where Evaluation International as proposed by Clements comes into play. The need for the creation of this office had never crossed my mind, but after reflecting on the issue this



requirement now seems quite obvious to me. Such is the result of teamwork. Clements favours more of a formal, norm-setting structure, while I think one should rely on a less rigorously structured, more consensus- and persuasion-oriented central “secretariat.” It is not necessary, however, to decide this question right now. Applying Chianca’s “first-things-first” motto, here is what I should like to propose as a first step:

The task force, supported by Chianca and Clements, should draft a “strategy” or, better still because less pretentious, an “orientation paper” containing, on some ten to twelve pages, three main sections: (a) Reasons for the insufficient quality of evaluations in IDC; (b) what to do to improve the situation by (i) creating a system of operational feedback-oriented evaluation, (ii) organizing a coordinating EI unit, and (iii) selecting a series of contact organizations that would be invited to critically review the proposals under (i) and (ii) and then enter into an international network, allowing to run the system and to constantly improve its effectiveness. I could help with establishing the terms of reference of the orientation paper and with reviewing its draft version prior to the formulation of the final version.

These are, then, my observations whose formulation was prompted by the study of Chianca’s article, together with that of previous writings by Clements. I have read these contributions with lively interest, and I had to correct some of my views in their light, while incorporating a series of suggestions contained in them into the PCM system. Maintaining a healthy mistrust as to one’s own opinions is always conducive to progress, I think, and I hope that my correspondents hold similar convictions. This attitude is, I know from experience, a perfect means to avoid losing face. If one doesn’t lose face, one will not lose faith; and if one doesn’t lose faith, one will have a real chance to become a Whynot instead of a Yesbut. You don’t know what that is? Then

turn to the “Evaluation Poetry” as published in *JMDE* volume 4, issue 7 (Eggers, 2007). We, a handful of people in a frail sailing vessel, are trying to nudge an ocean liner into course correction, you say? That’s right. Chances to succeed are dim, you say? Yesbut whynot try?

## References

- Chianca, T. (2008). The OECD/DAC criteria for international development evaluations: An assessment and ideas for improvement. *Journal of MultiDisciplinary Evaluation*, 3(9), 41-51.
- Clements, P. (2008). An association to improve development aid. *Journal of MultiDisciplinary Evaluation*, 5(9), 52-62.
- Eggers, H. W. (2006). Planning and evaluation: Two sides of the same coin. *Journal of MultiDisciplinary Evaluation*, 3(6), 30-57.
- Eggers, H. W. (2007). On yesbuts and whynots. *Journal of MultiDisciplinary Evaluation*, 4(7), 108.