

Relating Outputs, Outcomes and Impact in the Evaluation of International Development Projects

Robert Kenyon Walker
Free-lance evaluator

Background: The recent controversies on the pages of the JMDE regarding UNDP evaluations, as well as the DAC criteria, are discussed in the light of two UNDP/GEF evaluations in Latin America for which the author was primarily responsible.

Purpose: The author defends the utilization of all five evaluation criteria of the Development Assistance Committee of the OECD (the DAC criteria), and their integration via Theories of Change.

Setting: The pine/oak forest of Honduras and the mangrove regions along the coast of Brazil.

Intervention: Two projects of the Global Environment Facility (GEF) and the United Nations Development Programme (UNDP), with the respective governments.

Research Design: Mid-term evaluations with site visits and interviews.

Data Collection and Analysis: In the Brazilian case, longitudinal analysis of available data was conducted and related to the findings of the interviews and observation, as well as published reports and studies. A Theory of Change (action model) of the Honduran project was structured and graphically portrayed based on desk review of the project document and other documentation, adapted following initial interviews, and field tested.

Findings: In Brazil, preliminary evidence derived in part from Management Effectiveness Tracking Tool (METT) questions with reference to results indicates a relative lack of project effectiveness and of preliminary signs of impact. The state of Pará offers one possible exception, related in part to the fact that a certain momentum had already been built up in that state prior to project start-up and in part to the early adoption through the project of a new method of transportation of the fiddler crab (*Ucides cordatus*), in baskets covered with wet sponges rather than in sacks. The Honduran project, with the strong support of the national government and the UNDP field office as well as GEF, successfully adapted the provisions of the project document to pursue a more community- and community organization-centered approach, rather than relying on the outside consultants originally specified. Project-supported gathering of impact data raised environmental awareness, strengthened the local university and established a baseline for future ex-post impact evaluation.

Keywords: *Theory of Change; DAC criteria; capacity development; displacement of goals; UNDP/GEF.*

Introduction

With a view to conceptual clarity and uniqueness, it has been suggested (Chianca, 2008) that the concept of effectiveness be subsumed under that of impact in the OECD/DAC criteria (relevance, efficiency, effectiveness, impact and sustainability). The suggestion aims at reducing redundancy, if effectiveness is taken to be equivalent to the achievement of objectives, while expected policy, program or project results (objectives) are generally included under impact. However, in the six years since the suggestion was published in the JMDE, it has not been taken up by the Development Assistance Committee (DAC) or the international development community in general, the member organizations of which normally continue to require evaluators to use the five criteria in international development and environmental initiatives.

Traditionally, the specific or “immediate” objectives of UNDP projects have been mostly related to capacity development, while the “development” objective is often stated in the form “To achieve an impact through capacity development,” or “To develop capacity in order to achieve an impact.” This seems consistent with the fact that the United Nations General Assembly has defined capacity development as a core function of the United Nations development system. Thus, we may say, basically, the “effectiveness” of UNDP projects as currently defined for the most part has to do with capacity development. The UNDP results chain postulates that the efficient production of outputs leads to effective capacity development, which will at some point have positive impacts on human welfare and the environment. Although UNDP is responsible for monitoring the Millennium Development Goals, seven of which call for impacts (the eighth calls for strengthening the international system), it seems to have often assumed that impacts can only be produced in the long term, (i.e., well after project conclusion). This despite the fact that the official definition (OECD, 2014, p. 1) makes no reference to time: “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.” It is also widely assumed that the assessment of impact requires expensive experimental or quasi-experimental designs, which, while perhaps affordable for big international loan agreements,

are seen as beyond the scope of most UNDP project evaluations.

Lempert (2010), in his rather polemical critique of UNDP and other international development organizations, mentions, among other alleged shortcomings, that “there is no attempt to show how inputs change behaviors or incentives to lead to long-term changes that are part of the organization’s mission (sustainable development) that will improve the lives of beneficiaries” (p. 102). Note that the rejoinder by the deputy director of the UNDP Evaluation Office (Uitto, 2010) does not deal directly with this point.

In his Table 8, Scoring of UNDP on the 25 Component Questions of the Indicator, Lempert cites the UNDP Website, which, referencing U.N. General Assembly Resolution 59/250 of 2004, calls for “the systematic evaluation of United Nations system operational activities by assessing their impact on poverty eradication, economic growth and sustainable development of programme countries. It further mandates the United Nations system to promote national ownership and capacity development and to make system-wide progress in collaboration in evaluation” (p. 97). Lempert (2010) implicitly characterizes UNDP as a “failed type,” because it “defines interventions on the basis of a specific tool that the organization uses (‘capacity building’) rather than on a comparative advantage of the organization in solving the root causes of a specific problem” (p. 82). He calls this “a clear failure of the UNDP system” (p. 102).

If the main goal is to impact poverty eradication, economic growth, and sustainable development, and capacity development is seen as a means to that end, then focusing exclusively on the latter might be seen as an instance of what Etzioni (1964) called “goal displacement,” the ‘mildest and most common form’ of which “is the process by which an organization reverses the priority between its goals and means in a way that makes the means a goal” (p. 10). More recently, Senge (1990) lists “personal mastery” as one of the five disciplines which characterize a “learning organization,” and points out that people with high levels of personal mastery “focus on the desired result itself, not the ‘process’ or the means they assume necessary to achieve that result” (p. 164). In the public health area, for example, increasing the regularity of disease reporting is no doubt an outcome of some importance. However, had Jonas Salk’s vision been limited to building a world in which every case of polio would be reported, he and his colleagues might not have been so motivated to achieve success; and even if they had been successful in that more narrow

endeavor, that would have been of little consolation to the future polio victims. “Letting our vision erode” is a common strategy for coping with our belief in our own powerlessness or unworthiness (Senge, 1990, p. 157), but it leaves the individual, his or her organization and the world at large so much the poorer.

There is one approach that focuses explicitly on outcomes, in the sense of capacity development: “Outcome Mapping (OM) focuses on one particular category of results: changes in the behaviour of people, groups, and organisations with whom a programme works directly. These changes are called ‘outcomes’” (Roduner, Schläppi, & Egli, 2008, p. 12). The main difference between the approaches of UNDP and OM would seem to reside traditionally in their generally perceived target groups (macro or grassroots) and their respective understanding of the expression “directly.” For UNDP (n.d. a), “Direct Implementation (DIM) is the modality whereby UNDP takes on the role of Implementing Partner” (p. 1). In contrast, Financial Regulation.27.02 of the UNDP Financial Regulations and Rules (FRR) defines National Execution (NEX) (now National Implementation Modality – NIM) as “The overall management of UNDP programme activities in a specific programme country carried out by an eligible national entity of that country.... National implementation is used when there is adequate capacity in the national authorities to undertake the functions and activities of the programme or project.... On the other hand, it is expected that the implementation of the national execution modality contributes to build national capacities” (UNDP, n.d. b, p. 1). This further building of “national capacities” is what UNDP calls “capacity development.”

UNDP is responsible for administering the projects of the Global Environment Facility (GEF). Recently the author of this paper was responsible for doing the mid-term evaluations of two GEF/UNDP projects, one together with Vag-Lan Borges in Brazil (Kinzo, Walker and Borges, 2013) and one in Honduras (Mainstreaming Biodiversity Conservation into the Management of Pine/Oak Forests). It is considered important for GEF-supported projects to have a positive environmental impact, or to slow or reverse environmental degradation (and, in fact, there does seem to be a rather recent trend to include biodiversity impact indicators in GEF-supported UNDP projects). How may such impacts be determined or estimated, particularly in largely formative evaluations such as ours?

For protected area systems such as those targeted by the Brazilian Mangroves Project, GEF

has a tracking tool called the Management Effectiveness Tracking Tool (METT) (Belokurov, 2007), which we were able to utilize in part in our evaluation of that project (see below). For other kinds of GEF-supported projects, such as the Honduran one, this is not used. One proposal, called Review of Outcomes to Impacts or ROTI (GEF, 2009), involves the use of Theory of Change applications to assess the likelihood that the strategies in place will produce positive impacts. It is this approach that we applied in the Honduran evaluation.

The ROTI Handbook (GEF Evaluation Office, 2009) presents the following “Generic project results chain underlying the Theory of Change approach”:

Strategy {Activities → Outputs → Outcomes
→ Impacts → Means → Ends},

where impact is defined as “A fundamental and durable change in the condition of people and their environment brought about by the project,” and outcome as “The short to medium term behavioural or systemic effects that the project makes a contribution towards, and that are designed to help achieve the project’s impacts” (p. 1). In the “Schematic of the GEF impact evaluation framework,” a box entitled “Intermediate States” is inserted between the Outcomes and the Impacts boxes.

Although stating that “The views expressed herein are the responsibility of the [anonymous] authors and do not necessarily reflect the views of the GEF Evaluation Office,” the paper does seem to reflect how GEF-supported programs in general are supposed to work in addressing social and environmental problems, i.e., in Chen’s (2005) terminology, their “change model” (Riemer and Bickman, 2011), based on descriptive assumptions. In the text below, an “action model,” which takes the change model as its starting point and makes prescriptive assumptions for the Honduran project, is portrayed.

A Theory of Change (ToC) is defined as “a systematic exploration of the links between activities, outcomes and context” (Actknowledge, 2011, p. 2). It is often used for program and project planning and monitoring, and thus for effectiveness evaluations (as opposed to efficacy evaluations, which address the question of whether a program or intervention can have effects under research-like or “ideal” conditions – see Donaldson, 2003).

For its part, “Program Theory-Driven Evaluation Science is the systematic use of

substantive knowledge about the phenomena under investigation and scientific methods to improve, to produce knowledge about, and to determine the merit, worth and significance of evaluands such as social, educational, health, community, and organizational programs” (Donaldson, 2007, p. 9). Effectiveness evaluations of programs or interventions employing a ToC may be considered to fall within this definition, to the degree that they use “substantive knowledge about the phenomena under investigation and scientific methods.”

Methodology Employed in Honduras

Rather than the Theory of Change Online (or TOCO) software (Center for Theory of Change, n.d.), it was decided to employ the rather more adaptable and comprehensive DoView software (see <http://doview.com/>) to outline the Theory of Change (action model) of the Honduran project, based on desk review of the project document (see Figure 1). The basic model was shown to key individuals and groups at the beginning of the fieldwork, in the national capital Tegucigalpa and the provincial capital Juticalpa, and in Gualaco, in northern forest area of the department of Olancho. In general, the respondents were in agreement with the draft ToC.

Much of the (two week) fieldwork consisted of highly productive observation (and, with permission, recording) of regular meetings of groups of co-executors and local partners, in which there was often no time for more than a brief verbal interaction with the evaluator. In such cases, it was felt that projecting and asking about the ToC, which had already been pretty much substantiated and adapted, would have been a superfluous imposition. Observations were complemented by interviews with some of the participants.

The evaluation report begins with methodological considerations and goes on to a description of the project and its development context; findings and results; conclusions, recommendations and lessons learned; and annexes. The draft report was reviewed by UNDP staff and revised accordingly.

Methodology Employed in Brazil

In addition to observations and interviews in the federal capital and the state capitals and project areas of the five project conservation units studied, it was possible, to a certain degree, to assess impact, using available data. The long-term project

goal was the conservation and sustainable use of the mangrove ecosystems and the environmental functions and services required for national development and the welfare of the coastal communities.

Baseline and comparison line data derived from the METT surveys conducted in 2006/2007 and 2012 were analyzed to verify the effectiveness of administration and management of the previously selected protected areas with mangrove forests. There are questions related to each of the following METT categories: Context, Planning, Inputs, Outputs, and Outcomes. There were 30 questions, but their assignment to the categories varied considerably between the 2006/2007 and the 2012 versions. These changes between applications (instrument decay) make longitudinal comparison quite cumbersome.

Because of the general problem of instrument decay, a special effort was made in our evaluation to compare the baseline and comparison line on each of three specific questions for which there are both baseline and comparison line data: impact on ecological and cultural values and on biodiversity, economic benefits, and access to protection systems. These were the only questions considered by the evaluators to refer to impact phenomena (to be sure, only the first two may be said to refer to impacts, strictly speaking). Note that the three respective tables were omitted from the final report by the editor.

Seeking triangulation, questions from the METT were adapted and employed in the interviews with each site manager.

Findings for Honduras

The first page of the ToC assembled by the evaluator, and slightly modified after the initial interviews, is presented in Figure 1. Note that R1 refers to outputs, R2 to outcomes and R3 to impacts. The boxes in the first column refer to the three basic areas of activity of the project, those in the second column to outputs, those in the third column to outcomes, those in the fourth column to intermediate states and those in the fifth column to expected positive impacts. Arrows show the posited results chain. Using the DoView software, each box is broken down into linked components on subsequent pages. Questions and indicators are inserted at appropriate places, with comments in the corresponding spaces below the diagram.

As for planned participation in the current project, networks of stakeholders were a key element in project success. The meetings the evaluator was able to observe reflected noteworthy

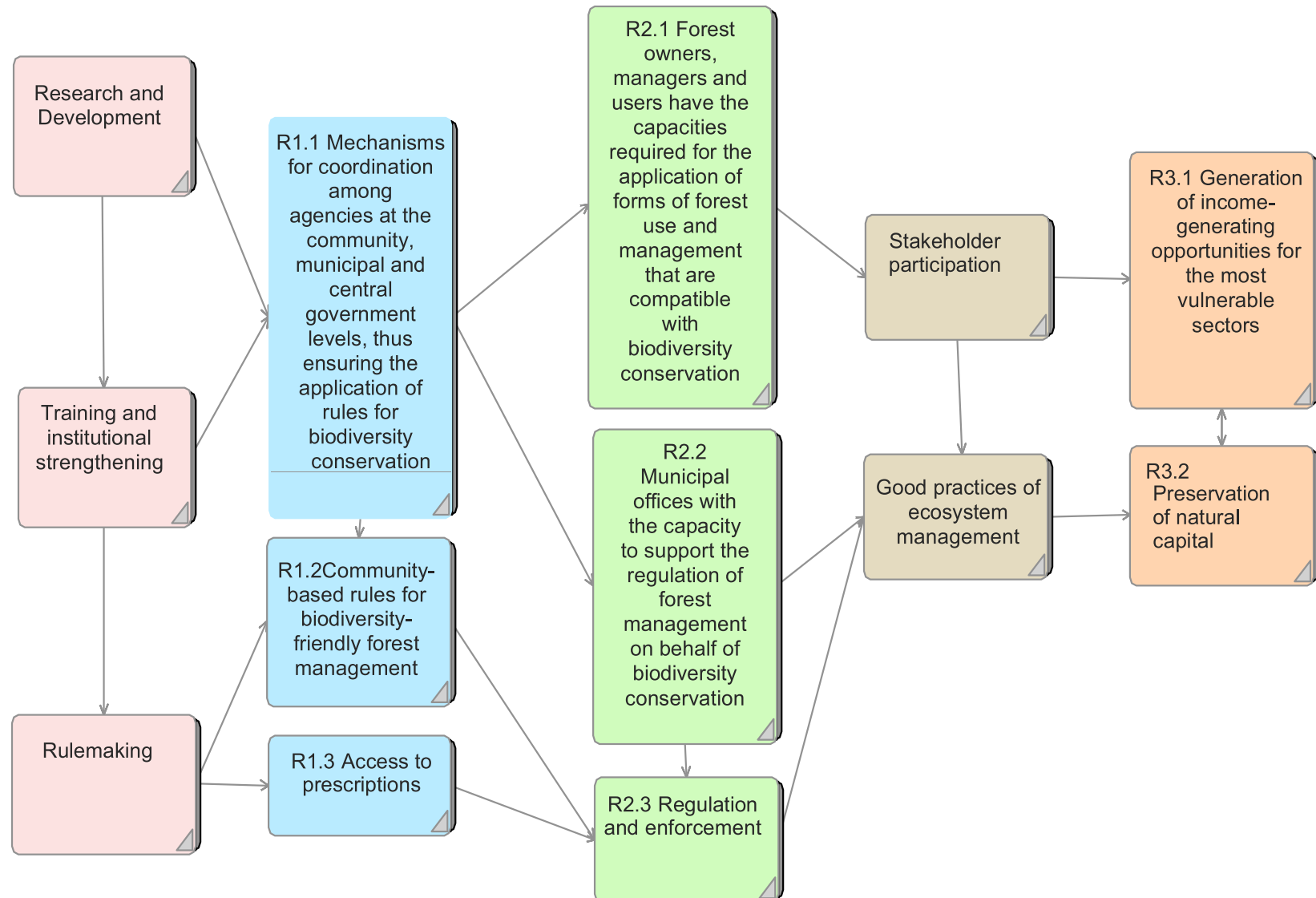


Figure 1. Theory of Change: Maintaining biodiversity conservation into the management of pine/oak forests

and widespread stakeholder participation. We concluded that the project had successfully adapted the provisions of the project document to pursue a more community- and community organization-centered approach, rather than relying on the outside consultants originally specified; this also happened to be more in line with current national policy. This has helped make the project more efficient, effective and sustainable. The approach will likely be continued and disseminated to new communities in the province, the nation and the Meso-American region. It should also be considered for adoption in similar settings worldwide. This can be done openly in future project documents by simply budgeting more for subcontracts and less for consultants (experts), provided competent and dedicated partner organizations can be identified.

The success of the approach recommended above depends on managerial competence and dedication by the implementing agency, both of which have certainly characterized this project. The UNDP office will need to consider to what degree it has the needed manpower to supervise several projects in such a careful manner, especially if the service providers (subcontracted parties) prove untrustworthy or unable to work well together (neither of which was the case in this project). The competence of the project coordinator proved to be another key factor.

Findings for Brazil

Because of the results available from the METT, we were able to include a discussion of impacts in our report. With regard to the first impact question, impact on ecological and cultural values, we observed that the mode fell, between 2006/2007 and 2012, from response 2 ("Some biodiversity, cultural and ecological values are being partially degraded, but a majority are not significantly impacted") to response 1 ("Some biodiversity, cultural and ecological values are being severely degraded"). Thirteen conservation units went from response 2 in 2006/7 to response 1 in 2012. For the other two questions, the mode remained the same.

The state of Pará offers one possible exception to the paucity of evidence for project effectiveness and positive impacts, related in part to the fact that a certain momentum had already been built up in that state prior to project start-up and in part to the early adoption through the project of a new method of transportation of the fiddler crab (*Ucides cordatus*), in baskets covered with wet sponges rather than in sacks.

Discussion and Analysis

It should not be assumed that a stress on project impact, particularly in the environmental area, implies a reduced emphasis on capacity development and social capital. Roduner, Schläppi and Egli (2008) point out that "Since the Logical Framework Approach was introduced in the 1970ies and 1980ies, a series of fundamental changes has taken place in the ways in which development assistance is delivered" (p. 8). These changes are said to include four partly overlapping areas or trends:

- from direct poverty alleviation to capacity building and social development,
- from direct implementation to a multi-stakeholder approach,
- from direct cooperation with beneficiaries to 'vertical integration,' and
- from implementing donor-driven projects to supporting partner programs.

The authors propose "a synthesis model combining LFA [the Logical Framework Approach] and OM", aiming at "bringing together the strengths of OM as an approach focusing on capacity building and LFA with its focus on development results." While we wholeheartedly agree, we would express this idea in a rather different fashion, as outlined in our book, *Impacting social problems: Writing and evaluating international development projects* (Walker, 2000). In the first chapter, it is pointed out that the UNDP Office of Evaluation and Strategic Planning (OESP, 1997) refers collectively to outputs, outcomes and impact as "results," where outcomes are "results of a programme or project relative to its immediate objectives that are generated by the programme or project outputs" (p. 13). If, as we have argued above, most immediate objectives of UNDP projects (at least in the Latin America and Caribbean area) traditionally have to do with capacity development, then "capacity" is an important "result," typically the key result pursued during the lifetime of a project. While it may be seen as a means to an end (impact), it is valued in its own right as well. Of the five DAC criteria, "effectiveness" best expresses the accomplishment of such outcomes.

The four trends listed above are well reflected in the Pino-Encino and Mangroves projects. Note that the third, vertical integration of levels of intervention (micro- meso- and macro-levels),

which was not discussed above, is also evident, through the integration of project efforts in Olancho province of Honduras with the Alliance for the Conservation of Pine and Oak in Meso-America (see <http://www.alianzapinoencino.com>). As for Brazil, it is so big that macro-level diversity is already reflected in the project's five sites, in the North, Northeast, Southeast and South of the country. Obviously, the wording of the first trend above might well be expanded a bit, to include mention of "sustainable development" or "the environment."

In addition to the fact that environmental projects typically seek positive environmental impacts (or the reduction of negative impacts), what is the rationale for the specification and assessment of impact phenomena? In the Pino-Encino Project, measurable targets were set for the desired biodiversity impact, focusing on key species. Continued monitoring should help identify trends, although co-executor specialists consulted discarded the possibility of confirming positive impacts in this area by project termination. The main benefits were awareness raising at the local level and program development at the local university.

The Mangroves project had the data needed to relate potential impacts to the other levels of effectiveness and efficiency, but simple questions of data management and lack of encouragement largely kept this from happening.

Even for non-environmental projects, we would argue for the importance of specifying impact-related targets and measuring trends. How long it may reasonably take to meet a target varies from case to case; it may require ten years for HIV infection to manifest itself as AIDS, but dehydration had better be stopped immediately; thus the argument that impacts can only be long term, while projects are of short duration, is to a certain degree fallacious.

For the two evaluations reported here, we would concur in principle with the conclusion of another UNDP/GEF mid-term evaluation report in the environmental area: "The Kazakhstan Steppe project is making good progress toward achieving the outcomes that will eventually lead to impacts, but a more detailed ROTI analysis at the end of the project (or ex-post) would be required to make a more concrete assessment of the project's contribution to impacts" (Brann and Zhakupova, 2012, p. 106).

Hopefully, any remaining ambivalence in the international community regarding the timely assessment of the full spectrum of policy, program, and project results, from efficiency in the production of outputs to effectiveness in the

accomplishment of outcomes to impacts, will be resolved, in part, through the use of Theory of Change applications such as the Review of Outcomes to Impacts (ROTI) approach.

References

- Actknowledge (2011). Introduction to TOC online. Retrieved from http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/2011_-_ActKnowledge_-_Theory_of_Change_Online_-_overview.pdf
- Belokurov, A. (2007). Management Effectiveness Tracking Tool: Reporting progress at protected area sites. Second edition. Gland, Switzerland, WWF International. Center for Theory of Change. Center for Theory of Change. Retrieved from www.theoryofchange.org/toco-software/
- Brann, J. and Zhakupova, A. (2012). Steppe conservation and management, Kazakhstan. PIMS 3835. GEF Agency: United Nations Development Programme. Executing Agency: Ministry of Agriculture: Committee on Forestry and Hunting. Astana, UNDP. Retrieved from https://www.google.com.br/?gws_rd=ssl#q=Brann+Zhakupova%2C+2012
- Chen, H. (2005). Practical program evaluation: Assessing and improving planning, implementation, and effectiveness. Thousand Oaks, CA: Sage.
- Chianca, T. (2008). The OECD/DAC Criteria for international development evaluations: An assessment and ideas for improvement. *Journal of Multidisciplinary Evaluation* 5(9), 41-45. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/issue/view/25.
- Donaldson, S. I. (2003). Theory-driven program evaluation in the new millennium. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 111-142). Mahwah, NJ: Lawrence Erlbaum Associates.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York, Taylor & Francis Group - Psychology Press.
- Etzioni, A. (1964). *Modern organizations*. Englewood Cliffs, NJ: Prentice Hall. Retrieved from http://biblsrc.btk.ppke.hu/Szociologia/o8EtzioniAmitai_Modern_Organizations.pdf.

- GEF Evaluation Office (2009). The ROTI handbook: Towards enhancing the impacts of environmental projects. Global Environment Facility and Conservation Development Centre. Retrieved from https://www.thegef.org/gef/sites/thegef.org/files/documents/M2_ROT%20Handbook.pdf
- Kinzo, D., Walker, R. and Borges, V-L (2013). BRA/07/G32, Conservação e Uso Sustentável Efetivos de Ecossistemas Manguezais no Brasil, PIMS 3280 – Projeto do Atlas 000559992, Avaliação de Meio Termo. Brasília, UNDP.
- Lempert, D. H. (2010). Why government and non-governmental policies and projects fail despite 'evaluations': An indicator to measure whether evaluation systems incorporate the rules of good governance. *Journal of MultiDisciplinary Evaluation*, 6(13), 58-108. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/243
- OECD (2014). Evaluation of development programmes: DAC criteria for evaluating development assistance. Retrieved from <http://www.oecd.org/development/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>
- Riemer, M. & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. In M. Mark, S. Donaldson and B. Campbell, *Social psychology and evaluation*. New York: The Guilford Press, 2011, pp. 104-138.
- Roduner, D., Schläppi, W., and Egli, W. (2008). Logical Framework Approach and Outcome Mapping: A constructive attempt of synthesis. Zurich, AGRIDEA and NADEL, ETH. Retrieved from http://www.outcomemapping.ca/forum/files/Discussion_Paper_OM_LFA_Synthesis_2008-1_126.pdf
- Senge, P. (1990). *The fifth discipline: The art & practice of the learning organization*. New York: Currency Doubleday.
- Uitto, J.I. (2010). Credibility and independence of evaluation in UNDP: A response to David Lempert. *Journal of MultiDisciplinary Evaluation*, 6(14), 153-158. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/285
- UNDP (2009). *Handbook on planning, monitoring and evaluating for development results*. New York, United Nations Development Programme.
- UNDP (n.d. a). *Direct Implementation (DIM) Modality*. New York, United Nations Development Programme. Retrieved from <https://info.undp.org/global/popp/frm/pages/direct-implementation-dim-modality.aspx>
- UNDP (n.d. b). *National implementation (NIM) finances and NGO implementation finances*. New York, United Nations Development Programme. Retrieved from <https://info.undp.org/global/popp/frm/pages/national-implementation-nim-Finances.aspx>
- Walker, R. (2000). *Impacting social problems: Writing and evaluating international development projects*. Brasília, EMAD.
- Walker, R. (2014). *Mid-term review of the project, Mainstreaming Biodiversity Conservation into the Management of Pine/Oak Forests*. PIMS 4210. Tegucigalpa, UNDP.