

**Evaluation of the Paris Declaration
Country/Development Partner Evaluations of phase 1**

Survey of lessons learnt during evaluation conduct

December 2008

Table of Contents

1	Introduction and methodology	3
2	Evaluation governance and management	3
3	Evaluation design and implementation	5
4	Evaluation dissemination and follow-up	6
5	Evaluation utility	7
6	Strengths/weaknesses	8
7	Proposals for phase 2	11

Table of Appendices

Annex 1	List of respondents
Annex 2	Applied methodology
Annex 3	Data collection instruments (The Questionnaires are available on request to edu@diis.dk)
Annex 4	Survey results

Acronyms

DP	Development Partner
HLF	High Level Forum
MDG	Millennium Development Goals
M&E	Monitoring and Evaluation
OECD/DAC	Organization of Economic Co-operation and Development/Development Assistance Committee
PC	Partner Country
PD	Paris Declaration
TOR	Terms Of Reference

1 Introduction and methodology

The evaluation of the Paris Declaration (PD) is not only one of the larger and more complex evaluations recently conducted; it is also unique in regard to its approach to stakeholder involvement as exemplified by the delegated TOR development for 19 partner countries (PC) and development partners (DP) as well as the establishment of national reference groups/advisory boards. This is the reason why the Evaluation Management Group has wished to conduct a survey to capitalize on the experiences from phase 1 of the evaluation- especially regarding lessons learnt from evaluation decentralization/joint conduct.

Consequently, a survey was launched in September 2008 directed at the evaluation managers¹ and evaluators of the 19 PC/DP Paris Declaration evaluation studies as well as at the synthesis team members (see annex 1 for list of respondents). Table 1 below shows the response rates for PC/DP stakeholders. The synthesis members all answered in full.

Table 1 Overview of partner country/development partner respondents

	Number of countries/partner groups contacted	Number of countries/partner groups responding	Response rate	Number of respondents
Partner countries	8	6	63%	19
Development partners	11	10	91%	21

It is noted that the response rate of the partner countries is too low to be representative with only four of eight partner countries answering in full and only three of these in a timely manner to allow for full statistical treatment.² This is why the survey report only holds limited comparative analyses of country partner/development partner responses.

This report summarizes the findings of the survey structured around the key evaluation parameters as per the questionnaire design; namely: evaluation governance and management (chapter 2); evaluation design and implementation (chapter 3); evaluation dissemination and follow-up (chapter 4); and evaluation utility in the form of capacity building and usage of evaluation results (chapter 5). The identified strengths and weaknesses are presented in chapter 6 and the proposals for PD evaluation phase II in chapter 7.

For further information on the applied survey methodology and instruments, please consult annexes 2 and 3. The detailed survey findings presented in table format are to be found in annex 4.

2 Evaluation governance and management

The vast majority of the surveyed evaluation managers (91% of PC and 89% of DP) found that the TOR framework was a useful tool for evaluation management set-up as it “clearly identified the scope and nature of the task” and “provided a common backbone to individual studies”. For those managers assessing the framework less favorably, the respondents would have preferred “more guidance (...) on methodologies”, and enhanced clarity in regard to applied terminology: “on ground level definitions and relevance of clarity, relevance and coherence were often misleading”.

Generally, the various management entities were positively assessed:

¹ Evaluation managers include national evaluation coordinators, in-country reference/advisory group members and sponsors.

² For Bangladesh and Uganda only responses from the evaluation sponsors have been received. The response from the evaluation managers and evaluators responsible of the Senegal evaluation were received too late to be included in the statistical analyses presented in chapters 2-5 below. The qualitative responses, however, have been included in the analysis as relevant.

- The *international reference group* was regarded as useful by the majority of the surveyed evaluation managers (55% PC and 88% of DP). The favorable assessment was subscribed to the experience that the reference group “gave clarity on TOR, methodology, and approaches” in the form of continuous sharing of experiences: “discussions on how to implement the evaluation required a constant dialogue – the reference group offered a venue for this dialogue to happen”. Furthermore, one in-country reference group member found that the guidance of the international reference group helped facilitate the management work of the national reference group. These positive parameters were also stressed in the assessment of the three reference group events (Copenhagen, Johannesburg, Paris) which were viewed favorably by those respondents participating in the events – both in regard to guidance, sharing of lessons learnt and networking. Especially the Johannesburg conference was regarded as critical in regard to lesson learning – not least for the synthesis team who saw “the Johannesburg workshop as absolutely essential for the Team to get a common understanding and understand the context and dynamics within the management group and steering committee”. For the respondents who found the reference group less useful, the PC respondents found the group “too ad-hoc and too distant for meaningful interaction” and this especially in regard to “coordination with donor field offices”. For the DP respondents “much more input” was expected. Both respondent groups subscribed this weak performance to the unclear role of the international reference group.
- The performance of the *management group/evaluation secretariat* was favorably assessed by the majority of the respondents, albeit for different reasons. For the evaluation managers the group/secretariat’s role in guiding the dissemination work was particularly stressed (88% of PC and 78% of DP); but also the general leadership, including facilitation of thematic papers for HLF in Accra, was viewed positively: “the management group and its leader played an important role in keeping the ball rolling in the right direction”. The synthesis team assessed the management group performance less favorably mainly due to the limited interaction between the synthesis team and the management group which did not allow for “building-up the necessary trust and mutual understanding”. Furthermore, “the introduction of an ‘advisory panel’ late in the process caused concern and generated mistrust. In terms of management group composition “the balance of donor and partner countries was found to be useful although some members were found to have too much ‘vested interest’ - too directive in how to write the report and what to write”.
- The *peer review* mechanism was positively assessed by the majority of the respondents both by evaluation managers and evaluators with especially the latter group finding the mechanism useful for “sharing of views and suggestion of changes” (100% of PC and 89% of DP). For those DP evaluators finding that the peer review mechanism did not work satisfactorily the reviews were found “to be theoretical and not contextual”.
- *In-country reference/advisory groups* were established for the majority of the evaluations (80% of PC and 78% of DP) and this primarily by appointment. The assessment of the reference/advisory groups’ performance, however, was mixed with the PC groups being most positively assessed in regard to the three parameters: speedy and transparent decision-making, productive working relations, and quality control (evaluation managers: 75%, 100%, and 88%; evaluators: 50%, 50%, 50%). The reasons stated for the less favorable assessment of DP reference/advisory groups were subscribed to lack of commitment – “engagement fell sharply after the first draft has been circulated” – as well as continuity issues: “getting the final report ready proved more cumbersome as some stakeholder comments came in very late and questioned some of the methods used”. Also “the slow transfer of information and non-adherence to agreed-upon evaluation implementation-plans” was mentioned. For those

countries/partners not establishing reference/advisory groups this was primarily due to time constraints.

- The *national evaluation coordinators* were also positively assessed by the surveyed evaluators (100% of PC and 90% of DP) and this in particular in regard to “good and timely comments on draft reports” and, for PC, “facilitation of access to key stakeholders including government officials and donors”.
- Finally, in regard to the *sponsors* of the PC evaluations it was found that where specific impact on evaluation quality was recorded, this was only in a positive sense in the form of “good feedback and help with access to key stakeholders” and “constructive inputs which didn’t try to gloss over weaknesses”. And this in spite of the fact that for 75% of the PC evaluations, one single donor financed the evaluations.

3 Evaluation design and implementation

The *specific PC/DP TOR* were positively assessed by both the surveyed evaluation managers (82% of PC and 83% of DP) and evaluators (75% of PC and 100% of DP) as an effective tool for management of evaluation results. In particular, the scope-setting of the TOR was stressed: “The TOR clearly set out the scope and expectations for the exercise, and honestly reflected the resource limitations of the exercise.” Also, the specific TOR focus was highlighted - “overall focus on commitment, competencies and incentives proved useful” –as was the overall objective: “the TOR was a yardstick by the evaluation team, the reference group as well as the two peer reviewers”. For those assessing the TOR less favorably it was stated that “it proved difficult to distinguish PD-related issues from other aspects including PD-related results since many initiatives were taken prior to the launching of the PD”. Furthermore, some found the TOR too constraining: the TOR were “too structured – did not allow for country-specific utilisation focus – only aimed to meet needs of overall report”. In terms of compromises made in the formulation of the specific TOR, the PC reported on negotiations regarding the selection of sector/cross-cutting themes for evaluation.

The *TOR for the synthesis report* were negatively assessed by all members of the synthesis team as summarized in the following statement: “The TORs were too complicated, and the elements that read across the donor and partner country studies were not strong enough to provide a fully robust comparative study. Hinging the Synthesis TORs to the 12 Survey indicators was a major, unrecognized mistake that we had to quietly rectify. The experiment with trying to assess commitment, capacity and incentives was a partial success, although most studies did/could not grapple with the concept of incentives effectively.” Furthermore, several synthesis team members would have liked “to be part of the process of developing TOR”. In regard to the specific PC/DP TOR, the synthesis team found that “the different TORs for the different types of reports meant a missed opportunity for using one to validate evidence from the other”.

There was general agreement on the speediness and effectiveness of the *bidding processes* for the partner countries with 75% of the surveyed evaluation managers assessing the process positively. However, for those countries applying national procedures, the process was in at least one instance seriously delayed and attempts of process “manipulation” were recorded. For the Netherlands no bidding process was conducted as the evaluation was done in-house by the Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs with the evaluation team leader also serving the role as evaluation coordinator.

In regard to *communication measures* between evaluation managers and the evaluation teams, the surveyed PC mainly applied workshop modalities (initial briefing workshop 89% and workshop to discuss preliminary findings 78%). The DP, however, also applied other communication measures such as the recruitment of support persons to assist the evaluation department with the

communication tasks. The communication was deemed satisfactory by all evaluators participating in the survey. For the synthesis team, the consultants would have liked to expand the communication to national evaluation coordinators in order to “ensure synthesisability”.

There was general satisfaction with the *applied methodology* from surveyed evaluation managers (89% for PC and 86% for DP) and evaluators (100% for PC and 88% for DP) alike. For those respondents being dissatisfied with the applied methodology, the overall evaluation timeframe was found to be constraining. Furthermore, the linkage to the Paris Declaration monitoring survey outcome was found not to be sufficiently explored.

The *commenting on draft inception- and evaluation reports* was generally given in a consolidated manner (75% for PC and 67% for DP), however still leading to 3-4 rounds of comments, especially for partner countries. The majority of the comments were not conflicting (56% for PC and 63% for DP), although especially the evaluation reports went through several drafting rounds.

In terms of the *PC/DP specific evaluation output*, the majority of the surveyed evaluation managers (83% for PC and 100% for DP) and evaluators (100% for PC and 80% for DP) were satisfied with the end-results. For those finding the quality of the reports wanting, it was found that “the reports lacked sector specific focus” or suffered from inadequate donor sampling leading to too general analyses; and (for DP) “that embassies could have been more involved”. Furthermore, PC and DP alike found that the tight evaluation time-frame affected the evaluation results negatively in terms of insufficient quality control.

In regard to the *synthesis report*, the majority of the surveyed evaluation managers (71% for PC and 89% for DP); evaluators (50% for PC and 80% for DP), and synthesis team members (100%) expressed satisfaction with the final output. In terms of identified short-comings, limited cross-referencing to the PC/DP studies was mentioned.

4 Evaluation dissemination and follow-up

The majority of the survey evaluation managers developed a *dissemination plan* (55% for PC and 63% for DP). For those not developing dissemination plans this was either “not considered”; not found relevant (“no need for specific plan”); or not implemented due to change of staff: “We have suffered 4 changes of authorities, Ministries and Vice-ministries”. In terms of guidance for the development work, only the evaluation managers of the PC (40%) applied the dissemination note of the evaluation secretariat, whereas the surveyed evaluation managers of the DP either tapped into the dissemination of the synthesis report, applied existing in-house procedures or found the note not relevant: “The dissemination guidelines were more suitable for country evaluations”.

In terms of specific *dissemination activities*, the PC applied meetings/workshops as their main measure whereas DP also applied web-sites (see table 2 below).

Table 2 Dissemination activities undertaken for evaluation reports (% agreed)

Which dissemination measures were applied?	Conduct of meetings/workshops	Distribution of press kit	Development of web-sites	Other
Partner country	100%	17%	17%	33%
Development partner	57%	0%	57%	57%

In terms of other measures, reference was made to newsletters, presentations at the European Association of Development Research and Training Institutes (EADI) and distribution of evaluation reports to key stakeholders, including universities and libraries. None of the surveyed evaluation managers applied mass media (radio, television, newspapers) as dissemination measures. Evaluators

(50% for PC and 30% for DP) and synthesis team members (25%) were involved in the dissemination efforts to some extent.

In terms of the *targeted audiences*, government officials constituted the main group for PC and DP alike (see table 3 below).

Table 3 Targeted audiences for evaluation reports (% agreed)

Who were the targeted audiences?	Government officials	Parliamentarians	Donors	Civil Society	Private sector	Media	Other
Partner country	100%	14%	86%	86%	14%	14%	14%
Development partner	71%	57%	0%	57%	14%	14%	57%

For the PC, civil society and donors were also regarded as key target groups. For the DP, the parliamentarians were regarded as important a target group as civil society. Furthermore, DP also viewed their own staff and researchers as important target groups. All targeted audiences were reported to be met by both surveyed PC and DP evaluation managers; albeit a number of these highlighted that the meeting of audiences had not been subject to measurement/assessment as such.

A majority of the surveyed DP (40% for PC and 75% for DP) had developed *action-plans* addressed to the Minister for Development for follow-up. For the majority of the PC/DP presently not holding action-plans, development of such plans were scheduled: “(plan) in process to consolidate findings not only of the evaluation but also of other interventions, which will be combined in one Aid Effectiveness action plan to be finalized by beginning 2009”. Finally in terms of *follow-up actions* already taken, reference was primarily made to process results such as the “convening a core action group post-Accra” and the approval of “a pilot programme (...) in the pool funding mechanism at the request of member states”.

5 Evaluation utility

Evaluation utility refers to two aspects, namely evaluation contribution to capacity building and usage of evaluation results.

In regard to the *building of evaluation capacities*, the majority of the surveyed PC evaluation managers (80% for PC and 33% for DP) reported a positive impact in the form of enhanced understanding of several stakeholders: “The consultants recruited for PD evaluation process was local consultants and through the process, capacity of local consultants as well as government officials were greatly enhanced. Moreover, the process also served a lot to enlighten DP officials.” Other responses were more nuanced in the form of the longer term capacity impact attained: “The experience gained will probably be of great value to team members and probably to some officials which participated of the process, however there is little evidence that this will contribute to strengthen national or government capacities.”

In terms of the *usage of evaluation results*, the majority of the surveyed PC/DP evaluation managers reported on the two pre-identified usages: evaluation contribution to country/agency preparation for the Accra meeting on aid effectiveness (80% for PC and 100% for DP); and utilization of the evaluation as an instrument for domestic dialogue on the Paris Declaration (80% for PC and 63% for DP). Furthermore, other usages were reported such as “possible inputs into sector evaluations”; “promotion of the aid effectiveness public debate”; input into “the ongoing elaboration of the national policy for external aid”; and input to guideline revision: “Ministry of foreign affairs plans to revise aid management guidelines following Accra meeting. It is expected that the lessons learned from the evaluation will be used in this connection.”

Finally, in regard to evaluation insights to be applied for future joint evaluations, the majority of the surveyed PC/DP evaluation managers (100% for PC and 75% for DP) identified insights for replication such as those relating to management experiences: “The joint management group put in place to coordinate several evaluation offices of participating UN agencies can be further replicated in other joint evaluations. A reference group that discusses the process and implications of joint evaluations is valuable. The peer review mechanism is another good practice for joint evaluations.” Others identified experiences for non-replication: “... the lesson relates to the international reference group (should be more active) as well as to the discipline of those who conduct the parallel evaluations (should be more stringent for synthesis purpose”); and process/methodology issues “At least there should be adequate time for each phase of the process. There should be allowance to replace some methods, if they have not yielded adequate response interest.”

6 Strengths/weaknesses³

In the following, the strengths and weaknesses of the evaluation process as identified by the respondents are presented:

Strengths

- *Evaluation framework/methodology.* Several evaluation managers and evaluators of both PC and DP found the “comprehensive analytical framework” to be a key strength of the evaluation process. One respondent qualified the statement further by adding: “I had expected a larger potential for 'mirroring' the donor evaluations and the partner country studies”. Another respondent appreciated the possibility of TOR “tailoring to specific interests” while noting the trade-off in terms of “difficulty for synthesis team”. Two synthesis team members, however, found that the “methodology (was) well defined” and that the TOR held “clear objectives”.
- *Wide-scaled engagement.* Several evaluation managers stressed overall evaluation *participation*: “Ample participation of countries and development agencies in an independent evaluation process. Twenty (19) evaluations conducted under common TOR is a landmark achievement”. Others (evaluation managers) stressed the spirit of *partnership* in which the evaluation was conducted; such as the fact that the evaluation was viewed as “a common donor/partner initiative” as well as the experiences that “the voices of donors and partner countries were listened (to) alike. One respondent also highlighted the importance of “a committed sponsor”. The evaluation itself was an expression of mutual accountability.” Finally, engagement in the form of (joint) *ownership* was referred to by several respondents as reflected in the “participatory approach” of stakeholders including “in-country reference group members mobilized (by self-appointment) in the (evaluation) process”.
- *Evaluation management.* Several evaluation managers and evaluators praised the overall management of the evaluation whether in the form of feed-back – “regular updates on OECD/DAC web-page”; or in the form of time management: “Clear and precise dateline for various stages of the evaluation process”. Also, the quality of the secretariat and the management group was highlighted: “Vigorous and constructive role of the evaluation management group and its chair”. Furthermore, one evaluation manager assessed the evaluation decentralization as critical: “Division of labor leaving individual evaluations to national coordinators”.

³ Chapter 6 and 7 have been drafted with the purpose of presenting the maximum amount of inputs (even if some of the inputs only represent the view point of a single respondent) with the overall objective of maximizing the lesson learning.

- *Forum for lesson sharing.* Several evaluators appreciated the lessons sharing in the context of the international reference group: “Brainstorming at international level between donors and recipients (allowed for) international exchange of lessons learnt and streamlining of perceptions”. Also the manner in which the findings were shared was appreciated: “Open/transparent manner in which findings were shared with all involved”.
- *Peer review.* Several evaluators and evaluation managers appreciated the inputs from the peer reviewers: “Useful inputs, feedback and guidance provided by the two peer reviewers”.
- *Evaluation team.* Several evaluators and synthesis team members highlighted “team diversity” as a key strength. Also “excellent selection of team leader” was mentioned. The evaluation managers stressed “the professional quality of the evaluators” and “the utilization of local consultants”.
- *Access to relevant material.* Several respondents highlighted the access to “substantial relevant material” as a key strength.
- *Candor.* The candidness of the evaluation reports was another strength mentioned by several of the respondents: “it is noteworthy that in the end there was still an impressive degree of candor in most reports”.
- *Reference to political economy.* Others (evaluation managers) stressed similar factors by highlighting the identification of political economy parameters: “The conclusion that realizing the principles of national ownership, alignment, harmonization, managing for results and mutual accountability is a political process and requires commitment, capacities and incentive systems from both sides”. Also “evidence on which political action could be based” was stressed.
- *Capacity building.* For PC the building of evaluation capacities was stressed by one evaluation manager. Others stressed the learning in regard to the PD principles.
- *Dissemination/follow-up.* Finally, some evaluation managers highlighted the reaching of the targeted audiences as critical.
- *Increased awareness of the PD agenda.* Several evaluators found that the conduct of the evaluation has helped increase the awareness and ownership of the aid effectiveness agenda both at PC and DP level as well as at international level (DAC/HLF 3). Moreover, the evaluation helped “initiate continuous and formal Paris Declaration follow up process into the donor community” in several of the PC.
- *New insights.* One evaluation manager found that the evaluation helped “identify some indicators that were neither appropriate nor adequate”. Also, “focus areas for improvement” were identified.

Weaknesses

- *Too tight time-frame.* Many of the surveyed evaluation managers and evaluators found the evaluation time-frame too tight: “timing was inadequate for in-depth study” and as a consequence “we relied on secondary data and in many cases one simply had to accept what we were told (e.g. that all missions were joint, that reviews were done jointly and so on)”. Also it was found that the preparatory work was “too slow (...) at international level compared to the time given to the actual evaluation work at country level”. Also the synthesis work suffered due to the tight time-frame: “The inordinate delays in producing the individual

reports (due to) time frame too compressed (and) comments received late (led to) insufficient time to absorb/ synthesize and reflect on the material submitted”.

- *Imperfect sample.* The PC/DP sampling was found to be inadequate by several respondents: “Insufficient coverage of donors (some important players were not in which may have discredited the overall evaluation!!!) as well as important partner countries (e.g. under representation of Latin American countries)”. One respondent also stressed the “difficulties in choosing and addressing the sample of donors and projects” at country-level.
- *Uneven quality.* Some evaluation managers regretted the uneven quality of the PC/DP evaluations.
- *Narrow evaluation focus.* Several evaluation managers found that the evaluation focus did not allow for other issues such as “collaboration on monitoring and evaluation” to be evaluated.
- *Superficial analyses.* At the same time, several of the respondents found the analyses too superficial either due to “unfortunate choice (of sectors) leading to generalized findings”; or due to “evaluation focus on trends and not concrete results” leading to “over-generalized findings about aid processes”; or due to a “too comprehensive approach”.
- *Weak guidance/methodology.* Some respondents found that the quality/validity of the findings suffered due to “methodological weaknesses” and the “somewhat confusing guidance from the international level”. For the synthesis team the “inconsistency in the reporting format” was highlighted as a weakness.
- *Evaluation management.* The majority of the synthesis team members found the mistrust between the management group and the synthesis team to be a key weakness.
- *Scarcity of pertinent evaluators.* One respondent highlighted as a weakness the “scarcity of international evaluators with expertise in the area of aid effectiveness” leading to imperfect selection processes.
- *Insufficient understanding of evaluation theme.* Several evaluation managers found that the “newness” of the aid effectiveness theme required pre-evaluation awareness work including “national workshops (to) be held to familiarize the subject and also to indicate the forthcoming evaluation, explaining why it is important to evaluate”. Also the evaluators regretted the absence of pre-evaluation learning sessions: “Opportunity should be provided at the onset to allow evaluation teams (or at least the respective team leaders) to meet for a couple of days to ensure that everyone has the same understanding of the methodology to be used. We were fortunate in that we started late and could learn from the mistakes made by some of the evaluation teams that had started before us.”
- *Low engagement from reference group.* Some evaluation managers found the engagement of the reference group wanting leading to “a very slow emergence of understanding what this evaluation was about! This also led to a very low level (in fact non-existing!) of commenting on each others evaluation reports and prior to that a very limited sharing of specific TORs “.
- *Weak provincial participation.* One of the partner country evaluation managers regretted the weak provincial participation in the evaluation.

- *Language barriers.* Some of the respondents found the language barrier an important weakness: “most of the documents were just in English, and the translations took us much time”.
- *Excessive resources requirements.* One evaluation manager found that the evaluation was “very resource consuming (in terms of) time and travel budget.”
- *Weak dissemination.* Finally, several respondents found the dissemination of the final report lacking.

7 Proposals for phase 2

The following recommendations for phase 2 were identified by the respondents:

- *Launch evaluation process early.* Several respondents proposed to start the evaluation process as early as possible.
- *Apply uniform and simple methodology.* It was recommended by several respondents to apply uniform and simple methodology for all evaluation reports for phase 2. Although a respondent also stressed the need for some contextualization: “some thought should be given to the context within which donor harmonization etc is happening and how this could be better accommodated next time round”. Contextualization of evaluation questionnaires was proposed as well. Further, it was proposed to “have the shape and objectives of the synthesis available from the outset, so participants can understand how the evidence they provide will be used”.
- *Maintain management structure.* It was proposed to “keep the same management structure: international reference group, small management group and peer reviewers (while) trying to engage the international management group much more in the process”. Furthermore, it was also proposed to complement the peer reviewers at central level with “other donors – partners” peer reviewers.
- *Include M&E theme for evaluation.* One PC evaluation manager proposed to include “collaboration on M&E between donors and partner countries in the next evaluation”.
- *Consider joint evaluations.* It was proposed to include “novel patterns of joint evaluations (such as) joint evaluation between a donor, multi-agency and a partner country, which would look at all three dimensions in one particular partner-country and vice versa“.
- *Apply cross-country teams.* It was also proposed to have cross-country teams conducting the evaluations.
- *Enhance quality control.* Several respondents recommended an enhanced focus on quality control for phase 2 “if not the synthesis process will be much more difficult than has been the case in phase one. Why? Because phase two focuses on RESULTS (and) will (hence) be of a less formative nature than the phase one evaluation”. Furthermore, it was recommended to establish clear procedures up-front including those relating to review/quality control mechanisms.
- *Capacitate evaluators.* One evaluator proposed to involve potential evaluators to participate in workshops, as relevant, in order to broaden their knowledge. Furthermore, it was proposed to establish an evaluator database with the purpose of ensuring lessons sharing.

- *Institutionalize follow-up.* It was proposed to institutionalize evaluation follow-up for both PC and DP.
- *Conduct scoping study.* Finally, it was proposed to “open the box of what is currently called "public development aid" and analyze it before going further. Mixing global public goods, MDGs, promotion of donors' interests etc. in the same box of "development aid" and applying the Paris Declaration approach to all that without further analysis is not effective and could be dangerous, notably for the incentives to increase the resources and the effectiveness of actions undertaken to address Global Public Challenges.”

Annex 1 List of respondents

Country/development partner evaluated	Name	Role
Partner countries		
Bangladesh	Kayo Torii	Sponsor (Japan)
Bolivia	Elizabeth Ascarrunz	National Evaluation Coordinator
	Jose Manuel Mariscal	Sponsors (Spain)
	Laura González Garcés	
	Eduardo Zeballos	Evaluation team leader
	Oscar Angulo	Evaluation team member
Senegal	Amadou Tidiane Dia	Deputy ⁴ National Evaluation Coordinator and in-country reference/advisory group member
	André Ndecky	In-country reference/advisory group member
	Werner Meier ⁵	Evaluation team leader
	Madeleine Guay	Evaluation team member
South Africa	Elaine Venter	National evaluation coordinator
	Mary Cole	In-country reference/advisory group member
	Lisa du Toit	
	James Watson	Sponsor (US)
	Matthew Smith	Evaluation team leader
	Kerstin Waddell	Evaluation team member
Sri Lanka	V. Sivagnanasothy	National evaluation coordinator
	A. S. Gunawardena	Evaluation team leader
Uganda	Michael Weiner	Sponsor (Austria)
Development partners		
Asian Development Bank	Samjhana Shrestha	Evaluation team member
	Leah Gutierrez	Evaluation team member
	Michael Pilbrow	Evaluation team leader
Denmark	Margrethe Holm Andersen	National evaluation coordinator
	Lars Koch	In-country reference/advisory group member
	Peter Ellehøj	
	Jørgen Billetoft	Evaluation team leader
	Dorte Kabell	Evaluation team member
Finland	Aira Paivoke	National evaluation coordinator
	Jyrki Salmi	Evaluation team leader
France	Benoît Chervalier	National evaluation coordinator
	Luc Lefebvre	Evaluation team leader
Germany	Michaela Zintl	National evaluation coordinator
	Guido Ashoff	Evaluation team leader
Luxembourg	Georges Ternes	National evaluation coordinator
	Robert Kremer	Evaluation Team Leader
Netherlands	Ted Kliest	National evaluation coordinator and Evaluation team leader
United Kingdom	Alison Girdwood	National evaluation coordinator

⁴ Mr. Dia was in charge of leading the PD evaluation exercise on behalf on the National Coordinator

⁵ The two evaluators of the Senegal evaluation have submitted one joint questionnaire.

Country/development partner evaluated	Name	Role
Development partners		
United Nations Development Group (UNDG)	Oscar A. Garcia	Evaluation coordinator
	Hans Wyss	Evaluation team leader
	Janie Eriksen	Evaluation team member
	Neddy Matshalaga	Evaluation team member
Synthesis team		
Team members	Bernard Wood	Team leader
	Francisco Sagasti	Team member
	Nansozi K. Muwanga	Team member
	Dorte Kabell	Team member

Annex 2 Applied methodology

To meet the study focus – lessons learnt in regard to evaluation decentralization/joint conduct – the scope of the study has been limited to those stakeholders involved in the conduct of country/development partner evaluations including a total of 19 different stakeholder groups; namely: Asian Development Bank (ADB), Australia, Bangladesh, Bolivia, Denmark, Finland, France, Germany, Luxembourg, Netherlands, New Zealand, Philippines, Senegal, South Africa, Sri Lanka, United Kingdom, United Nations Development Group (UNDG), Uganda, and Vietnam.

In order to further delimit study scope, the applied sampling criteria for themes, instruments, and respondents have been as follows:

The themes covered in the study are meant to represent the key evaluation elements; namely: evaluation management, evaluation design/implementation, evaluation dissemination/follow-up, and in the extent possible evaluation utility (capacity building/usage of evaluation results).

The data collection instrument has been that of questionnaires, for select respondents.. Three distinct sets of questionnaires have been applied to meet the different targeted respondent groups (see below). However, in order to facilitate speedy response rate, all questionnaires have applied the same design format; namely: closed questions with opportunity to offer qualitative responses as needed (see annex 3).

In regard to respondents, three different groups have been targeted for survey:

- Evaluation managers: national evaluation coordinators, in-country reference/advisory group members, and sponsors.
- Evaluators: evaluation team leaders, national evaluation team members, and international evaluation team members.
- Synthesis team: synthesis team leader and synthesis team members.

In terms of data treatment the database SPSS Version 16.00 was applied.⁶

The limitations having a potentially damaging effect on the overall quality of the study have been identified as follows:

- *Low response rate.* The response rate of the partner countries is too low to be representative. This is why in the survey report only holds limited comparative analyses of country partner/development partner responses.
- *Questionnaire design.* Some of the questions asked were not clear to all respondents leaving to potentially faulty answers. Retrospectively, key terms should have been defined up-front. Furthermore, for some questions more response categories should have

⁶ Specific interpretation rules were formulated prior to the keying of the results. Missing values, defined as a question without a clear answer mark or a double answer mark (were not allowed), were coded with “9999”. “Yes” and “No” questions were coded “1” and “2” respectively, and questions with more response categories were coded “1, 2, 3, 4” and so forth. Where it was possible to respond with a qualitative answer, the results were listed in the continuation of the specific question. The answers were not coded or reformulated in any sense. The analysis of the questionnaire provided the following data: response rates for each question together with a per cent value of how many that agreed with the specific answering categories. Each analysis was performed with the distinctions of partner country and a development partner, due to the possible significant differences in their perception of the whole evaluation process; and this per questionnaire category: evaluation managers, evaluators, and synthesis team

been included such as “I did not participate in the workshop” for the questions asking evaluation managers about the quality of the Copenhagen/Johannesburg/Paris workshops.

- *Timing.* Given that the finalization of the various country/development partner evaluations is still fairly recent (maximum of 6 months), evaluation utility in terms of built evaluation capacity and/or evaluation results usage might be too early to detect.

The team carrying-out the evaluation is Rikke Ingrid Jensen (team leader) and Simon Feldbæk Kristensen (student). Finn Skadkær has been responsible for the quality assurance.

Annex 3 Data collection instruments

The Questionnaires are available on request to edu@diis.dk

Annex 4 Survey results)⁷

Table 4 Survey responses of evaluation managers (%agreed)

Q1 - Was the Paris Declaration evaluation TOR framework an effective tool for evaluation management set-up?	Yes	No	Response rate	
Partner countries	91%	9%	92%	
Development partner	89%	11%	82%	
Q2 - Did you view the utility of the international reference group as evaluation organ as satisfactory?	Yes	No	Response rate	
Partner country	55%	45%	92%	
Development partner	88%	13%	73%	
Q3 - Was the June 2007 Copenhagen inception workshop satisfactory in terms of o*?	Guidance	Sharing of lessons learnt	Networking	Response rate ⁸
Partner country	80%	40%	80%	42%
Development partner	40%	40%	100%	55%
Q4 - Was the January 2008 Johannesburg workshop satisfactory in terms of	Guidance?	Sharing of lessons learnt?	Networking?	Response rate
Partner country	63%	88%	50%	67%
Development partner	50%	83%	50%	55%
Q5 - Was the April 2008 Paris workshop satisfactory in terms of	Guidance?	Sharing of lessons learnt?	Networking?	Response rate
Partner country	100%	67%	67%	25%
Development partner	86%	43%	43%	64%
Q6 - Did the management group/evaluation secretariat work satisfactorily in terms of	Advice/support in regard to identification/selection of consultants?	Development and implementation of dissemination strategy?	Response rate	
Partner country	50%	88%	67%	
Development partner	33%	78%	82%	
Q7 - Did the peer review mechanism work satisfactorily in terms of quality assurance?	Yes	No	Response rate	
Partner country	89%	11%	75%	
Development partner	100%	0%	73%	
Q8 - Did you establish an in-country reference/advisory group?	Yes	No	Response rate	
Partner country	80%	20%	83%	
Development partner	78%	22%	73%	
Q9 - How were the members of your in-country reference/advisory group selected?	By appointment	Auto-selection	Response rate	
Partner country	56%	44%	75%	
Development partner	63%	38%	73%	

⁷ The response from the evaluation managers and evaluators of the Senegal evaluation were received too late to be included in the statistical analyses presented.

⁸ A number of respondents could not attend the workshops, which explain the low response rate in Q3, Q4 and Q5.

Table 4 Survey responses of evaluation managers (%agreed) , continued

Q11 ⁹ - Did the in-country reference group work satisfactorily in terms of	Speedy and transparent decision-making?	Productive working relations among key stakeholders?	Quality control?	Response rate
Partner country	75%	100%	88%	67%
Development partner	27%	36%	45%	55%
Q12 - How was the evaluation financed?	Single sponsor	Multiple sponsors	Auto-financed	Response rate
Partner country	75%	25%	0%	100%
Development partner*	38%	13%	63%	73%
Q13 - Who was the lead drafter of the specific partner country - development partner TOR?	National evaluation coordinator	Development Partner(s)	Other	Response rate
Partner country	50%	0%	50%	83%
Development partner	50%	13%	38%	73%
Q14 - Did compromises by stakeholders have to be made in the development of the TOR?	Yes	No	Response rate	
Partner country	44%	56%	75%	
Development partner	13%	88%	73%	
Q15 - Did the specific country/development partner TOR serve as an effective instrument for managing of evaluation expectations /results?	Yes	No	Response rate	
Partner country	82%	18%	92%	
Development partner	83%	17%	55%	
Q16 - Whose bidding procedures were applied in the selection of evaluators?	National partner procedures	Sponsor/development partner procedures	Response rate	
Partner country	75%	25%	67%	
Development partner	Not relevant for DP's			
Q17 - Did the bidding procedures allow for a speedy and effective selection process?	Yes	No	Response rate	
Partner country	80%	20%	83%	
Development partner	Not relevant for DP's			
Q18 - Which measures were put in place to ensure effective communication between evaluation managers and evaluators?	Initial briefing workshop	Workshop to discuss preliminary evaluation findings	Other	Response rate
Partner country	89%	78%	44%	75%
Development partner	67%	67%	67%	55%
Q19 - Were conflicting comments to the proposed methodology/inception report put forward?	Yes	No	Response rate	
Partner country	44%	56%	75%	
Development partner	38%	63%	73%	

⁹ Question 10 responses (names and titles of in-country reference/advisory group members) have not been listed in the table due to their qualitative nature.

Table 4 Survey responses of evaluation managers (% agreed), continued

<i>Q20 - Was the applied methodology adequate for delivering the desired results?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	89%	11%	75%
<i>Development partner</i>	86%	14%	64%
<i>Q21 - Was sufficient guidance/back-up provided to the evaluation team to allow for adequate data collection?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	100%	0%	75%
<i>Development partner</i>	80%	9%	45%
<i>Q22 - Were conflicting comments to the evaluation report put forward?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	44%	56%	75%
<i>Development partner</i>	50%	50%	45%
<i>Q23 - Are you satisfied with the final evaluation output?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	83%	17%	100%
<i>Development partner</i>	100%	0%	73%
<i>Q24 - Was a dissemination plan for the country evaluation designed?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	55%	45%	92%
<i>Development partner</i>	63%	38%	73%
<i>Q24a - If yes, did the dissemination guiding note guide the development?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	40%	60%	42%
<i>Development partner</i>	0%	100%	36%
<i>Q27¹⁰ - Were the targeted audiences met?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	100%	0%	50%
<i>Development partner</i>	100%	0%	36%
<i>Q28 - Were action plans for implementing the evaluation recommendations issued?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	40%	60%	42%
<i>Development partner</i>	75%	25%	36%
<i>Q29 - Were any follow-up actions taken?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	80%	20%	42%
<i>Development partner</i>	100%	0%	36%
<i>Q30 - Did the evaluation contribute to the building of in-country evaluation capacities?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	80%	20%	83%
<i>Development partner</i>	33%	67%	55%
<i>Q31 - Has the evaluation contributed to country/agency preparation for the Accra meeting on aid effectiveness?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	80%	20%	83%
<i>Development partner</i>	100%	0%	82%
<i>Q32 - Has the evaluation been utilized as an instrument for domestic dialogue on the PD?</i>	<i>Yes</i>	<i>No</i>	<i>Response rate</i>
<i>Partner country</i>	80%	20%	83%
<i>Development partner</i>	63%	38%	73%

¹⁰ Please consult chapter 4 for detailed responses to questions 25 and 26.

Table 4 Survey responses of evaluation managers (% agreed), continued

Q33 - Is there any other planned use of the evaluation and its result?	Yes	No	Response rate
Partner country	100%	0%	50%
Development partner	33%	67%	55%
Q34 - Will the insights from the evaluation process be of use in the future joint evaluations that you may undertake?	Yes	No	Response rate
Partner country	100%	0%	67%
Development partner	75%	25%	73%
Q35 - Are you satisfied with the manner in which your country evaluation findings are reflected in the synthesis report?	Yes	No	Response rate
Partner country	71%	29%	58%
Development partner	89%	11%	82%

(n=23, Partner countries n=12, Development partners n=11)

Table 5 Survey responses of evaluators (% agreed)

Q1 - Did the in-country reference group work satisfactorily in terms of	Speedy and transparent decision-making?	Productive working relations among key stakeholders?	Quality control?	Response rate
Partner country	50%	50%	50%	50%
Development partner	73%	64%	64%	100%
Q2 - Did the national evaluation coordinator impact the evaluation quality?	Yes	No	Response rate	
Partner country	100%	0%	75%	
Development partner	90%	10%	91%	
Q3 - Did the composition of reference group impacted the evaluation quality?	Yes	No	Response rate	
Partner country	25%	75%	100%	
Development partner	90%	10%	91%	
Q4 - Did the sponsor impact the evaluation quality?	Yes	No	Response rate	
Partner country	50%	50%	100%	
Development partner	56%	44%	82%	
Q5 - Did the peer review mechanism worked satisfactorily?	Yes	No	Response rate	
Partner country	100%	0%	100%	
Development partner	89%	11%	82%	
Q6 - Did the country-specific TOR serve as an effective instrument for managing of evaluations results?	Yes	No	Response rate	
Partner country	100%	0%	100%	
Development partner	89%	11%	82%	
Q7 - Did the bidding procedures allow for a speedy and effective selection process?	Yes	No	Response rate	
Partner country	50%	25%	75%	
Development partner	57%	43%	64%	

Table 5 Survey responses of evaluators (%agreed), continued

Q8 - Which measures were put in place to ensure effective communication with the evaluation managers and evaluators?	Initial Briefing Workshop	Workshop to discuss preliminary findings	Response rate	
Partner country	100%	75%	100%	
Development partner	78%	56%	82%	
Q9 - Were the communication measures satisfactory?	Yes	No	Response rate	
Partner country	100%	0%	100%	
Development partner	100%	0%	82%	
Q10 - Were conflicting comments to the proposed methodology/inception report put forward?	Yes	No	Response rate	
Partner country	67%	33%	75%	
Development partner	60%	40%	91%	
Q11 - Was the applied methodology adequate for delivering the desired results?	Yes	No	Response rate	
Partner country	100%	0%	75%	
Development partner	88%	13%	73%	
Q12 - Was the sufficient guidance/back-up provided to evaluation team to allow for adequate data collection?	Yes	No	Response rate	
Partner country	75%	25%	100%	
Development partner	89%	11%	82%	
Q13 - How many rounds of comments were given to the evaluation report?	1-2 times	3-4 times	≥5 times	Response rate
Partner country	0%	67%	33%	75%
Development partner	50%	20%	30%	91%
Q14 - Were the comments given in a consolidated manner?	Yes	No	Response rate	
Partner country	75%	25%	100%	
Development partner	67%	33%	82%	
Q15 - Are you satisfied with the final evaluation output?	Yes	No	Response rate	
Partner country	100%	0%	100%	
Development partner	80%	20%	91%	
Q17 - Are you satisfied with the manner in which your country evaluation findings are reflected in the synthesis report?	Yes	No	Response rate	
Partner country	50%	50%	100%	
Development partner	88%	13%	73%	
Q16 - Were you directly involved in the planning/conduct of the dissemination efforts?	Yes	No	Response rate	
Partner country	50%	50%	100%	
Development partner	30%	70%	91%	
n=15, Partner countries n=4, Development partners n=11				

Table 6 Survey responses of synthesis team (% agreed)

Q1 - Did the in-country reference group work satisfactorily?	Speedy and transparent decision-making	Productive working relations among key stakeholders	Quality control	Response rate
Synthesis team	100%	0%	0%	25%
Q2 - Did the composition of the management group impacted the evaluation quality?	Yes	No	Response rate	
Synthesis team	100%	0%	100%	
Q3 - Did the country-specific TOR serve as an effective instrument for managing of evaluations results?	Yes	No	Response rate	
Synthesis team	0%	100%	50%	
Q4 - Did the bidding procedures allow for a speedy and effective selection process?	Yes	No	Response rate	
Synthesis team	100%	0%	50%	
Q5 - Which measures were put in place to ensure effective communication with the evaluation managers?	Initial Briefing Workshop	Workshop to discuss preliminary findings	Other	
Synthesis team	33%	100%	33%	
Q6 - Were the communication measures satisfactory?	Yes	No	Response rate	
Synthesis team	75%	25%	100%	
Q7 - Were conflicting comments to the proposed methodology/inception report put forward?	Yes	No	Response rate	
Synthesis team	50%	50%	100%	
Q8 - Was the applied methodology adequate for delivering the desired results?	Yes	No	Response rate	
Synthesis team	75%	25%	100%	
Q9 - How many rounds of comments were given to the evaluation report?	1-2 times	3-4 times	≥5 times	
Synthesis team	0%	50%	50%	
Q10 - Were the comments given in a consolidated manner?	Yes	No	Response rate	
Synthesis team	33%	67%	70%	
Q11 - Are you satisfied with the final evaluation output?	Yes	No	Response rate	
Synthesis team	100%	0%	100%	
Q12 - Were you directly involved in the planning/conduct of the dissemination efforts?	Yes	No	Response rate	
Synthesis team	25%	75%	100%	
(n=4)				