

Comments received by 18:00 on 20 November 2009

**Australia:**

No substantive comments from us.

Virginia

**Denmark:**

- I think the TOR are generally well written and should provide a good framework for the individual country case studies. Major challenges are linked, I think, to the methodology and how, in particular, to measure e.g. whether effectiveness of aid has improved (core question no. 2) + whether the implementation of PD has strengthened the contribution of aid to development results (core question no. 3).
- The mix of different approaches seems a useful way to go about, but I believe getting quantitative data which can shed light on questions 2 and 3 will be difficult in many countries. Links to the monitoring survey and other ongoing exercises to be explored?
- The other major challenge will of course relate to making sure the core team gets comparable data from the various countries. Again this is difficult to ensure, but I believe the early contracting of the core team and their possibilities to assist the national coordinators should be essential in this respect.
- On the management side: Not sure it is practical for the national reference groups to be involved in the selection and contracting of consultants? (section 24, c).
- On p. 11 under iii (country evaluation report....), you might want to insert a deadline (tentative).

**Specific comments from the Danish Embassy, Mali**

- As for the TOR for country evaluations it is desirable to highlight the aspect of utilisation of country systems more sharply – e.g. at article 14 (page 5) to add under bullet “Implementation or process – assessing changes of behaviour of countries and donors” a specific point to “analyse if national systems, joint financial arrangements and sector budget support are increasingly used and the eventual obstacles and barriers in this respect (including an analysis of the quality of national systems).
- As for article 22 h. (page 9): It is not clear what is meant by this point, should be re-worded!
- It appears that it is not specifically/explicitly mentioned that the concrete indicators associated with the PD + AAA should be at the very heart of the evaluation.

Hope the regional workshops have been productive and look forward to seeing you in Paris end of November- early December!

Margrethe Holm Andersen

## German Comments on the Draft Generic ToRs for Country Evaluations (Version Oct. 20, 2009) 28 October 2009

**Note:** These comments focus on the **substance** of the Country Level Studies (ToRs pages

1-9, Appendices A, B & G) which seems to be the most important issue to be addressed at this stage.

### 1. Overall orientation of PD Evaluation Phase 2, including the core evaluation questions

PD Evaluation Phase 1 focused on implementation, i.e. inputs, process and outputs. Phase 2 will now focus on **results (outcomes and impacts) in terms of aid and development effectiveness**. The conceptual clarity and strategic orientation of the *Linkages Study* and the Approach Paper with regard to the “impact chains” (inputs => outputs => outcomes => impacts), which had been visualized in two framework diagrams, could be better reflected in the Draft Generic ToRs.

- ⇒ We suggest to **visualize these impact chains** in the ToRs, attaching both diagrams elaborated during Phase 1 since they complement each other and can also be of didactic use during the regional workshops.<sup>1</sup> This would also help stressing the importance of the **five PD principles** around which “PD interventions” are normally grouped and discussed (ownership-alignment-harmonization-management for results-mutual accountability, see first diagram) and which should be updated according to the **Accra Agenda for Action (AAA)**. Also, the Synthesis Report pointed at **potential trade-offs** between these principles which seems to be a topic worth exploring more in Phase 2 (as mentioned briefly in Appendix B).

The **core evaluation questions** (p. 7) have been shortened and rephrased considerably in comparison to the Approach Paper. In part, they now follow the **impact chain** (Question 1: context; Question 2: process & intermediate outcomes; Question 3: development outcomes). However, they are now mainly put as “**yes/no questions**”, although the sub-questions in Appendix B point at the necessary complexity of the assessments.

- ⇒ Since this phrasing could limit the depth and quality of analysis and presentation of results, it should be reconsidered. **Questions 2 and 3** should rather be phrased “To what extent, and how has ...”. **Question 4** could return to the more encompassing phrasing in the Approach Paper and directly address potential alternatives to the PD.

In terms of substance, **Question 1** focusing on what is now called the “aid arena” seems to include both initial **country conditions** (which could be used for a comparative analysis of country study results) and “**PD inputs**” produced by countries and donors together (“PD configurations” according to the Approach Paper). The latter – as in Phase 1 – will need to be described in some detail and analyzed according to various criteria (adaptation to country circumstances, priorities among the five PD principles, changes after the AAA etc.).

---

<sup>1</sup> See attachments: Phase 1 Framework ToRs 4/2007:57 with diagram from Booth & Evans 2006 and Approach Paper 5/2009:8.

- ⇒ It might be useful to **separate** the issues dealt with in **Question 1** into **two** core evaluation questions on **country context** (questions 1a, aspects of 1c,e) and **strategies of PD implementation** (questions 1b,d,f). Then, what is now Question 2 would also no longer need to focus on “process” since this would be covered under the second aspect of Question 1.

⇒

## 2. Draft Evaluation Matrix (Appendix B)

**Note:** comments at this point are mainly on the sub-questions, less on indicators, sources and methods. A general suggestion would be to better specify and make use of existing sources of data, like on-going monitoring processes in the context of the EU Code of Conduct on Division of Labour and Complementarity or the Monterrey Consensus.

### Questions related to the country context:

- **Question 1a:** relevant issues also to be considered are the **history of aid** in the particular country, the country’s current state of “**aid dependence**” (comparing different indicators, e.g. ODA or CPA (Country Programmable Aid)/GNI, ODA or CPA/capita, percentage of ODA or CPA in total national budget etc.) and if it falls into the classifications of “**aid darlings**” or “**aid orphans**”.
- **Question 1c:** pay attention to “**new donors**” and funding mechanisms
- **Question 1e:** pay attention to issues like the emerging **climate change regime** (new funding opportunities)

### Questions related to PD country-level implementation (inputs):

- **Question 1b: development cooperation is more than development finance.** The evaluation needs to trace changes in the overall aid system, including other aid modalities like technical cooperation.
- **Question 1d** is interesting but more descriptive and should be the **first** in this block.
- **Question 1f** is important especially in “Phase 2 only countries” (see below under point 4), should also include the implementation of AAA commitments and be described and analyzed in quite some detail with regard to the five PD principles and their subcomponents (e.g. division of labour among donors in the case of harmonization strategies). A key issue will be to trace if different degrees of progress in the five PD principles (e.g. much harmonization but little ownership and alignment) have consequences for overall progress in aid and development effectiveness.

### Question 2 (intermediate outcomes):

- The proposal to “test against the **original expected aid effectiveness outcomes**” (eleven commitments in the “Statement of Resolve” of the PD) is new and interesting but methodologically not very well developed yet (the Approach Paper only lists the 21 “indicative propositions” of the Linkages Study in Appendix 2; it remains unclear if these have been dropped completely). However, they should not only be “clustered by the major AAA themes” but need to be **expanded according to the new and/or stronger AAA commitments**. With regard to this assessment (not really a test!), given the ongoing methodological problems of the PD indicators and monitoring

surveys, the PD evaluation should not strive for defining new indicators. The “third-best solution” seems to be the only feasible option.

- **Question 2a:** important issue; good initiative to design a specific survey tool (no comments on Appendix G which covers all relevant aspects).
- **Question 2b:** should be integrated into the overall assessment against PD and AAA commitments, since several address aspects of country capacity.

#### **Question 3 (development outcomes):**

- Again, it would be helpful not to phrase these questions as “yes/no” questions (rather: “**To what degree and how ...**”)
- “Implementation of the PD” is also a **matter of degree** (different progress with regard to the five principles etc.) – how can this more detailed analysis of impact chains be ensured?
- Overall, this block will depend less on individual evaluation questions and much more on solid contextual analysis of **plausible** connections.

#### **Question 4 (alternative approaches):**

- The phrase “is the PD the best way” is **misleading** since it seems to focus the evaluation on the **declaration** itself and not the considerable **structural and procedural changes** that have happened in the international aid system as a consequence. On the other hand, it is an interesting question if the international community has other instruments than a declaration for this purpose (e.g. stronger instruments like conventions or codes of conduct), but these are so far not addressed in the questions.
- In **questions 4a and b**, application of the PD suddenly seems to get reduced to certain new “modes of delivery” (like programme based approaches, budget support etc.). While the questions are relevant, they do not seem sufficient under this heading (**assessing “alternative approaches” is more than a comparison to “business as usual”**).
- **Question c:** this question seems to be less on “added value” of the PD than on the (very relevant) issue of the **relative contribution** the PD implementation can make to international development, given its limited adherence among other increasingly important actors and other important trends (e.g. economic & financial crisis).
- As stated above (under 2.), the **two additional questions in the Approach Paper** (Other strategies to achieve the same results? Same effects without the PD?) are also relevant in this block.

### **3. Different Treatment of Phase 1 and Phase 2 Countries**

**Phase 1** covered nine countries (including Zambia), **Phase 2** will cover 22 countries (eight repetitions, excluding the Philippines).

- ⇒ The ToRs should at some point make reference to the difference between “**Phase 1 countries**” (which need only an update on PD implementation and can also track follow-up on recommendations) vs. “**Phase 2 only countries**” (which need to describe and analyze in more details also the “PD inputs”).

**Samoa:**

I look forward however to the finalizations of the TOR for the evaluations which we did not have difficulties with in the first place.

Noumea Simi

**Sweden:****Comments on ToR's from the Swedish Reference Group for the Evaluation on the Paris Declaration**

Sweden has established a reference group for the second phase of the evaluation of the Paris Declaration. The group is chaired by the Secretariat for Evaluation at Sida and has members from SADEV, Sida and the Ministry of Foreign Affairs. It has agreed on the following comments on the Generic Terms of Reference for Country Evaluations and the Generic Terms of Reference for Donor/Agency HQ Evaluations.

Terms of Reference for Country Evaluations

A general observation is that the emphasis put on aid effectiveness and development effectiveness does not capture the fact that the PD is also about changing relationships, which is both of political and moral value. It is important that the evaluation does not lose sight of these "softer" values of the PD.

Generally, the ToR would be clearer if it would explicitly state that the evaluation will a) evaluate to what extent the PD has been implemented, and b) *in as far* it has been implemented evaluate what the results have been in terms of aid effectiveness and development. The main reason why this has to be very explicit is that the PD principles in many cases have not been implemented or only partially been implemented. In such cases question 2, 3 and 4 in the ToR may become redundant.

In cases when the PD has not been implemented question 1 could be elaborated in more detail. Issues around commitment, incentives and capacity among different stakeholders in the aid context (not only among donors) could give some explanations why the PD principles have not been implemented. Power analysis could be a useful tool to dig deeper into that issue.

Regarding question 1 power analysis could generally serve as a good methodology to analyse factors that work against the implementation of the PD. In terms of context, it would also be useful to look into how political factors in donor countries affect the aid arena in the partner country. For example, political changes or financial problems in donor countries will have repercussions on the aid context in partner countries.

Regarding question 3 human rights is not explicitly addressed in any of the three sub questions. Question 2 b touches upon this, including the dimension of social exclusion and gender equality, but it does not address the general issue on whether implementation of PD principles has led to a better HR-situation in the country.

The question on whether the implementation of the PD has led to changes for the poorest people is an important issue which might not be addressed properly if it is embedded as a component in 3b. Since this is a contested issue it would be worthwhile considering a supplementary study to elaborate on this theme.

As regards the tracer sectors the education sector may be a better choice than health. The reason is that aid to education generally is more along the principles in the PD than health, where other aid modalities persist alongside aid channeled through country systems.

#### **Switzerland:**

- We find it essential that the evaluation move significantly beyond the voices of the government ministries and ensure that the perspectives of civil society, of Parliament and of decentralized government entities be adequately heard and analysed.
  
- A value-added of the second phase is moving beyond aid effectiveness toward assessing the PD contribution towards development effectiveness. This is a tall order that absorbs resources and runs into tremendous attribution / contribution issues. In order to ensure robust results, we feel that the emphasis in this regard should be on the “tracer studies” in a very limited number of sectors (e.g., health). Better to do less in terms of broad coverage but more in terms of depth in a limited number of areas.

Anne Bichsel

#### **UK**

#### **UK Reference Group Comments on PD2 Draft Generic ToRs for Country Evaluations**

##### **Key Recommendations**

- Ownership and mutual accountability merit a higher and more explicit profile in the evaluation matrix to ensure balance between political and technical elements of Paris.
- Additional guidance needed to support standard approaches e.g. on data handling and analytical steps for each stage; counter-factuals and attribution; ensuring appropriate balance of quantitative and qualitative analysis.
- Clarification of expert evaluation support available to teams, quality assurance, and accountabilities for sign-off and dispute, along with clear explanation of links between national and international governance arrangements for the study.
- The national communications plan should be directly linked to key points in the national and international dialogue on aid effectiveness and MDG trends over the coming two years to build policy engagement with the study and ensure its timely contribution to the debates. Would the Management Group consider presentation of preliminary findings to the MDG stocktaking summit in late 2010?
- Clarity on the final number of country studies and budgets needed.

#### **Overall**

##### **Aid effectiveness is a political and not a technical agenda.**

1. The Paris Declaration ‘agenda’ is arguably more concerned with changing the wider partnership between countries and the associated ‘rules of the game’ than with the technicalities of aid management efficiency. However, key

principles such as ownership, mutual accountability and results are often crowded out by the focus on more readily measurable indicators of harmonisation and alignment as reflected in the Paris Monitoring Framework.

2. This study is an important opportunity to rebalance the evidence on the changing dynamics of country leadership, aid in relation to countries' development outcomes, and changes in international accountability and power relations. While implicit in parts of the evaluation matrix, **ownership and mutual accountability merit a higher and more explicit profile**. Without this there's a risk that the evidence collected will overly focus on the narrow technical interpretation of Paris. This would limit the relevance of the study to important international discussions through 2010 and 2011 on the future 'international social contract' in which aid plays an important part.

### **Expectations and deliverables**

3. Expectation of this complex study are high and set to increase with the approaching MDG Stocktake in late 2010 and the 2011 HLF4. In the current global political and economic context, donors are under increasing pressure to demonstrate results and this study will form a vital part of the evidence base informing the debate on the impact of, and most effective arrangements for managing ODA. **Would the Management Group consider presentation of preliminary findings to the MDG stocktaking summit in late 2010?**
4. The TORs recognise the challenges facing the study, but more detail would be helpful on how these challenges will be handled, especially balancing the focus and relationships between **intermediate and development outcomes, handling of different aid instruments, counter-factuals and attribution** for the PD's proposed linkage between aid and development effectiveness. Further comments on the methodological issues are provided below.

### **Standardised methodology and guidance required**

5. The TOR clearly recognise the need for as robust an international evidence base as possible requiring a sufficiently standardised approach across countries, which we support. However, there is a tension between this analytical need and diverse country prioritisation of the various Paris elements, often clearly articulated in country level partnership arrangements such as PAF, JAFs, and Aid Management Policies. Further explanation of how the study will manage the **balance between analytical comparability and the diverse context** will be helpful.
6. The challenges and complexity of this study can't be underestimated and do argue for a fairly standardised approach to analysis across countries, the case for which has been supported in the regional workshops. We agree with this, and propose **further Guidance be provided to country evaluation teams and their stakeholders** to help make this happen. Although this may well be planned for 'internal' orientation with actual teams when appointed, such guidance at this stage would assist a shared understanding among all stakeholders and particularly those responsible for the recruitment of evaluation teams, of the **skills and capacities needed and further support requirements**.

7. The proposed Guidance would help develop the current draft Evaluation Matrix, and particularly the Methods Column 4 from its current listing of possible methods and tools, to a **more precise identification of the analytical methods for each study element** and ensure clear understanding of all the steps involved.
8. The Guidance would provide useful further support for country evaluation teams, the national coordinator and reference groups in establishing clear expectations of the scope, approach and standards required. Such a shared understanding will be critical to their success and credibility.

### **Governance and Management**

9. The TORs refer early on to the Paris Declaration's embedded commitment to this evaluation. There is little further reference to this nor the institutional architecture around Paris. The TOR rightly proposes a national **communications plan**. We suggest this is **directly linked to key points in the national and international dialogue on aid effectiveness and MDG trends over the coming two years** to build the wider policy stakeholder engagement with the study and ensure its full and timely contribution to the debate.
10. The TORs make little reference to the Paris Monitoring Framework and its associated country and international institutional arrangements. Further reference in the TOR of **how this study will build on monitoring arrangements** would be helpful both in communicating this study with policy stakeholders, and in minimising possible duplications and data handling issues. The wider contribution of the studies to country and DAC level aid partnership discussions would also merit more explicit reference particularly to ensure **timely use of findings to inform key decision points up to HLF4 at country and international levels**.
11. The TOR would benefit from further explanation of the governance arrangements including most importantly the **precise accountabilities** between the team, coordinator and reference groups, and the arrangements for **quality assurance, sign-off responsibilities and handling in case of dispute**.
12. Explanation of how the country studies management arrangements connect with the international management and wider governance arrangements would be useful, either in the TOR's Responsibilities and Accountabilities section or in an annex.
13. The challenges this study faces are already formidable and risk being increased by what appears to be uncertainty on the **budget situation** and even in the number of country studies expected.

**Further details and specific recommendations on some of these points are provided below.**

### **Management and Governance Arrangements**

1. We **propose additional guidance** is attached to the TOR covering e.g. interpretation and specification of the evaluation questions, measurement and

data sources of the evaluation questions; construction of data sets across aid modalities; handling of counter-factual, attribution and plausible contribution; and the application of analytical techniques to the various data sets.

2. Para 26a leaves the question of budgets rather open to question - **clarification on budgets & funds** planned and available for an agreed set of country studies and associated groups such as the NRAGs would be helpful and essential for the governance of the overall study.
3. Accountabilities at all levels are going to be complex and will need careful management. These **could be made clearer and more comprehensive in the next draft. Para 29iii** for instance is quite vague on external quality assurance and sign-off responsibilities of e.g. Inception, Draft and Final Reports, and greater clarity on the final accountability point in case of dispute would be useful. **Para 23** should explicitly include quality assurance as a key management consideration and the arrangements for it be more clearly set out.
4. The role of the National Coordinator (NEC) is very significant both with respect to overall process and quality. **Could more clarity be given on relationships and accountabilities between NEC, NRAG, Country and Core Teams, and the links between these arrangements at national and international levels.** E.g. Is the membership of the NRAG mirrored in that of the IRG? Where is the **final point of accountability** in case of dispute? What is the role of the DAC/WPEFF in the overall study and the response to it?
5. The composition, application and responsibilities of the **external Peer Reviewers** would also be helpfully explained and consideration of further **expert support** through this mechanism could be useful, given the stated need for international comparability.
6. Given the large numbers of donors with interest in this study, will the **IRG have the opportunity to comment** at interim stages not currently written into the Schedule e.g. at **Inception Report Stage or draft report stage.**

## Methods

7. Given the highly diverse country contexts for, interpretations of and approaches to implementing PD, more explicit explanation of how these will **actually be factored into the analysis** would help understanding of how the overall international evidence base will be constructed.
8. The core questions while generally clear, generate high expectations for the study's ability to provide categorical answers e.g. on whether we did as we undertook to, what effects has it had and to what extent the investment has contributed to better and faster development results.
9. We suggest adding into the relevant core questions **'why and why not' type** questions and giving more **detail of the comparative analysis** considered possible between sectors, aid instruments, donors, national and local, beneficiaries, and ultimately through the Synthesis, between countries.
10. We would like the TORs to be clearer on the difficult issues e.g. **handling of the baseline issue, attribution, counter-factuals and 'plausible contribution' and what the theory-based approach will mean for the studies in practice and envisaged type of results.** In the absence of neatly defined counter-factuals, could a set of alternative scenarios that will be tested through the study be used?
11. Having opted for standardised approaches the current draft stops short of setting out how this will be achieved. For instance it's not clear from the TORs whether common indicators will be applied to the 'expected outcomes' or if a common framework for structured and semi-structured surveys/questionnaires will be provided. **While lists of possible indicators, data, methods and analysis are provided, their actual handling, selection and application remains unclear.**

## Evaluation Matrix

12. The 11 expected outcomes (appendix B evaluation matrix) do not include anything specific on **ownership and partnerships** which para 8 says the study should pay particular attention to. While these may be implied in some of the other outcomes, **more explicit reference merited**.
13. **MDG** results could be mentioned in as evidence sources in section 1a, as well as at development outcome and indicator levels. **MDG's merit more mention throughout TORs**.
14. *Process and intermediate outcomes*
  - While the third option of key informant responses on change trajectories against each outcome may prove most practical, **further exploration of the first option to agree on selected key indicators** for each outcome would strengthen the cross-sector and cross-country comparability.
  - The use of a survey to gather information on less data-rich outcomes sensible and should definitely be complemented by key informant interviews and qualitative analysis.
  - **Item iv** in the second set of issues could be more tightly **specified** to focus on global programmes and initiatives' **use of country budgetary, implementation and monitoring systems** rather than the vaguer suggested 'integration into..agenda'.
  - **Mutual Accountability**. The issue of asymmetric power relationships between donors and partner countries and the extent to which under PD implementation these have changed and with what effect, is **not explicit enough** given its prominence in the PD/AAA. A useful sub-question could be framed around "*how far donors and partner countries have become more accountable to each other.*"
  - **Transaction costs** (Question 2a). That views range widely on this variously understood key aspect of the Paris agenda is well known. We welcome the note under methods column to '*bear in mind benefits as well as costs*' but are **not convinced that the instrument presented in Appx G will adequately assess** this area. The definition ignores benefits, the comparator is unclear, the reference period too short and the questionnaire is rather simplistic. The questionnaire as proposed (Q2) omits the crucial variable – ownership, around which much of the added transaction costs typically associated with PD are invested. **A revised questionnaire could be a first step** for focusing the search for more robust evidence but more detail needed to breakdown types of costs.
15. *Development outcomes*
  - Although the following development outcomes section of the matrix covers impact on **gender and excluded groups**, the intermediate outcome would also benefit from more explicit examination of how far arrangements for the planning, management and delivery of aid have targeted development outcomes, including gender equality. Entry points for this could be through the proposed additional guidance e.g. definitions/criteria for "stronger national strategies and frameworks" referred to in the table, which could include **analysis of voice and representation in strategy definition** and disaggregated monitoring.
  - In sources of data use of **sex disaggregated data** should be included where appropriate.
16. *Paris the best approach?*
  - While the sub questions are largely highly relevant to what is in many ways the key and most challenging core question, the matrix provides

**very limited information on methods and analysis** for this section of the study.

- Not sure 4b is appropriately specified – while budget support is typically associated with it, the PD itself is deliberately silent with respect to budget support.
- More broadly posed the predictive question is welcome if challenging methodologically, and addresses a **key issue of how relevant the Paris approach is to addressing emerging global issues such as climate change**. In particular how can the Paris Principles to be applied to the climate investment funds that are already established and those likely to be developed post COP-15 in Copenhagen.

#### **UNDP:**

Thank you very much for sharing the draft document with all of us. I am sure there will be a lot of input from the regional workshops and that will be very important as the countries are the ones conducting the evaluation. So I will just keep my comments to the minimum.

Since this is an evaluation as opposed to the formative evaluative study we conducted last year, I feel the evaluation criteria and corresponding questions should be much clearer and more searching. Moreover the management structure for all the country and donor evaluations should address issues of independence and credibility.

With regard to the criteria, I do not think fundamental issues of relevance are addressed. The link between aid and development effectiveness should be explored in the country evaluations from a national standpoint of how priorities were set; how the need for aid was established and how it was used.

In order to have a useful country-specific it is insufficient to establish the context of aid. Other issues of development context need to be addressed. More and more developing countries are in a position to fund the bulk of their development. How has this changed the perception and use of aid?

There need to be many more specific questions on donor behavior and practice.

In this connection the content of the Paris Declaration should be placed under rigorous scrutiny. Throughout the document there appears to be an assumption that the PD is given, the indicators and targets are useful, and that the direction is clear. With a full-fledged evaluation there is the opportunity to ask whether the PD itself addressed the right issues in order for aid effectiveness to contribute to development effectiveness.

On the management arrangements, I do not believe the earlier arrangement of a single National Evaluation Coordinator who is responsible for design, contracting and quality assurance of the evaluation is sufficient either for the country evaluations or for the donors. The reference group would only be sufficient for validation but not quality assurance. In short I do not think this is a credible arrangement.

Even for the UN Delivering As One Pilot country led evaluations which are now underway, the countries have agreed on a much more structured, representative and independent evaluation management group, a consistent and independent quality assurance process covering the TOR, inception report and the draft report; and an approach to disclosure which meets international standards.

Saraswathi Menon.