

# **Paris Declaration Country Evaluations: How Solid is the Evidence?**

META-Evaluation of the Country Evaluations of the Phase II Paris Declaration Evaluation

Krizelda Marie **Songco**  
Nathalie **Holvoet**  
Liesbeth **Inberg**

Working Papers are published under the responsibility of the IOB Thematic Groups, without external review process. This paper has been vetted by Nadia Molenaers, convenor of the Thematic Group Aid Policy.

Comments on this Working Paper are invited. Please contact the author at [nathalie.holvoet@ua.ac.be](mailto:nathalie.holvoet@ua.ac.be), [liesbeth.inberg@ua.ac.be](mailto:liesbeth.inberg@ua.ac.be) or [klsongco@yahoo.com](mailto:klsongco@yahoo.com).

Instituut voor Ontwikkelingsbeleid en -Beheer  
Institute of Development Policy and Management  
Institut de Politique et de Gestion du Développement  
Instituto de Política y Gestión del Desarrollo

Postal address:	Visiting address:
Prinsstraat 13	Lange Sint-Annastraat 7
B-2000 Antwerpen	B-2000 Antwerpen
Belgium	Belgium

Tel: +32 (0)3 265 57 70  
Fax: +32 (0)3 265 57 71  
e-mail: [iob@ua.ac.be](mailto:iob@ua.ac.be)  
**<http://www.ua.ac.be/iob>**

**WORKING PAPER / 2012.01**

# **Paris Declaration Country Evaluations: How Solid is the Evidence?**

META-Evaluation of the Country  
Evaluations of the Phase II Paris  
Declaration Evaluation

Krizelda Marie **Songco**

Nathalie **Holvoet**

Liesbeth **Inberg**

January 2012



## TABLE OF CONTENTS

<b>ABBREVIATIONS</b>	<b>6</b>
<b>ABSTRACT</b>	<b>7</b>
<b>1. INTRODUCTION</b>	<b>8</b>
<b>2. BACKGROUND ON METHODOLOGY AND RESEARCH QUESTIONS</b>	<b>12</b>
2.1. META-EVALUATION	12
2.2. OECD/DAC EVALUATION QUALITY STANDARDS	14
2.3. EVALUATION FRAMEWORK OF THE PARIS DECLARATION EVALUATION PHASE TWO	16
2.4. SELECTION OF META-EVALUATION QUESTIONS, STANDARDS, RATING SYSTEM AND SAMPLE.	18
<b>3. HOW WELL DO THE COUNTRY EVALUATION REPORTS PASS THE TEST OF THE OECD/DAC EVALUATION QUALITY STANDARDS?</b>	<b>20</b>
3.1. EVALUATION METHODOLOGY	21
3.2. INFORMATION SOURCES	22
3.3. INDEPENDENCE	22
3.4. QUALITY ASSURANCE	23
3.5. RELEVANCE OF THE EVALUATION RESULTS	24
3.6. COMPLETENESS	25
3.7. STRENGTHS AND WEAKNESSES	26
<b>4. HOW WELL DO THE COUNTRY EVALUATION REPORTS COMPLY WITH THE EVALUATION FRAMEWORK?</b>	<b>28</b>
4.1. PARIS DECLARATION IN CONTEXT	30
4.2. PROCESS AND INTERMEDIATE OUTCOMES	30
4.3. DEVELOPMENT OUTCOMES	32
4.4. FRAMEWORK FOR OVERALL CONCLUSIONS	33
4.5. STRENGTHS AND WEAKNESSES	33
<b>5. CONCLUSIONS AND RECOMMENDATIONS</b>	<b>37</b>
5.1. CONCLUSIONS	37
5.2. IMPLICATIONS FOR FUTURE SIMILAR EVALUATIONS	38
5.3. POSSIBLE ROUTES FOR FURTHER RELATED RESEARCH	39
<b>REFERENCES</b>	<b>40</b>
<b>ANNEX I: PARIS DECLARATION INDICATORS: BASELINES, TARGETS AND STATUS</b>	<b>42</b>
<b>ANNEX II: OECD/DAC EVALUATION QUALITY STANDARDS</b>	<b>43</b>
<b>ANNEX III: TALLIED SCORES OF THE 15 COUNTRY EVALUATION REPORTS ON THEIR COMPLIANCE WITH THE EVALUATION FRAMEWORK (AGGREGATED BY SUB-QUESTIONS)</b>	<b>46</b>
<b>ANNEX IV: STRENGTHS OF EVIDENCE FOR INTENDED OUTCOMES UNDER CORE QUESTION TWO.</b>	<b>48</b>

**ABBREVIATIONS**

CPIA	Country Policy and Institutional Assessment
MDG	Millennium Development Goal
OECD/DAC	Organisation of Economic Cooperation and Development – Development Assistance Committee
PIU	Parallel Implementation Unit
PFM	Public Financial Management

## ABSTRACT

The evaluation of the Paris Declaration (PD) is one of the most important and challenging evaluative undertakings of the past decade in the aid sector. The PD evaluation commissioned by the OECD/DAC Evaluation Network consists of a set of independent cross-country and donor evaluations which were carried out in two phases. The scope and importance of this evaluation makes it a particularly suitable subject for a meta-evaluation. Our 'evaluation of the evaluation' complements the official meta-evaluation of the synthesis report in that it assesses all country evaluation reports available in English (15 out of 21 reports) using the OECD/DAC Evaluation Quality Standards. Two research questions are central in our undertaking: Is the quality of the country evaluation reports good enough to be included in the synthesis report? Do the reports properly comply with the evaluation framework to permit comparison of evaluation across countries? The findings of the meta-evaluation demonstrate that comparability of country evaluation reports is satisfactory. The quality of evidence, however, is questionable, due to various limitations and constraints that plagued several country evaluations. Therefore, the inclusion of some of the country reports in the evaluation synthesis report is questionable.

## 1. INTRODUCTION

In the history of international cooperation the 2000 Millennium Development Goals (MDGs) and the 2002 Monterrey Consensus are both important agreements which concretise the recognition of the need for increased efforts in battling poverty and for better cooperation between developed and developing countries. In 2005 the Paris Declaration for aid effectiveness further pushed the international community to commit more concretely on pursuing better delivery and management of aid, thus bearing more impact on sustainable development (Canlas et al., 2008). The Paris Declaration, which is endorsed by 52 donor and partner countries and 30 multilateral agencies and non-governmental organisations (IODPARC, 2009), aims to “strengthen ‘partnerships’ between donor countries and countries receiving aid in order to make aid more effective and to maximise development results” (IODPARC, 2009: 3). The Paris Declaration contains 56 partnership commitments aggregated into five core principles, defined as:

‘Ownership`	Developing countries set their own strategies for poverty reduction, improve their institutions and tackle corruption.
‘Alignment`	Donor countries align behind these objectives and use local systems.
‘Harmonisation`	Donor countries coordinate, simplify procedures and share in formation to avoid duplication.
‘Managing for Results`	Developing countries and donors shift focus to development results and results get measured.
‘Mutual Accountability`	Donors and partners are accountable for development result (OECD/DAC, 2005).

During the third High-Level Forum in Accra (Ghana) in 2008 donors and partner countries took stock of their progress in implementing the Paris Declaration commitments and reinforced the agenda to accelerate the advancement towards the Paris targets. In the Accra Agenda for Action three main areas for improvement are proposed:

‘Ownership`	Countries have more say over their development processes through wider participation in development policy formulation, stronger leadership on aid co-ordination and more use of country systems for aid delivery.
‘Inclusive partnerships`	All partners - including donors in the Organisation for Economic Cooperation and Development/ Development Assistance Committee (OECD/DAC) and developing countries, as well as other donors, foundations and civil society - participate fully.
‘Delivering results`	Aid is focused on real and measurable impact on development (3rd High Level Forum on Aid Effectiveness, 2008)

During the fourth High-Level Forum in Korea (Busan) in November 2011, participants, including ministers of developed and developing nations, emerging economies, providers of



South-South and triangular co-operation and civil society, signed the Busan Partnership for Effective Development Cooperation. While recognising the diversity of the different signatories, four common principles which form the foundation of effective development co-operation are formulated in the document: 'ownership of development priorities by developing countries', 'focus on results', 'inclusive development partnerships' and 'transparency and accountability to each other' (4th High Level Forum on Aid Effectiveness, 2011).

To ensure the achievement of the targets established in the Paris Declaration and to determine the impact of the pursuit of aid effectiveness on development, donors and partner countries committed themselves to monitor, evaluate and document progress on implementing the Paris Declaration (OECD/DAC, 2005). To pursue this commitment, a set of independent cross-country and donor evaluations were carried out in two phases, in 2008 and in 2010, coordinated by an International Reference Group and supported by a Core Evaluation Team commissioned by the OECD/ DAC Evaluation Network (OECD/DAC Evalnet) (IODPARC, 2009). In addition to the evaluations, three rounds of monitoring surveys (2006, 2008 and 2011) were to track progress on 12 indicators and targets set for 2010 (see annex I for an overview of the indicators, their baselines, targets and status). While progress has been made in most of the indicators between 2005 and 2010, only one indicator (4: Strengthen capacity by co-ordinated support) has met its target (OECD, 2011).

The evaluation of the Paris Declaration is the largest joint evaluation ever undertaken. The first phase of the evaluation took place from March 2007 to September 2008 and examined the start-up and implementation of the Paris Declaration from March 2005 to late 2007 (Wood et al., 2008). It aimed at answering the 'how' and 'why' questions, providing emphasis on the input and output levels, to capture the "incremental and incidental behaviour changes" (IODPARC, 2009: 3) related to the implementation of the Paris Declaration. The evaluation consists of 8 country assessments and 11 donor studies at the headquarters level (voluntary participation) and was designed to draw practical lessons on improving implementation of the five principles and to track progress of the implementation to serve as an input to the 3rd High-Level Forum on Aid Effectiveness in Accra (2008). The evaluation concludes that the participating countries have the enabling conditions in place (commitment, leadership, capacities and incentives), with varying magnitude, yet experience great difficulties in implementing the Paris Declaration principles. Carrying out the Paris Declaration was emphasised to be a highly political undertaking, dealing with power and political economy struggles which require political solutions for both donors and partner countries. Donors, for example, need to accept and manage the risks involved in aligning with partner countries' systems and harmonising with other donors instead of imposing their own systems (Wood et al., 2008). The evaluation presents the extremely varied performance in implementing the Paris Declaration, but also recognises that aid effectiveness is only a small part of the development agenda within governments of the partner countries. Hence, the Paris Declaration serves a variety of purposes for different countries as well (Wood et al., 2008).

The second phase of the evaluation was used as an important input into the 4th High Level Forum in Busan (2011). This evaluation synthesizes the results from two phases of voluntary country evaluations (8 in phase one and 21 in phase two), donors/agency institutional studies (11 in phase one and 7 new and 7 updates in phase two) and five supplementary studies. It also aims to highlight policy-relevant findings, conclusions, possible lessons and recommendations for the key audience and expected users of the evaluation results (Wood et al.,

2011). The evaluation demonstrates that most progress has been made within the ‘ownership’ principle and least within the ‘managing for results’ and ‘mutual accountability’ principles. While the evaluation concludes that changes made in the context of the Paris Declaration did not lead to a reduction in the overall burdens of aid management, the Paris Declaration has contributed to a better quality of aid, to more transparent and effective partnerships, and to the rise of volumes of aid. Beyond the scope of the Paris Declaration (aid by non OECD/DAC donors), however, these improvements have not yet been made, which could decrease the benefits of the Paris Declaration (Wood et al., 2011). From this vantage point, it is positive that the recent Busan Partnership has not only been endorsed by traditional donors, but by a diverse set of other actors as well.

Monitoring and evaluation of the Paris Declaration, as recognised by the evaluation itself, is a challenging task, because its implementation involves a complex web of actors, intentions and contexts. Moreover, attribution to development results, say poverty reduction, is tricky as the efforts and processes involved in attaining such is not directly linked to the implementation of the Paris Declaration. The complexity of the evaluation inevitably raises questions on the validity of the evaluation findings. Guion (2002) associates the validity of a study to whether its findings are true and certain, meaning it accurately reflects the real situation and it is backed by sound evidence. Given the importance of the evaluation of the Paris Declaration, the management group of the second phase of the evaluation commissioned an independent assessment of the evaluation (Patton and Gornick, 2011), a so-called meta-evaluation. The meta-evaluation team not only assessed the synthesis evaluation report, but the evaluation process as well. To this end, the team reviewed data collection instruments, templates and processes and the partner country and donor evaluation reports on which the synthesis is based; observed two meetings of the International Reference Group; engaged participants of the International Reference Group in a reflective practice lessons learned session; surveyed participants in the evaluation process; and interviewed key people involved in the evaluation (Patton and Gornick, 2011). While the meta-evaluation team acknowledges some limitations inherent to a complex and comprehensive evaluation like this one, the team concludes that “the Final Synthesis Report can be trusted as independent, evidence-based, and adhering to international standards for quality evaluation” (Patton and Gornick, 2011: iv).

Complementary to the official meta-evaluation, which focuses mainly on the synthesis report, this study aims to specifically assess the country evaluation reports<sup>1</sup> for phase two of the Paris Declaration evaluation, using the OECD/DAC Evaluation Quality Standards. As crucial as the evaluation of the Paris Declaration is to the pursuit of aid effectiveness, its value and worth also highly depend on the quality of the country evaluation reports, as an evaluation synthesis is “only as good as the evaluation findings they synthesise” (Cooksy and Caricelli, 2005: 32). A meta-evaluation is normally conducted prior to the preparation of a synthesis report, using the findings of the meta-evaluation as basis for choosing which evaluation report to include in the synthesis report (Cooksy and Caricelli, 2005). However, as outsiders, we were only able to access the country evaluation reports after the release of the synthesis report and we thus analyse ex-post whether the country evaluation reports pass the test of the OECD/DAC Quality Standards to be included in the synthesis. Unlike the official meta-evaluation, we were only able to assess the final country evaluation reports and not the evaluation processes.

---

[1] 15 out of 21 country evaluation reports are included in the study as these were the only ones available in English during the time of the study.

The structure of this paper is as follows: section 2 of the study briefly introduces the methodology of meta-evaluation, the evaluation framework of the Phase Two Evaluation of the Paris Declaration, the research questions and specific methods used in this meta-evaluative exercise. Section 3 summarises the findings with respect to the application of the OECD/DAC Evaluation Quality Standards on each country evaluation report, while section 4 highlights the extent to which different country evaluation reports comply with the overarching evaluation framework. Section 5 lays out conclusions from our study and offers a few recommendations for further studies.

## 2. BACKGROUND ON METHODOLOGY AND RESEARCH QUESTIONS

This chapter introduces the methodology of meta-evaluation (2.1.) and discusses evaluation standards (2.2.). The evaluation framework of the second phase of the PD evaluation is presented in 2.3 while specific research questions and methods used in our meta-evaluation are discussed in 2.4.

### 2.1. Meta-evaluation

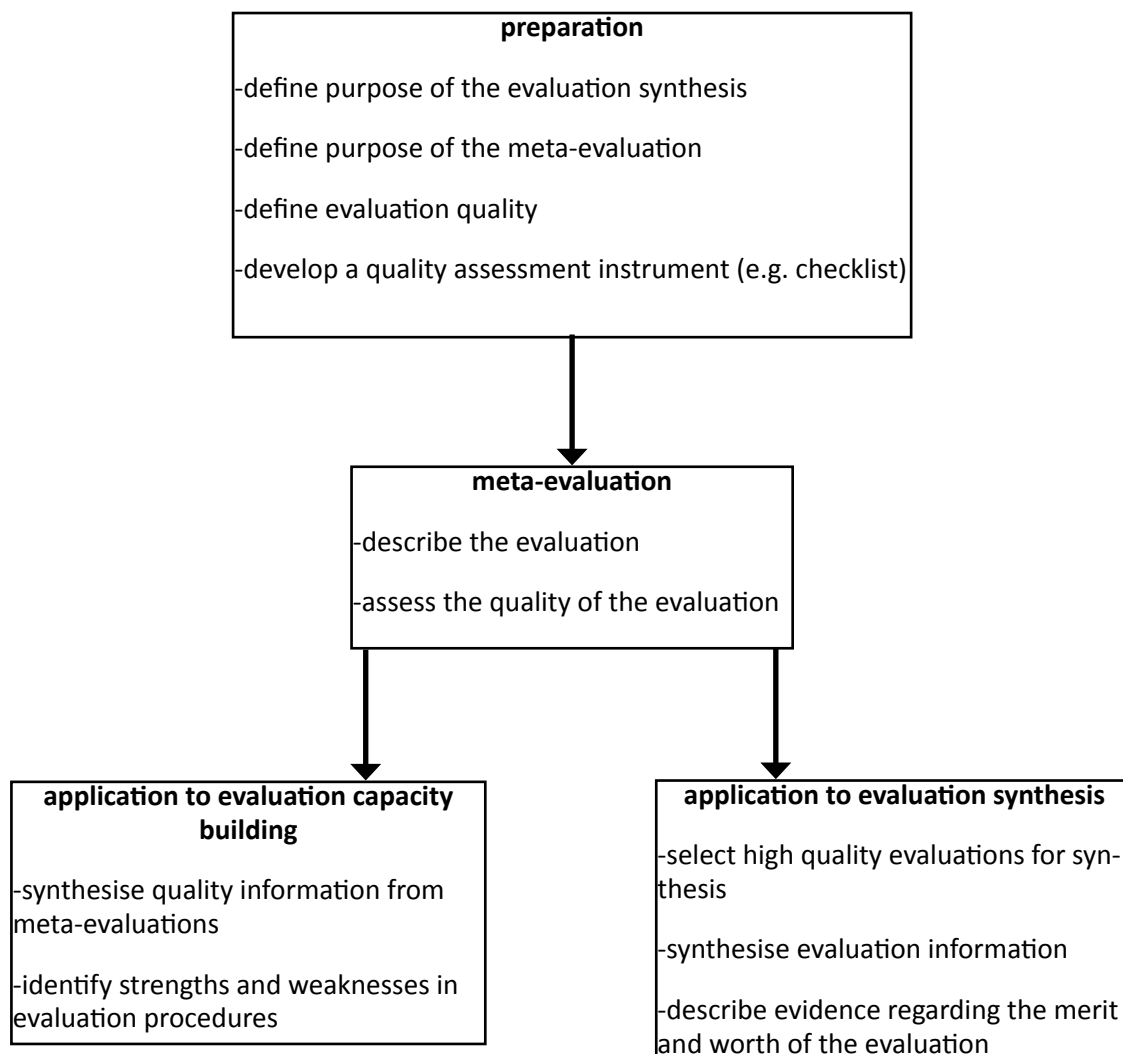
The concept of ‘meta-evaluation’ was popularised by Scriven (1969), who used the term for his evaluation of a plan for an evaluation of education products (Stufflebeam, 2001). He defines meta-evaluation as “any evaluation of an evaluation, evaluation system, or evaluation device” (Stufflebeam, 2001: 185) and strongly argues the need for such evaluations to screen out inaccurate or biased evaluation reports which are misleading for consumers. Stufflebeam (2001) further specifies the definition of meta-evaluation as “the process of delineating, obtaining, and applying descriptive information and judgmental information about an evaluation’s utility, feasibility, propriety, and accuracy and its systematic nature, competence, integrity/honesty, respectfulness, and social responsibility to guide the evaluation and publicly report its strengths and weaknesses” (Stufflebeam, 2001: 185). Since the launch of the term meta-evaluation by Scriven, the practice of meta-evaluation has slowly gained ground. The evaluation field has recognised the imperative for meta-evaluations as much as the need for evaluations.

Meta-evaluations can be either formative or summative. Formative meta-evaluations, as Stufflebeam (2001) defines it, are used by evaluators to assist them to “plan, conduct, improve, interpret, and report their evaluation studies”, while summative meta-evaluations are carried out to judge an evaluation’s merit and worth, and to identify its strengths and weaknesses (Stufflebeam, 2001: 183). Summative meta-evaluations are conducted after an evaluation whereas formative ones take place during an evaluation (Stufflebeam, 2001).

As mentioned in the introduction meta-evaluations are important for deciding which evaluations to include in an evaluation synthesis. The existence of an evaluation implicitly implies the demand for a meta-evaluation to determine its quality and validity, as evaluation reports feed into decisions which affect communities or even whole countries. As Stufflebeam (2001) puts it, “[m]etaevaluation...is a professional obligation of evaluators” (Stufflebeam, 2001: 183). Furthermore, identifying strengths and weaknesses of evaluations are also valuable to the evaluation field as it provides information for the continuous improvement of the evaluation practice. Stufflebeam (2001) believes that “meta-evaluations are in public, professional, and institutional interest” to guarantee the improvement of evaluation practice and the efficient and effective administration of evaluation systems (Stufflebeam, 2001: 183). Cooksy and Caracelli (2005) consider the improvement of the evaluation practice to be crucial for avoiding poor quality evaluations, which can result when evaluators are poorly trained and/or unfamiliar with criteria and standards of good evaluation. Evaluations of poor quality increase the possibility of the misuse of the evaluation reports and decrease the likelihood of attaining the desired impact of the evaluation to bring about behavioural changes in society (Cooksy and Caracelli, 2005; Alkin and Coyle, 1988; Datta, 2000; Hofstetter and Alkin, 2003; Stevens and Dial, 1994; Weiss, 1998; Henry and Mark, 2003).

The diagram shown in figure 2.1. presents Cooksy and Caracelli’s (2005) proposal of the steps and purposes of a meta-evaluation.

**Figure 2.1. Steps and purposes of meta-evaluation**



Source: Cooksy and Caracelli, 2005

Stufflebeam (2001) as well defined several steps which should be considered in the execution of a meta-evaluation:

1. Determine and arrange to interact with the meta-evaluation's stakeholders.
2. Staff the meta- evaluation team with one or more qualified evaluators.
3. Define the meta-evaluation questions.
4. Agree on standards, principles, and/or criteria to judge the evaluation system or particular evaluation.
5. Develop the memorandum of agreement or contract to govern the meta-evaluation.
6. Collect and review pertinent available information.
7. Collect new information as needed, including, for example, on-site interviews, observations, and surveys.
8. Analyse the qualitative and quantitative information.

9. Judge the evaluation's adherence to appropriate standards, principles, and/or criteria.
10. Convey the meta-evaluation findings through reports, correspondence, oral presentations, etc.
11. As needed and feasible, help the client and other stakeholders interpret and apply the findings.

Stufflebeam's list basically mirrors the list of Cooksy and Caracelli, but it further specifies steps requiring interaction with the stakeholders. Both Stufflebeam and Cooksy and Caracelli include the identification or development of standards or instruments to assess the quality of the evaluation (system) as one of the first steps. Cooksy and Caracelli put forward that "[c]reating a meaningful set of quality criteria that will provide a substantive and credible representation of quality requires that the criteria selected be tailored to the purpose of the meta-evaluation" (Cooksy and Caracelli, 2005: 35).

One of the important elements in a meta-evaluation are indeed the standards, principles or criteria used to judge the quality of particular evaluations or the entire evaluation system. One of the best known set of standards are the Programme Evaluation Standards of the American Joint Committee on Standards for Educational Evaluation<sup>2</sup> (Yarbrough et al., 2011), which were originally elaborated in 1994 and which have been revised recently. Over the years a number of standards have been elaborated with particular reference to the development evaluation field, including the UN Standards for Evaluation<sup>3</sup> (United Nations Evaluation Group, 2005), the evaluation checklist of the Evaluation Office of the Netherlands Ministry of Foreign Affairs<sup>4</sup> (IOB, s.a.) and the OECD/DAC Evaluation Standards (OECD/DAC, 2006; 2010). Different sets of standards differ with respect to the level of detail but they all somehow capture issues of validity, reliability and usefulness. As the Paris Declaration Evaluation was commissioned by the OECD/DAC Evaluation Network, we have selected the OECD/DAC evaluation quality standards as the guiding standards for our meta-evaluative exercise.

## 2.2. OECD/DAC Evaluation Quality Standards

The members of the OECD/DAC identified in 2006 a set of 'Evaluation Quality Standards' for a three-year test phase with the aim to establish a benchmark for quality evaluation process and products expected from DAC members, albeit not binding. The set of standards intends to serve as a guide to improved development intervention evaluations, which are crucial in determining the real value of development initiatives. Sound and accurate evaluation findings are critical for DAC members to serve as basis for their decisions. Moreover, the standards are drawn up to contribute to the principles of the Paris Declaration on harmonising efforts in evaluation (OECD/DAC, 2006). Specifically, the standards aim to:

---

[2] The thirds edition of the standards include 30 standards subdivided over 'utility standards' (8 standards), 'feasibility standards' (4 standards), 'propriety standards' (7 standards), 'accuracy standards' (8 standards) and 'evaluation accountability standards' (3).

[3] The UN Standards for Evaluation include 50 standards subdivided over 'institutional framework and management of the evaluation function' (7 standards), 'competencies and ethics' (8 standards), 'conducting evaluations' (17 standards) and 'evaluation reports' (18 standards).

[4] The checklist elaborated by the Dutch Evaluation Office includes 33 standards subdivided over 'validity' (17 standards), 'reliability' (10 standards) and 'usefulness' (6 standards).

- Provide standards for the process (conduct) and products (outputs) of evaluations;
- Facilitate the comparison of evaluations across countries (meta-evaluations);
- Facilitate partnerships and collaboration on joint evaluations;
- Better enable member countries to make use of each others' evaluation findings and reports (including good practice and lessons learned); and
- Streamline evaluation efforts (OECD/DAC, 2006: 3)

The principles of evaluation that the OECD/DAC advocates served as basis for the formulation of the standards, hence both should be read in conjunction with each other. The standards have ten major components, namely: 'rationale, purpose and objectives of an evaluation'; 'evaluation scope'; 'context'; 'evaluation methodology'; 'information sources'; 'independence'; 'evaluation ethics'; 'quality assurance'; 'relevance of the evaluation results'; and 'completeness' (OECD/DAC, 2006). The complete listing of the standards and its descriptions are presented in annex II.

On the basis of a 2008 survey on the use of the standards, a workshop in Auckland (2009) and comments of the DAC network on Development Evaluation members, the standards were finalised in 2010 (OECD/DAC, 2010). Instead of the earlier ten major components, the 'Quality Standards for Development Evaluation' currently have only four major components: 'overarching consideration'; 'purpose, planning and design'; 'implementation and reporting'; and 'follow-up and learning'. While many standards are still included in one of the four major components, some standards have been incorporated into other standards<sup>5</sup>. If anything, methodology issues seemingly get less attention in the new version of the OECD/DAC Quality Standards as compared to the 2006 version. This is somehow surprising, as compared to other frequently used evaluation standards, since the 2006 OECD/DAC evaluation standards already pay relatively little attention to methodology issues. The recent third edition of the Programme Evaluation Standards of the American Joint Committee on Standards for Educational Evaluation has e.g. even strengthened its attention on methodology and information sources with three standards covering this issue ('valid information', 'reliable information' and 'information management') (Yarbrough et al., 2011). As we feel that the 2006 version of the OECD/DAC standards is somehow closer to the other standards which are commonly used in the aid sector and beyond, we have opted to use the 2006 version as the benchmark. However, it would certainly also be a useful undertaking to redo the exercise using the 2010 version of the standards. If anything, the outcome is likely to be more positive against the 2010 standards as they include a number of new standards which relate to the commitments of the Paris Declaration and Accra Agenda for Action, including e.g. 'partnership approach', 'co-ordination and alignment',

---

[5] 'Assessment of results' (including references to attribution and contribution) and 'sampling' which were earlier included under the 'evaluation methodology' component are now included in 'explanation of the methodologies used' and 'validity and reliability of information sources' respectively. Moreover, this 'validity and reliability of information sources' standard includes both former two standards related to information sources ('transparency of information sources' and 'reliability and accuracy of information sources').

‘evaluation capacity building’ and ‘systematic consideration of joint evaluation’; issues on which the Paris Declaration Evaluation is expected to score high.

### **2.3. Evaluation framework of the Paris Declaration Evaluation Phase Two**

The Core Evaluation Team of the Phase Two Paris Declaration evaluation elaborated an evaluation framework based on earlier preparatory work documented in a paper produced in May 2009. This paper synthesis inputs from a major workshop of the International Reference Group held in February 2009 and from commissioned studies on the Paris Declaration prepared in 2006 and 2008 (IODPARC, 2009). It discusses the evaluation’s conceptual framework, the evaluation model and methodological issues as well as the governance, management and operational structures and processes.

The evaluation framework is used by the individual country evaluation teams to allow for comparison among various country experiences. Taking into account the extreme complexities of the evaluation of the implementation of the Paris Declaration and the numerous possible approaches from which the evaluation can be tackled, the evaluation framework focuses on a manageable set of core questions and sub-questions (IODPARC, 2009). The three core evaluation questions are:

1. What are the important factors that have affected the relevance and implementation of the Paris Declaration and its potential effects on aid effectiveness and development results? (The Paris Declaration in context)
2. To what extent and how has the implementation of the Paris Declaration led to an improvement in the efficiency of aid delivery, the management and use of aid and better partnership? (Process and intermediate outcomes)
3. Has the implementation of Paris Declaration strengthened the contribution of aid to sustainable development results? How? (Development outcomes) (IODPARC, 2009: 9).

These three core questions together with the framework for conclusions (listed in box 2.1.) provide a comparative common structure for the individual country evaluations as well as the final synthesis report.



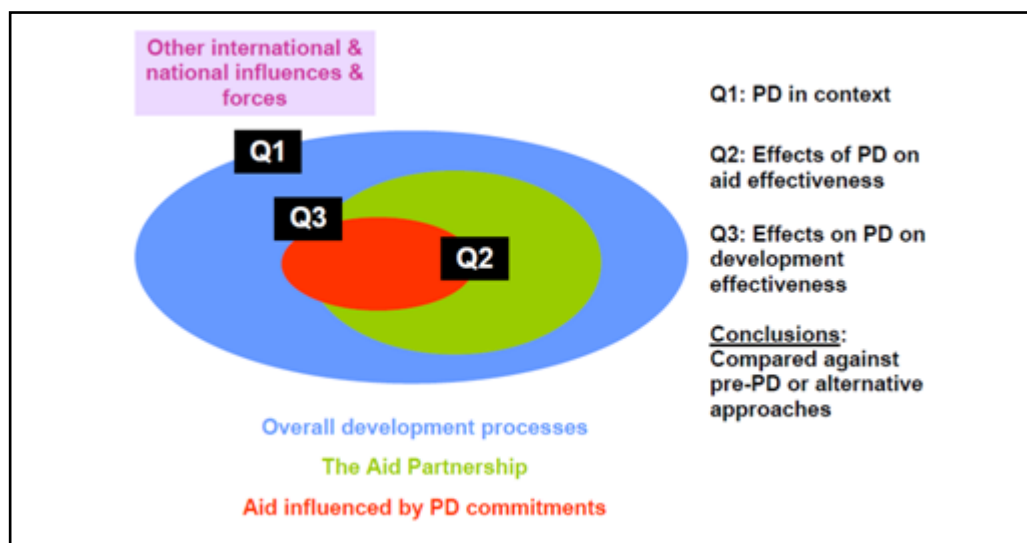
**Box 2.1. The Framework for Conclusions**

- i. What has been the relevance of the Paris Declaration and the ways it has been implemented to the challenges of aid effectiveness?
- ii. To what extent has each of the five principles of the Paris Declaration been observed and implemented, and the Accra Agenda priorities reflected? Why? Have there been conflicts or trade-offs between them?
- iii. What has the Paris Declaration achieved for aid effectiveness and development results? How significant are these contributions? How sustainable? Is there evidence of better ways to make aid more effective and contribute more to development results, for women and men and for those who are excluded?
- iv. What effects has the implementation of the Declaration had on the respective burdens of aid management falling on the partner country and its respective donors, relative to the changing volumes and quality of aid and of the aid partnership itself? Are these effects likely to be transitional or long term?
- v. What has been the added value of Paris Declaration-style development cooperation compared with the pre-PD situation, and seen alongside other drivers of development in the country, other sources of development finance and development cooperation partners beyond those so far endorsing the Declaration?
- vi. What are the key messages for a) national stakeholders, and b) donor countries and agencies?
- vii. What are the key implications for aid effectiveness in the future taking account of new challenges and opportunities (e.g. climate change) and new actors and relationships?

Source: IODPARC, 2009

The logic behind the core evaluation questions is illustrated in figure 2.2. Ideally, the structure of the core questions properly contextualises and sets out the aid subject to the Paris Declaration principles in relation with other sources of development assistance and drivers of development within each country (IODPARC, 2009).

**Figure 2.2. The Logic of the Core Questions**



Source: IODPARC, 2009

The evaluation framework also enumerates the types of evidence and indicators to be used, the anticipated data sources, and the proposed sources, methods and techniques for data collection, analysis, triangulation and validation (IODPARC, 2009).

#### **2.4. Selection of meta-evaluation questions, standards, rating system and sample.**

Both Stufflebeam (2001) and Cooksy and Caracelli (2005) include the formulation of the meta-evaluation purpose/questions as one of the first steps in performing a meta-evaluation. For this meta-evaluation we formulated two questions:

- Is the quality of the country evaluation reports good enough to be included in the synthesis report?
- Do the reports properly comply with the evaluation framework to permit comparison of evaluations across countries?

The first question will be answered by performing a meta-evaluation. As the country evaluation reports had already been completed prior to the start of our study, this study serve as a summative meta-evaluation of the country evaluation reports.

While according to the meta-evaluation steps defined by Stufflebeam (see 2.2.) the research questions should define what set of quality criteria will be selected, in this study we selected the quality criteria before the questions were formulated. This is because the OECD/DAC Evaluation Quality Standards are the most logical choice for an independent evaluation of an OECD/DAC initiative. Our study focuses on six of the ten OECD/DAC criteria, i.e. 'evaluation methodology', 'information sources', 'independence', 'quality assurance', 'relevance of the evaluation results', and 'completeness' as these are deemed to be the most relevant for our study. Other criteria, such as the 'evaluation scope' and 'rationale, purpose and objectives of an evaluation' were omitted because these parts of the country evaluation are more or less the same

for all reports. 'Context' was also not included in the criteria as this will overlap with the meta-evaluation of the 'completeness' of the country evaluation reports (see section 4. of the study), given that contextualising the Paris Declaration implementation is one of the core questions of the evaluation. Thus, assessment of the evaluations' response to the core evaluation questions indirectly measures its adherence to the 'context' standard. Lastly, 'evaluation ethics' was excluded in our study as well since assessment of this criterion requires information usually not written in reports, but which can be obtained through interviews with the evaluation teams and relevant stakeholders in each country. A checklist of the criteria from the OECD/DAC Evaluation Quality Standards is used and rated as weak (=1), partially satisfactory (=2), satisfactory (=3), and excellent (=4). An index is computed for each criteria to identify the strengths and weaknesses of the evaluation overall.

The same rating system is used for answering research question two. The evaluation of the Paris Declaration worked around three core questions, for which the evaluation framework provided sub-questions, progress markers, sources of evidence, indicators, methods of analysis, and judgements to give guidance to country evaluation teams in attempting to answer the three core questions. Our study gives a rating for each sub-question based on the completeness of discussion as required in the evaluation framework and on the depth of rigour exerted implied in the sources of evidence presented. Adherence with the evaluation framework is important for the comparability value of each country report that feeds into the main synthesis evaluation report. Thus, checking for their level of compliance is critical for the quality and validity of the synthesis report.

The rating for both research questions are applied to 15 of the 21 country evaluation reports produced, namely Afghanistan, Bangladesh, Cambodia, Cook Islands, Ghana, Indonesia, Malawi, Mozambique, Nepal, Philippines, Samoa, South Africa, Uganda, Vietnam, and Zambia<sup>6</sup>. Six country evaluation reports (Benin, Bolivia, Cameroon, Colombia, Mali, and Senegal) were not included as these were not yet translated in English at the time of our study. While we have no reason to think that this selection has led to a substantial bias in the findings of our meta-evaluation, we would need an additional check on the 6 remaining reports before we can generalise the findings based on the 15 reports to the entire sample of country reports.

While interaction with stakeholders is important within a meta-evaluation (see Stufflebeam, 2001), one of the limitations of our study is the lack of communication with the evaluations' stakeholders. Although inputs from stakeholders could have had a great added value in the meta-evaluation, due to limited time and resources, communication with the country evaluation stakeholders was not attempted. Moreover, due to time and network constraints, other materials used in producing the country evaluation reports are not tapped as well. The research is mainly a desk study of the country evaluation reports, which are the main sources of information of our study.

The next two chapters summarise findings of the meta-evaluation exercise.

---

[6] The evaluation reports of these countries are included in the list with references.

### 3. HOW WELL DO THE COUNTRY EVALUATION REPORTS PASS THE TEST OF THE OECD/DAC EVALUATION QUALITY STANDARDS?

This section presents and discusses the meta-evaluation findings which result from the assessment of 15 country PD Phase Two evaluation reports against six OECD/DAC Evaluation Quality Standards. Table 3.1. provides a summative overview of the quantitative results, while sections 3.1. to 3.6. briefly discuss the compliance of the 15 country reports with each standard, highlighting significant trends or instances worth noting. Section 3.7. gives an overview of the top five strengths and the top five weaknesses of the country evaluation reports.

**Table 3.1. Tallied scores of 15 Country Evaluation Reports for the phase two Evaluation of the Paris Declaration**

	Weak	Partially Satisfactory	Satisfactory	Excellent	Index	Ranking
<b>Evaluation methodology</b>						
Explanation of the methodology used	0	2	2	11	3.60	3
Assessment of results	1	3	5	6	3.07	6
Relevant stakeholders consulted	0	2	5	8	3.40	4
Sampling	5	1	3	6	2.67	13
Evaluation team	2	11	1	1	2.07	17
<b>Information Sources</b>						
Transparency of information sources	3	3	2	7	2.87	11
Reliability and accuracy of information sources	2	4	2	7	2.93	9
<b>Independence</b>						
Independence of evaluators vis-a-vis stakeholders	0	1	14	0	2.93	9
Free and open evaluation process	2	4	7	2	2.60	14
<b>Quality assurance</b>						
Incorporation of stakeholders' comments	4	5	2	4	2.40	15
Quality control	4	5	2	4	2.40	15
<b>Relevance of the evaluation results</b>						
Formulation of evaluation findings	0	6	5	4	2.87	11

	Weak	Partially Satisfactory	Satisfactory	Excellent	Index	Ranking
Evaluation implemented within the allotted time and budget	6	3	5	1	2.07	17
Recommendations and lessons learned	0	2	2	11	3.60	2
Use of evaluation	0	0	0	15	4.00	1
<b>Completeness</b>						
Evaluation questions answered by conclusions	0	5	4	6	3.07	6
Clarity of analysis	1	3	6	5	3.00	8
Distinction between conclusions, recommendations and lessons learned	2	2	2	9	3.20	5
Clarity and representativeness of the summary	0	3	6	6	3.20	5

### 3.1. Evaluation methodology

The 'evaluation methodology' standard consists of the 'explanation of the methodology used', 'assessment of results', 'relevant stakeholders consulted', 'sampling', and 'evaluation team'. Average scores for each criterion varies significantly, with rankings at the top five and bottom five of all the criteria.

The 'explanation of the methodology used' received the highest ranking. It concerns the description of methods and processes involved as well as the limitations encountered by the evaluation. Most of the country evaluation reports did a satisfactory job in detailing information and evidence gathering needed for answering the evaluation's three core questions. Given that a single evaluation framework was followed by all country reports, methodologies employed by the evaluations were more or less the same for everyone. Notwithstanding, the evaluation reports gave explanations on the choice and process of the methodologies used.

'Assessment of results' and 'relevant stakeholders consulted' were easily conformed by the country evaluations. Except for Afghanistan, most reports explicitly specify sources of data especially if these are results of interviews and/or surveys. The organisational set up of the evaluation required by the Core Evaluation Team also provides for the involvement of all stakeholders in the evaluation through the National Reference Group. The National Reference Group should consist of representatives from government, donor partners, civil society and possibly academia (IODPARC, 2009). It is tasked to "ensure stakeholders' participation and buy-in to the evaluation process and results and to assure the independence of the evaluation" (IODPARC, 2009: 14).

'Sampling' and 'evaluation team', on the other hand, appear to be among the weaknesses of the country evaluation reports. Most reports do not mention the criteria for selection of respondents and interviewees, though a number of them includes a list of persons consulted. Sampling varies very much across countries; Bangladesh interviewed 181 relevant

stakeholders whereas Mozambique approached 37 persons for an interview. In terms of survey respondents, among those who indicated the response rate, Nepal had a response rate of approximately 70% for the questionnaires it distributed among secretaries of ministries and development partner officials. Indonesia, on the other hand, stresses its difficulty in obtaining a high response rate for its mail survey (electronic and hard copies) despite its efforts to contact respondents individually. Six countries receive a rating of 'weak' or 'partially satisfactory' due to the lack of reporting on the sample frame, criteria for selection of respondents/interviewees and the perceived insufficient number of respondents/interviewees.

All country reports list the members of the evaluation team contracted, with the exception of Mozambique and South Africa whose overview of team members is not complete. Nevertheless, only the country report of Bangladesh includes complete information on the qualifications and areas of expertise of its consultants for the evaluation. Hence, an assessment of the mix of evaluative skills and thematic knowledge cannot be undertaken. Moreover, determining the gender balance of the evaluation teams could not be comprehended with merely the names of the members as basis. Therefore, the lack of information on the team members grounds the low marks for this criteria.

### **3.2. Information sources**

The OECD/DAC Evaluation Quality Standards require sufficient description of information sources used, i.e. documentation, respondents, literature, etc., to determine the adequacy of the information gathered (OECD/DAC, 2006). Reliability and accuracy of information are also critical in the assessment of the quality and validity of evaluation reports; the standards particularly check the triangulation of methods and sources of information used.

'Transparency of information sources' is determined by the inclusion of a list of interviewees/respondents and a list of documents consulted. Out of 15 country evaluation reports, 9 have a complete list of interviewees/respondents, or at least the organisations consulted, and an inventory of the documents utilised for the report. The Philippines, Samoa and Vietnam do not annex a list of relevant stakeholders and documents consulted, although the report of Samoa includes footnotes of its secondary sources.

'Reliability and accuracy of information sources' is assessed on the basis of whether the country evaluation reports try to triangulate its claims and answers to the core evaluation questions and sub-questions by providing different sources of evidence. Given the emphasis of the evaluation framework on using different methods of data gathering and analysis, quite a number of the country reports (9 out of 15) do at least a satisfactory job in triangulating its data sources. The reports of Bangladesh, Cambodia, Cook Islands, Nepal, South Africa, Vietnam and Zambia should be applauded for excellently cross-validating their information and claims using interviews, surveys, statistics and document review.

### **3.3. Independence**

Within the independence standard the 'independence of evaluators vis-à-vis stakeholders' and the 'free and openness of the evaluation process' are assessed. Evaluations are supposed to report on the degree of independence of the evaluators from the relevant

stakeholders of the evaluation, such as the organisations related to the commissioning agent, implementers and beneficiaries, and to explicitly address conflicts of interest (OECD/DAC, 2006).

As the commissioning of independent evaluators is a requirement for the evaluation of the Paris Declaration, independence of the evaluation team can somehow be assumed. However, all country evaluation teams are assisted by the National Evaluation Coordinator, who is usually the head of the aid management unit of the government, as well as guided by the National Reference Group comprised of various stakeholders from the government, development partners and civil society. The National Evaluation Coordinator is tasked to initiate, facilitate, contract and manage the country evaluation. It can also engage with a development partner for facilitation of the evaluation and for funding (IODPARC, 2009). Therefore by design, the independence of the evaluation team is not excellent. For this reason, a ceiling rating of 'satisfactory' was given to all but one country report. Nepal reported to have included in its evaluation team two members of its Foreign Aid Coordination Division, which further exposes the evaluation to bias in favour of the government, thus getting a 'partially satisfactory' rating.

In terms of 'free and open evaluation process', a number of countries encountered several difficulties in either obtaining quantitative data or in conducting interviews/surveys. Afghanistan and Mozambique report on external events, i.e. parliamentary elections, the Ramadan period and riots, which significantly limited the evaluation process leading to the low number of stakeholders consulted by their respective evaluation teams. Others report on the unavailability of donor partner officials during the data gathering period as it coincided with the summer season of their country of origin. This delayed the evaluation of a number of countries. Zambia, Uganda, Samoa and Afghanistan mention the lack of quantitative statistical data as constraining their pre- and post- Paris Declaration analysis. Lastly, the low and/or delayed response of survey respondents limited the evaluation critically. Because of the many constraints in the evaluation process, this criterion is ranked at the bottom (14th on a total of 18 criteria).

### **3.4. Quality assurance**

Quality assurance, as established by the OECD Evaluation Quality Standards, is guaranteed through 'incorporation of stakeholders' comments' and 'quality control'. The evaluation process should provide for measures of quality control, internally or externally, and should allow for stakeholders to comment on the report findings, conclusions, lessons learned and recommendations. Inputs from the stakeholders' comments, including substantive disagreements, must be incorporated in the report (OECD/DAC, 2006).

The organisation of the country level evaluation of the Paris Declaration is designed to have a National Reference Group, comprised of various relevant stakeholders from the government, development partners and civil society, which ensures quality control of the evaluation process and findings. The rating for the 'quality control' criteria is determined by the explicit reporting of the evaluation team's process of seeking comments of the National Reference Group and other stakeholders. Bangladesh, Mozambique, Nepal, Philippines, South Africa and Uganda refer in their respective reports to the organisation of workshop/s for presentation of the evaluation's draft findings to the National Reference Group and/or other stakeholders. Most of these countries receive an excellent rating for 'quality control', except for South Africa and Uganda. The evaluation team of South Africa failed to receive substantive

comments on its draft findings due to low attendance of development partners in the workshop and limited inputs from its National Reference Group despite its efforts. Uganda, on the other hand, reports only on the intention of holding a workshop but not on the actual conduct of it.

‘Incorporation of stakeholders’ comments’ could not be verified concretely by our study as the reports do not refer to comments raised by stakeholders during quality control workshops nor do any of them includes an annex with comments from stakeholders consulted. Thus, our study resorted to equating the score of the ‘incorporation of stakeholders’ comments’ with that of the ‘quality control’, assuming in good faith that the conduct of quality control workshops equates with reflecting stakeholders comments in the final report. Meanwhile, four countries (i.e. Ghana, Malawi, Vietnam and Zambia) were given a ‘weak’ rating in this criterion for being silent on how they assured quality control for their evaluations.

### **3-5. Relevance of the evaluation results**

The relevance of the evaluation results is determined taking into account ‘formulation of evaluation findings’, ‘evaluation implemented within the allotted time and budget’, ‘recommendations and lessons learned’, and ‘use of evaluation’.

The evaluation framework (discussed in section 2.4) elaborated to guide the country level evaluations is designed to assist country evaluations in ensuring the relevance of the findings by providing and/or determining specific progress markers, indicators, methods of analysis and judgement to be utilised by the evaluation teams. Our study assessed the ‘formulation of evaluation findings’ by looking into the reports’ compliance with the evaluation framework, i.e. whether they answered all the sub-questions, discussed the relevant progress markers and/or indicators, used the proper methods of analysis, and utilised judgements. A separate scoring system for this exercise has been used, which is discussed in section 4. Most of the country reports cluster within the ‘partially satisfactory’ and ‘satisfactory’ range, given the extreme complexity and extensiveness of the evaluation framework. A few notable country evaluations, i.e. Bangladesh, Cambodia, Nepal and Vietnam, produced a truly thorough report that details most, if not all, progress markers and/or indicators, triangulated by various sources of evidence, and providing as well judgements about progress.

With respect to ‘evaluation implemented with the allotted time and budget’, only the time element has been assessed by our study as the budget for the evaluation is not mentioned in any of the 15 country reports. Several countries report delays in the conduct of the evaluation due to issues in contracting independent consultants and conflicts with the scheduling of more established government activities. South Africa indicates in its report that it only started with the actual evaluation in November 2010, a month prior to the submission of country evaluation reports. The Vietnamese evaluation report mentions it conducted the evaluation in merely three weeks. This criterion is ranked at the bottom two among all the other criteria.

‘Recommendations and lessons learned’ of evaluations, according to the standards, should be relevant, actionable and applicable for wider use. Most country reports receive an ‘excellent’ rating as most provide detailed lessons learned and recommendations stemming out of the analysis framed by the evaluation framework. ‘Use of the country evaluation reports’ is limited to direct use in the context of the production of the synthesis evaluation report of the Paris Declaration which fed into the Busan conference in November 2011. It would be interesting



to complement desk study with field study to study to what extent the country evaluation reports are used beyond the context of the Busan conference.

### 3.6. Completeness

The assessment of the completeness of each country report focuses on 'evaluation questions answered by conclusions', 'clarity of analysis', 'distinction between conclusions, recommendations and lessons learned', and 'clarity and representativeness of the summary'.

The Core Evaluation Team includes a framework for overall conclusions in the evaluation framework for the individual country evaluations. Thus, 'evaluation questions answered by conclusions' is rated against the compliance of the country evaluation report to this framework for overall conclusions. Scoring of this criterion is based on the approximate average of the scores for all the sub-questions in the framework for overall conclusions. Most country reports are able to comply and answer appropriately the questions specified in the framework. However, three country reports (i.e. Malawi, Uganda and Vietnam) do not follow the framework and instead directly answer the three core questions in their evaluation report's conclusion. Afghanistan and Zambia, on the other hand, structure their conclusion by discussing each Paris Declaration principle, its context, implementation and impact in the country. The countries that do not comply with the framework get a rating of 'partially satisfactory.'

'Clarity of analysis' concerns the logical flow of the discussion, with the data and information systematically presented and analysed. Furthermore, clearly identified findings and conclusions should stem logically from the analysis. Most country reports give a clear analysis of its data, especially when they strictly adhere to the evaluation framework; thus nine countries are rated as either 'partially satisfactory' or 'satisfactory' with regards to this criterion. Laudable reports with excellent discussion of analysis are Bangladesh, Cambodia, Nepal, Vietnam and Zambia.

The OECD/DAC Evaluation Quality Standards advocate the 'distinction between conclusions, recommendations and lessons learned' in evaluation reports, thus requiring that these elements are separately presented with clear logical distinctions among them (OECD/DAC, 2006). Among the 15 country reports, 11 are able to comply with this standard by clearly distinguishing the conclusions, recommendations and lessons learned from each other. However, the remaining four countries, (Ghana, Philippines, Samoa and Zambia), fail to meet this criterion satisfactorily. The reports of Ghana and Zambia do not distinguish their discussions of recommendations from lessons learned, while the Philippines and Samoa do not have sections (or implied in other parts) for lessons learned and recommendations.

'Clarity and representativeness of the summary' refers to the executive summary of evaluations. The standards require summaries to contain an overview of the report and to emphasise the main conclusions, recommendations and lessons learned. All country reports include an executive summary, with six reports (i.e. Cambodia, Cook Islands, Nepal, Samoa, South Africa and Zambia) giving an excellent summary of their evaluation study, while three (Ghana, Philippines and Vietnam) only partially meet the standards as they fail to give a substantive overview of the findings and highlights of the main conclusions, recommendations and lessons learned.

### 3-7. Strengths and weaknesses

Based on the results of the meta-evaluation of the country evaluation reports using the OECD/DAC Evaluation Quality Standards, the five elements that score best include:

- Use of evaluation (relevance)
- Recommendations and lessons learned (relevance)
- Explanation of methodology used (methodology)
- Relevant stakeholders consulted (methodology)
- Distinction between conclusions, recommendations and lessons learned (completeness); and clarity and representativeness of summary (completeness)

Thus the strongest elements are situated within the standards 'relevance of the evaluation results', the 'evaluation methodology' and the 'completeness'. This might to some extent be related to the long gestation period of the phase two evaluation of the Paris Declaration, which focused on the pursuit of a good evaluation methodology realising that the evaluation would be burdened by attribution complexities. The OECD/DAC commissioned a Core Evaluation Team composed of six international consultants to function as the core team members, and a handful of associated team members hired for specific tasks. The Core Evaluation Team gave valuable inputs during the planning and set-up stage and ensured the consistency and smooth operations of the country evaluations (IODPARC, 2009). This support is visible in the efforts of the individual country reports to adhere to the evaluation methodology. Moreover, the evaluation framework provides very detailed instructions on the possible indicators, sources of evidence and methods of analysis to be employed by the country evaluation teams. It highly emphasises the validation of data by using several methods of gathering information. The framework details such for each core question of the evaluation, hence the completeness of the reports and the relevance of the evaluation results.

In line with our findings, the official meta-evaluation includes the elaboration of the evaluation framework in its list of major strengths of the Paris Declaration evaluation. According to the official meta-evaluation, the framework was especially useful in providing coherence to the evaluation. A large majority of stakeholders involved in the evaluations (87% of national stakeholders, 96% of the International Reference Group) indicated in their responses to the survey of the meta-evaluation that the framework was 'somewhat to very useful/ important' (Patton and Gornick, 2011). The two elements in our top five strengths within the relevance standards ('use of evaluation' and 'recommendations and lessons learned') are also included in the list of major strengths of the official meta-evaluation under 'being utilisation-focused throughout'. The positive assessment of the evaluation's usefulness in both our own meta-evaluation and the official meta-evaluation stems from the fact that the Paris Declaration evaluation has been designed to be an input for the 4th High Level Busan Forum on Aid Effectiveness in November 2011. Moreover, the official meta-evaluation exemplifies the usefulness of the evaluation by referring to the inclusion in the evaluation of recommendations differentiated for the specific users: policymakers in partner countries, policymakers in donor countries, and policymakers in both partner countries and donor countries and their agencies (Patton and Gornick, 2011).

In addition to the five top strengths, our study also identifies the top five weaknesses

of the country evaluations. These include:

- Evaluation implemented within the allotted time and budget (relevance)
- Evaluation team (methodology)
- Quality control (quality assurance)
- Incorporation of stakeholders' comments (quality assurance)
- Free and open evaluation process (independence)

The weaknesses of the country evaluations are possibly connected with each other, with 'timing' as the main cause of the weaknesses. Several countries report on being constrained by time because of delays in contracting the country evaluation teams, conflict with other in-country activities of the government and development partners, and external events that prevented the teams to conduct the evaluation within the allotted timeframe.

Because of the time limits, the country evaluations were also constrained in data gathering. A handful of countries scores negatively on 'free and open evaluation process', which is part of the independence criteria. As this criteria consists of only two parts (the other is 'independence of evaluators vis-a-vis stakeholders') the picture arises that the country evaluations were not independent, while this is one of the strengths identified in the official meta-evaluation: 'establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout'. The difference in appreciation of 'independence', however, seems to be caused by a different interpretation of independence. The inclusion of this point in the list with major strengths in the official meta-evaluation relates to the establishment of a free-standing and independent secretariat, the governing structures, the management arrangements, the engagement of the International Reference Group of diverse stakeholders and the terms of reference for the Core Evaluation Team. The independence of the evaluation has been maintained and supported by the process of data collection, analysis, interpretation and stakeholders review (Patton and Gornick, 2011). The OECD/DAC standards describe the 'free and open evaluation process' as: "The evaluation team is able to work freely and without interference. It is assured of cooperation and access to all relevant information. The evaluation report indicates any obstruction which may have impacted on the process of evaluation" (OECD/DAC, 2006: 6). The low score relates to insufficient availability of quantitative statistical data, low and/or delayed response rate, unavailability of donor partner officials and external events, which could have been circumvented with more time allotted for the conduct of the evaluation. These elements are also referred to in the official meta-evaluation, which mentions that due to short timelines, inadequate resources, capacity limitations and administrative and implementation challenges, many partner country reports struggled with quality of data: "response rates on surveys were low. The number and quality of interviews were frequently poor, and analysis of interview data was often less than systematic and rigorous. Many country reports thus relied heavily on existing data sources and reports, government and agency documents, and a few key resource people with in-depth knowledge. In short, "primary data collection was often secondary; secondary data were often primary" (Patton and Gornick, 2011: 27). Moreover, one of the participant of the survey is quoted in the report: "Access to evidence and data from donors was difficult, hence not adequate reflected. Moreover, few donors appeared to be interested in extending full support" (Patton and Gornick, 2011: 25).

‘Quality assurance’ is another weakness of the country evaluations. Ratings for ‘quality control’ and

‘incorporation of stakeholders’ comments” are equated in our study as it was not possible to determine the latter with only the written reports as basis. The organisation design of the evaluation supposedly ensures quality assurance, having the National Reference Group (composed of representatives from the government, development partners and civil society organisations) check for the quality of the process and outputs of the country evaluation team. However, only a handful reports indicates the conduct of quality control activities/workshops, hence the low rating given to this criteria. Time constraint could also have impacted on quality assurance as country evaluation teams were pressing for time, trying to meet the global deadline could have been at the expense of conducting activities/workshops for quality check.

The criterion for ‘evaluation team’ is ranked low because of insufficient information in the reports on the qualifications and areas of expertise of the consultants commissioned by the individual countries for their respective evaluations. In addition, determining the gender balance of the evaluation teams could not be comprehended with merely the names of the members as basis.

#### **4. HOW WELL DO THE COUNTRY EVALUATION REPORTS COMPLY WITH THE EVALUATION FRAMEWORK?**

This section lays out the research findings with regard to the level of compliance of the 15 country reports with the evaluation framework developed by the Core Evaluation Team. One of the objectives of this study is to determine the degree of compliance of the country evaluation reports with the evaluation framework to ascertain whether comparison among country reports is possible and to determine the level of comparability. For each core question of the phase two Evaluation of the Paris Declaration, the evaluation framework specifies the type of evidence and/or indicators to be used as well as the proposed methods for data collection, analysis and triangulation (IODPARC, 2009). These are suggested to ensure comparability among the individual country evaluations. Thus, to be able to establish the level of comparability of the country reports, our study uses a simple rating system applied to each evaluation component. The three evaluation components of the evaluation framework are key characteristics (core question one), expected outcomes of the Paris Declaration (core question two), intended development results (core question three). Each country report is given a rating of ‘weak’ (=1), ‘partially satisfactory’ (=2), ‘satisfactory’ (=3) or ‘excellent’ (=4) depending on the degree of adherence with the discussion of evidence and/or indicators, appropriate use of methods of analysis and proper use of judgements on progress.

Shown in table 4.1. are the tallied scores of the country reports on their level of compliance with the evaluation framework, aggregated by the three core questions. Indicated in the second column of the table is the number of components that were given a rating for each of the 15 country evaluation reports.

**Table 4.1. Tallied scores of the 15 Country Evaluation Reports on their compliance with the Evaluation Framework (aggregated by the three core evaluation questions)**

	No. of components	Weak	Partially Satisfactory	Satisfactory	Excellent	Index	Ranking
<b>Core Q1:</b> What are the important factors that have affected the relevance and implementation of the Paris Declaration and its potential effects on aid effectiveness and development results? (The Paris Declaration in context)	20 key characteristics	60 (20%)	50 (17%)	80 (27%)	110 (37%)	2.80	<b>2</b>
<b>Core Q2:</b> To what extent and how has the implementation of the Paris Declaration led to an improvement in the efficiency of aid delivery, the management and use of aid and better partnerships? (Process and intermediate outcomes)	11 expected outcomes	18 (11%)	48 (29%)	62 (38%)	37 (22%)	2.72	<b>3</b>
<b>Core Q3:</b> Has the implementation of Paris Declaration strengthened the contribution of aid to sustainable development results? How? (Development outcomes)	9 intended development results	22 (16%)	24 (18%)	31 (23%)	58 (43%)	2.93	<b>1</b>
<b>Framework for Overall Conclusions</b>	7 sub-questions	40 (38%)	9 (9%)	8 (8%)	48 (46%)	2.61	<b>4</b>

For core question one, twenty key characteristics were identified for the five sub-questions formulated to answer the question on the context of the Paris Declaration implementation. Core question two, on the process and intermediate outcomes, has eleven expected outcomes aggregated by 'country ownership', 'inclusive partnership' and 'development results'. Meanwhile, for core question three, nine intended development results are specified to determine answers to the four sub-questions on development outcomes. Lastly, the conclusion is collapsed into seven sub-questions. Breakdown of table 4.1. into sub-questions is presented in annex III.

Presentation of the results of the rating system applied to the 15 country evaluation reports are tackled by giving a summary of the level of compliance per core evaluation question in 4.1., 4.2., 4.3. and 4.4. The top five strengths and top five weaknesses are presented in 4.5.

#### **4.1. Paris Declaration in context**

The evaluation framework determines five sub-questions to enable country evaluation teams to describe the context of the Paris Declaration implementation in their respective countries and to answer the first core evaluation question. Based on the rating system, this core evaluation question is second to be completely complied upon by the country evaluation reports.

A number of characteristics are identified by the evaluation framework as key to determining the context of Paris Declaration implementation. Sources of evidence/indicators, methods or forms of analysis and categories for analysis and judgement are also provided for in the framework. Our study gives marks on the basis of the completeness of utilisation of these instruments by the country reports as suggested by the evaluation framework, the provision of sufficient sources of evidence/indicators in the reports to describe each key characteristic and the use of categories for analysis and judgement, where necessary.

As shown in annex III, the country reports are weakest in identifying the key actors involved in aid-related decisions and the influence of the Paris Declaration and Accra Agenda for Action in relation with their priorities and incentives. Specifically, most countries fail to describe or completely omit to mention the degree of decentralised decision-making between donor headquarters and country field office, while Nepal misunderstood the question completely as decentralisation within the government bureaucracy. A number of country reports also give insufficient information on the mechanisms for parliamentary and civil society oversight on the budget and aid allocation, with seven countries (i.e. Afghanistan, Ghana, Malawi, Mozambique, Philippines, South Africa and Uganda) rated as 'weak' on this key characteristic.

On the other hand, many country reports (9 out of 15) give a full picture of the extent of the implementation of the Paris Declaration principles, detailing the evolution of the countries' engagement with the Paris Declaration initiatives, the policies put in place that directly/indirectly promote the Paris Declaration principles and the specific development and/or reform initiatives that reflect the principles, hereby making this characteristic the most complied upon by the reports.

For most country evaluations, information on the priorities and incentives of actors within the government and donor agencies could not be obtained easily through secondary data and should therefore be extracted from key informant interviews. The same applies to data on decentralisation from donor headquarters to country offices, which might be the reason why these key characteristics rank lowest in this category as its collection of data required more effort and cooperation from development partners than the others.

Overall, most of the country reports are able to present enough information to contextualise the Paris Declaration implementation, with facts and figures on pre- and post Paris Declaration events, trends and initiative.

#### **4.2. Process and intermediate outcomes**

Discussion of the process and intermediate outcomes has the lowest compliance for the 15 country evaluation reports. The second core question should be answered by providing analysis on the three main areas for improvement determined in the 3rd High Level Forum in

Accra, i.e. 'country ownership over development', 'building more inclusive and effective partnerships for development', 'and delivering and accounting for development results' (see section 1).

The evaluation framework lists the expected outcomes of the Paris Declaration per main area for improvement and identifies progress markers, potential indicators of change/ milestones, methods/forms of analysis and rating for judgement on progress made. Assessment on the adherence of the country reports to the sought information is undertaken by developing minimum criteria for each rating judgement:

'excellent'	the report discusses at least half of the progress markers, uses proper method/s of analysis and gives explicit judgement on the country's progress.
'satisfactory'	the report tackles at least half of the progress markers, uses proper methods of analysis, but with insufficient judgement of progress made/ the report gives details on at least half of the progress markers, gives sufficient judgement of progress, but uses insufficient methods of analysis.
'partially satisfactory'	the report discusses less than half of the progress markers, but gives sufficient explicit judgement on progress made and uses proper methods of analysis/ the report discusses more than half of the progress markers, but does not give judgements on progress made and does not use proper methods of analysis.
'weak'	the report discusses the progress markers very lightly, or none at all, and does not mention progress made over the years and does not use proper methods of analysis.

Application of the scoring system shows the strength in the discussion of the 'partnership for development' component. More than half of the country reports did at least a satisfactory job at explaining the collaborative nature of donors in terms of policies, procedures, multi-year commitments, delegation of authority and integration of global programmes. Discussion of most, if not all, of these process and intermediate outcomes necessitated documentary analysis, key informant interviews and survey questionnaires. Country reports getting high marks for these components use at least two of the methods of analysis instructed by the evaluation framework.

As indicated in the evaluation framework, the results of the Paris Declaration Monitoring Survey should also be incorporated in the discussion of the process and intermediate outcomes. With the exception of Cook Islands and Samoa, the countries participated in at least one round of the Paris Declaration Monitoring Survey. Incorporation of the monitoring survey results is inconsistent for most of the evaluation reports. Only the reports of Bangladesh and Nepal have a consistent effort in including the monitoring survey results in its discussion points.

Discussion of 'country ownership over development' receives the lowest ranking among the three main components with quite a number of country reports rated as 'weak' in presenting the 'national strategies and operational frameworks', 'alignment of aid with partner countries' and 'measures and standards of performance and accountability of partner country systems'. In addition, seven of the Paris Declaration Monitoring Survey indicators are expected to be discussed in this section. As discussed earlier, most countries are inconsistent in including the monitoring survey indicators in their discussion and analysis of process and intermediate outcomes. In addition, several countries fail to give judgements on the progress made since

the advent of the Paris Declaration, with most merely presenting the events and initiatives that have been undertaken.

Overall, the evaluation reports of Bangladesh and Nepal have the most complete discussion of the process and intermediate outcomes component of the evaluation with very good triangulation of data sources and explicitly expressing judgement on progress made since the first implementation of the Paris Declaration. Samoa and Uganda are on the other side of the spectrum. The Samoa report mainly fails to give judgements on the progress made in terms of achieving the progress and intermediate outcomes. Uganda's report does not discuss quite a number of the progress markers and focuses instead on only one or two markers. While this may not necessarily be a bad thing, as our study is concerned with the level of compliance with the evaluation framework, Uganda receives low ratings due to it.

#### **4.3- Development outcomes**

Core evaluation question three, regarding development outcomes, appears to be the most appropriately answered by the country evaluations reports from the viewpoint of adherence with the evaluation framework. Scores are given on the basis of the reports' 'completeness of discussion of interim development results', 'contribution of aid to the sector', and 'effects of Paris Declaration on the aid relationship'.

Case studies of selected sectors (at least the health sector and optionally one or two other sectors) were used to answer the first sub-question. All but one country report presents an extensive sector level analysis of development results that could have been enhanced through the application of the Paris Declaration principles. Regardless of the country evaluations' findings of the Paris Declaration's contribution or non-contribution to development results, the reports are given a satisfactory rating. As they discuss the matter and provide reasons for the findings, the country evaluations are credited for their efforts. While levels of depth of analysis vary among the country reports, all reports are able to detail the events and initiatives undertaken to achieve sector level results and subsequently deduce contribution or non contribution to the Paris Declaration. Effects of Paris Declaration on the aid management and delivery are tackled as well in the country reports.

In terms of answering the second sub-question regarding 'prioritisation of the needs of the poorest of the poor, including women and girls', most countries discuss the matter sufficiently with the exception of Afghanistan and Uganda. Both countries only present government initiatives in improving prioritisation of the poor and fail to mention the role and contribution of aid in those efforts, hence a 'weak' rating is given. In addition, Afghanistan fails to mention the gender aspect of this outcome.

The third and fourth sub-questions on institutional capacities and social capital, and mix of aid modalities have almost the same spread of scores with quite a number of the country reports scoring only partially or weakly. Some fail to mention social capital completely and focus only on the institutional capacity status and initiatives. Meanwhile, discussion on the mix of aid modalities is mainly focused on the increased trend in programme based approaches, itemising the programme based approaches that have been implemented. The reports lack details on the process leading up to the implementation of the programme based approaches,



more specifically whether it was an agreed optimal mix of modalities between country and development partner or not.

#### **4.4. Framework for overall conclusions**

The framework for overall conclusion has been formulated together with the three core evaluation questions to serve as the backbone of all country reports as well as the synthesis report for comparability. Seven sub-questions are framed to provide conclusions based from the presentation and analysis of data of the three core questions.

Out of 15 country reports, 10 follow the framework in preparing the conclusions of their respective reports. However, five country evaluations (i.e. Afghanistan, Malawi, Uganda, Vietnam and Zambia) formulate their own structure for this section of the report. Afghanistan and Zambia frame their conclusion by discussing the five Paris Declaration principles, while Malawi, Uganda and Vietnam uses the core evaluation questions to structure their conclusions. Because of the non-compliance of these country reports with the framework, indexing and ranking of the sub-questions for the overall conclusions are distorted, as this automatically give them a 'weak' rating in all sub-questions for the conclusion, regardless of whether they answer the sub-questions implicitly in their own-structured discussion. Hence, no matter the quality of conclusions of the other country evaluations, the ranking of the overall conclusions section is lowest. Nevertheless, taking out the five countries that do not follow the framework for conclusions, at least 60% of the remaining countries satisfactorily answer the sub-questions for the conclusion.

The country reports that comply with the framework are able to answer the relevance of the Paris Declaration and the extent of implementation of the five Paris Declaration principles in their respective countries. On the other hand, indexes for the questions on the contributions of the Paris Declaration on aid effectiveness and development results and on the value-added of the Paris Declaration are ranked low because of the distortions discussed above. Nevertheless most countries try to answer the questions using the facts and figures presented in the discussion of the three core questions.

In terms of the burdens of aid management, the country reports that followed the framework are able to answer this satisfactorily, albeit contrasting views for some. The report of the Philippines receive low marks in this component as this aspect is discussed very minimally.

#### **4.5. Strengths and weaknesses**

The official meta-evaluation of the evaluation synthesis of the Paris Declaration points at the variation in quality of evidence as one of the major weaknesses of the evaluation, but refers as well to the fact that the synthesis report is open and transparent in recognising this variation (Patton and Gornick, 2011). In fact, for core question two, the Paris Declaration evaluation provides an overview of the strength of evidence for each of the 11 intended outcome. Table 4.2. summarises the findings for the three main areas of core question two whereas full details for the 11 outcomes are presented in annex IV.

**Table 4.2. Strengths of evidence for intended outcomes in the three main areas of core question two.**

Main area	Good	adequate	poor
Country ownership over development	3 (75%)	1 (25%)	0
Building more inclusive and effective partnership for development	4 (57%)	2 (29%)	1 (14%)
Delivering and accounting for development results	2 (33%)	3 (50%)	1 (17%)

Source: Wood et al., 2011

The official meta-evaluation lists the strongest and weakest evidence ratings on the basis of a survey among partner country participants and members of the International Reference Group. Similar to our meta-evaluation, these ratings are not limited to the second core question, but include elements of the other core questions and overall conclusions as well. As a matter of comparison, we have listed the five major strengths and weaknesses identified in the two meta-evaluations in table 4.3<sup>7</sup>. The comparative exercise was somehow hindered by the fact that the official meta-evaluation's list of strongest and weakest evidence ratings includes (parts) of core questions as well as sub-elements of these same core questions<sup>8</sup>.

[7] To facilitate comparison we used the same formulation of the sub-questions and expected outcomes for our meta-evaluation and for the official meta-evaluation (in the official meta-evaluation report the formulation is shorter).

[8] For example the list of strongest evidence ratings includes 'improvements in the efficiency of aid delivery' and 'overall improvements in the management and use of aid', which are both parts of the second core question: 'To what extent and how has the implementation of the Paris Declaration led to an improvement in the efficiency of aid delivery, the management and use of aid and better partnerships?' Building more inclusive and effective partnerships for development', which is also included in the list of strongest evidence ratings, is one of the three main areas of question two.

**Table 4.3. major strengths and weaknesses identified in our meta-evaluation and the official meta- evaluation**

Our meta-evaluation	Official meta-evaluation
Strengths	
<ul style="list-style-type: none"> <li>- To what extent and where have the Paris Declaration principles been implemented? (sub-question for core Q<sub>1</sub>)</li> <li>- What are the most important national and international events that have affected the implementation of the Paris Declaration and the Accra Agenda for Action priorities, and how? (sub-question for core Q<sub>1</sub>)</li> <li>- Building more inclusive and effective partnerships for development (expected outcome for core Q<sub>2</sub>)</li> <li>- What are the key characteristics of the country that have been most relevant to the implementation of the Paris Declaration? (sub-question for core Q<sub>1</sub>)</li> <li>- Were results in specific sectors enhanced through the application of the Paris Declaration principles? (sub-question for core Q<sub>3</sub>)</li> </ul>	<ul style="list-style-type: none"> <li>- Country ownership over development (expected outcome for core Q<sub>2</sub>)</li> <li>- To what extent and where have the Paris Declaration principles been implemented (sub-question for core Q<sub>1</sub>)</li> <li>- To what extent and how has the implementation of the Paris Declaration led to an improvement in the efficiency of aid delivery (part of core Q<sub>2</sub>)</li> <li>- Building more inclusive and effective partnerships for development (expected outcome for core Q<sub>2</sub>)</li> <li>- To what extent and how has the implementation of the Paris Declaration led to an improvement in the management and use of aid (part of core Q<sub>2</sub>)</li> </ul>
Weaknesses	
<ul style="list-style-type: none"> <li>- What has been the added value of Paris Declaration-style development cooperation compared with the pre- Paris Declaration situation, and seen alongside other drivers of development in the country, other sources of development finance and development cooperation partners beyond those so far endorsing the Declaration? (sub-question for conclusion)</li> <li>- What has the Paris Declaration achieved for aid effectiveness and development results? How significant are these contributions? How sustainable? Is there evidence of better ways to make aid more effective and contribute more to development results, for women and men and for those who are excluded? (sub-question for conclusion)</li> <li>- Which are the key actors, in the country and among its development partners, who take major decisions on aid? What influence do the Paris Declaration and Accra Agenda for Action commitments have on them, in relation to their other priorities and incentives? (sub-question for core Q<sub>1</sub>)</li> <li>- What effects has the implementation of the Declaration had on the respective burdens of aid management falling on the partner country and its respective donors, relative to the changing volumes and quality of aid and of the aid partnership itself? Are these effects likely to be transitional or long term? (sub-question for conclusion)</li> <li>- Country ownership over development (component for core Q<sub>2</sub>)</li> </ul>	<ul style="list-style-type: none"> <li>- Did the implementation of the Paris Declaration help countries to improve the prioritisation of the needs (beyond income poverty) of the poorest people, including women and girls? (sub-question for core Q<sub>3</sub>)</li> <li>- Has the implementation of the Paris Declaration strengthened the contribution of aid to sustainable development results? (core Q<sub>3</sub>)</li> <li>- Has Paris Declaration implementation led to sustainable increases in institutional capacities and social capital at all levels to respond to development challenges? (sub-question for core Q<sub>3</sub>)</li> <li>- What are the key implications for aid effectiveness in the future taking account of new challenges and opportunities (e.g. climate change) and new actors and relationships? (sub-question for conclusion)</li> <li>- Delivering and accounting for development results (expected outcome for core Q<sub>2</sub>)</li> </ul>

Source: Patton and Gornick, 2011; section four.

Table 4.3 highlight substantial differences in the top 5 strengths and particularly in the weaknesses identified. These differences are probably related to the fact that our meta-evaluation is based on the country evaluation reports, while the official meta-evaluation represents the opinion of the national coordinators and evaluation team leaders. While on the one hand the national coordinators and evaluation team leaders might be biased, as it concerns 'their' evaluation, on the other hand these actors could put forward issues which are not incorporated in the reports and which thus escaped our attention.

A clear difference between the two meta-evaluations is that most of the strengths in our meta-evaluation are part of core question 1 (3/5) while most of the strengths in the official meta-evaluation are part of core question 2 (4/5). Our ranking might to a large extent be explained by the fact that country evaluations could heavily rely upon secondary data sources to address issues with respect to the contextualisation of the Paris Declaration implementation

(core question 1) while to answer satisfactorily core questions 2 and 3 secondary data needs to be complemented with primary data collection using surveys and/or structured or semi-structured interviews which are more challenging (in terms of time and methodology) (see 3.7).

A sharp contrast among the two meta-evaluations is the fact that 'country ownership over development' (expected outcome for core Q2) has the strongest evidence rating in the official meta-evaluation (and has no poor rating in the more detailed overview in the synthesis report, see table 4.2.), while it is included in the top five of weaknesses in our meta-evaluation. In the survey 80% of the national coordinators and 60% of the evaluation team leaders were of the opinion that there was strong evidence presented in their reports for 'country ownership over development' (Patton and Gornick, 2011). In our meta-evaluation 'country ownership over development' is included in the top five of weaknesses especially because of the inconsistent discussion of the indicators of the Paris Declaration Monitoring Survey. More than half of the monitoring survey indicators should be incorporated in this part of the report, however, most country evaluations only selected a few indicators to be discussed in their reports. In addition, judgements of progress since the implementation of the Paris Declaration are hardly included despite of it being required by the evaluation framework. Possibly countries feel that giving judgements overrides the report on the Paris Declaration Monitoring Survey, hence the omission of judgement and the monitoring survey indicators.

A weakness pointed to by both meta-evaluations is compliance with the evaluation framework's requirements for the identification of key actors involved in aid-related decisions. The information for this question, and particularly the 'sensitive' information regarding interests, capacities, priorities and motivations of actors, needs to be extracted from primary data. As mentioned above, primary data collection, particularly on sensitive issues, is often more challenging, which was also further compounded by the unavailability of donor officials.

## 5. CONCLUSIONS AND RECOMMENDATIONS

This section is based on the findings of the meta-evaluation (sections 3. and 4.) and feeds back to the research questions established at the beginning of the study. Implications for future similar evaluations are presented in 5.2. and, based on the conclusions, a few recommendations for further studies, are offered in 5.3.

### 5.1. Conclusions

In the pursuit of conducting a meta-evaluation of the country evaluations that fed into the phase two Evaluation of the Paris Declaration, our study tries to answer the questions: 'Is the quality of the country evaluation reports good enough to be included in the synthesis report?' 'Did the reports properly comply with the evaluation framework to permit comparison of evaluations across countries?'

Our study uses the OECD/DAC Evaluation Quality Standards to check for the level of quality and validity of the country evaluation reports. A simple scoring system is devised to arrive at the strengths and weaknesses of the country evaluations. Compliance with the evaluation framework is assessed using the same scoring system.

Findings from the meta-evaluation using the OECD/DAC Evaluation Quality Standards and the findings related to the evaluation framework compliance are consistent in pointing out the strength of the evaluation methodology. Nevertheless, as pointed to in 2.2., other meta-evaluation standards generally include more standards related to methodology and will consequently give more weight to these issues in meta-evaluation exercises. Therefore, if we would have used other evaluation standards in this meta-evaluation, our assessment might be more critical on methodology. On the other hand, it is obvious and also acknowledged in the official meta-evaluation, that the robust evaluation methodology developed by the Core Evaluation Team to ensure the proper and appropriate conduct of the individual country evaluations, contributes positively to the quality of the country evaluation reports. As validated by our findings with respect to the evaluation framework compliance, most of the country evaluations adhere to most of the instructions set forth in the framework. The adherence with the evaluation framework's specifications of types of evidence and indicators to use as well as the sources, methods and techniques for analysis and validation feeds into the high rating on 'evaluation methodology', 'relevance of evaluation results' and 'completeness'.

A robust methodological design enhances the quality of the evaluation, however, its proper application needs to be double checked. A number of the country evaluations encountered several problems in conducting the evaluations. Time was a main constraint in most evaluations, with some citing delays in contracting independent evaluators as the primary reason. Delays in contracting the country evaluation team snowballed into problems with data collection. Country reports also identify adverse events, unavailability of donor partner officials and low and/or delayed survey response as limitations of their evaluations. These issues could have been restricted or addressed if evaluation teams had more time to conduct the evaluation. With this in mind, the validity of the data gathered, especially through interviews and surveys, can be questioned. In addition, half of the country reports give insufficient information on the selection process of the interviewees and/or respondents that were consulted for the evaluation. Hence, there is less basis to check for the validity of the process employed in data gathering, which con-

sequently questions the quality of the evaluation product.

The organisational design of the country evaluations provides a platform for quality assurance of the evaluation process and product through the National Reference Group. This set-up was meant to ensure the quality of the individual country evaluations. However, since our study is limited to the reports of the country evaluations as the sole source of information, the utility of the National Reference Group is based on the details written in the reports regarding the quality assurance measures undertaken by the countries. The restrictions of our study also limits the assessment of the country evaluations' 'incorporation of stakeholders' comments' as these are not explicitly emphasised in any of the reports nor is an annex included containing the comments.

The evaluation framework developed by the Core Evaluation Team is properly complied upon by approximately 60% of the 15 country evaluations included in this study. The evaluations use, more or less, the same methodology and focus on the same aspects and issues related to the implementation of the Paris Declaration. The country evaluations are provided with a specific evaluation structure on how to tackle the three core questions formulated in trying to trace the contribution of the Paris Declaration in aid effectiveness and consequently on development results. The types of evidence and indicators to be used are specified in the evaluation framework as well as the proposed sources, methods and techniques for data gathering, analysis and validation (IODPARC, 2009). In this regard, it could be said that the comparability of the country evaluations is satisfactory as our study confirms that they properly follow the common approach in tackling the evaluation purpose, hence comparison across country evaluations can be better facilitated. This is, however, opposed by the several limitations and constraints that plague a number of the country evaluations. Using the 2006 OECD/DAC Evaluation Quality Standards as benchmark, the quality of the evidence presented is questionable, and consequently the soundness of the findings as well. As Cooksy and Caricelli (2005) indicate, synthesis reports are "only as good as the evaluation findings they synthesise" (Cooksy and Caricelli, 2005: 32). Our study suggests that the findings of some of the country evaluations are disputable, hence their inclusion in the evaluation synthesis report is questionable.

## **5.2. Implications for future similar evaluations**

Demand for evaluations of the same calibre as the Evaluation of the Paris Declaration will surely proliferate in the next few decades as international cooperation has been increasing over the past years. Information from research studies on this type of evaluations, e.g. joint evaluation among several partner countries and donors, will be useful for the development cooperation community and for the evaluation field as well. Our study, a meta-evaluation of the country-level evaluations of the Paris Declaration, identifies the strengths and weaknesses in conducting such joint evaluations.

Findings from our study establish that a robust common evaluation methodology benefits the whole undertaking in terms of comparability. This methodology, however, must be brewed from early preliminary work with participation from key stakeholders who will be involved in the evaluation. By doing so, a joint effort is forged and a common understanding is reached among stakeholders on how to conduct the evaluation. Our study confirms that efforts

in pursuing a quality evaluation at the high-level are translated with much enthusiasm into the country-level. Ensuring a free and open evaluation process is, however, critical to the process. The quality of evidence that can/should be gathered depends on the working conditions of the evaluators. Timing is especially an important factor to consider as the whole evaluation process hinges on the working deadlines imposed on the evaluators. The phase two Evaluation of the Paris Declaration was coordinated by a Core Evaluation Team, which also provided support to the country-level evaluations. This set-up could be utilised to give direction to country-level evaluation teams most especially when they are faced with several issues and bottlenecks.

### **5.3- Possible routes for further related research**

Based on the findings, nuances and limitations of our study, we recommend the following possible routes for further research:

- Meta-evaluation of the country evaluations using the 2010 version of the OECD/DAC Evaluation Quality Standards as the benchmark, analysing amongst others the degree to which the country evaluations have been used as M&E capacity building exercises;
- Meta-evaluation of the country evaluations with participation of the country evaluation teams to be able to get a broader picture of the evaluation processes that were undertaken;
- Case study of a single country evaluation to have an in-depth analysis of the country's evaluation capacity in relation to evaluating an international agreement aimed at reforming processes;
- Meta-evaluation of the donor studies and/or supplementary studies in terms of their quality and validity to be included in the final synthesis report on the phase Two Evaluation of the Paris Declaration.

in pursuing a quality evaluation at the high-level are translated with much enthusiasm into the country-level. Ensuring a free and open evaluation process is, however, critical to the process. The quality of evidence that can/should be gathered depends on the working conditions of the evaluators. Timing is especially an important factor to consider as the whole evaluation process hinges on the working deadlines imposed on the evaluators. The phase two Evaluation of the Paris Declaration was coordinated by a Core Evaluation Team, which also provided support to the country-level evaluations. This set-up could be utilised to give direction to country-level evaluation teams most especially when they are faced with several issues and bottlenecks.

## REFERENCES

- 3rd High Level Forum on Aid Effectiveness (2008) *Accra Agenda for Action*, 3rd High Level Forum on Aid Effectiveness, Accra.
- 4<sup>th</sup> High Level Forum on Aid Effectiveness (2011) *Busan Partnership for Effective Development Co-operation*, 4<sup>th</sup> High Level Forum on Aid Effectiveness, Busan.
- Alkin, M.C. and Coyle, K. (1988) "Thoughts on evaluation utilization, misutilization and non-utilization", *Studies in Educational Evaluation* 14: 331-340.
- Baawar Consulting Group (2010) "Joint Evaluation of the Paris Declaration Phase 2: Islamic Republic of Afghanistan", *Country Evaluation Report*, Ministry of Finance, Islamic Republic of Afghanistan.
- Canlas, D., Galvez-Tan, J., Abuyuan, L. (2008) "First Phase of the Evaluation of the Implementation of the Paris Declaration Country Level Evaluation Philippines", *Country Evaluation Report*, Government of the Philippines.
- Canlas, D., Galvez-Tan, J., Abuyuan, L. (2011) "Second-Phase Country Level Evaluation (CLE 2) of the Implementation of the Paris Declaration (PD) in the Philippines", *Country Evaluation Report*, Government of the Philippines.
- Cooksy, L. and Caracelli, V. (2005) "Quality, Context, and Use: Issues in Achieving the Goals of Metaevaluation", *American Journal of Evaluation* 26: 31-42.
- Cox, M., et. al. (2011) "Paris Declaration/Hanoi Core Statement Phase 2 Evaluation: Vietnam Country Evaluation", *Country Evaluation Report*, Government of Vietnam.
- Datta, L.E. (2000) "Seriously seeking fairness: Strategies for crafting non-partisan evaluations in a partisan world", *American Journal of Evaluation* 21: 1-14.
- Guion, L.A. (2002) "Triangulation: Establishing the Validity of Qualitative Studies", *FCS6014*, series of the Department of Family, Youth and Community Sciences, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida.
- Health Specialists Limited (2010) "Joint Evaluation of the Paris Declaration Phase 2: Cook Islands Country Evaluation", *Country Evaluation Report*, Cook Islands Government.
- Henry, G. T. and Mark, M. M. (2003). "Beyond use: Understanding evaluation's influence on attitudes and actions", *American Journal of Evaluation* 24: 293-314.
- Hofstetter, C. and Alkin, M. C. (2003) "Evaluation use revisited", in: T. Kellaghan & D. L. Stufflebeam (eds.) *International handbook of educational evaluation*, Dordrecht: Kluwer: 197-222.
- IOB (s.a.) "Beoordelingslijst Programma-evaluaties".
- IODPARC (2009) "Joint Evaluation of the Paris Declaration, Phase 2: Evaluation Framework and Work-plan", OECD/DAC, Paris.
- Jimat Development Consultants (2010) "A Country Evaluation of the Paris Declaration for Malawi Final Report", *Country Evaluation Report*, Ministry of Finance, Government of the Republic of Malawi.
- Jimat Development Consultants (2011) "Phase II Evaluation of the Implementation of the Paris Declaration in Uganda: Final Report", *Country Evaluation Report*, Office of the Prime Minister, Government of the Republic of Uganda.
- Jimat Development Consultants and RuralNet Associates Ltd (2011) "Country Evaluation of the Implementation of the Paris Declaration in Zambia Phase II", *Country Evaluation Report*, Ministry of Finance and National Planning, Republic of Zambia.
- KPMG Mozambique (2010) "Final Report – Paris Declaration Evaluation Phase 2 Mozambique", *Country Evaluation Report*, Ministry of Planning and Development, Government of Mozambique.
- Natural Resources Planners Ltd. (2010) "Evaluation of the Implementation of the Paris Declaration, Phase – II, Country Evaluation Bangladesh", *Country Evaluation Report*, Ministry of Finance of Bangladesh.
- OECD (2008) *2008 Survey on Monitoring the Paris Declaration. Effective Aid by 2010? What it Will Take*, Paris, OECD.
- OECD (2011) *Aid Effectiveness 2005-10: Progress in Implementing the Paris Declaration*, Paris, OECD.



- OECD/DAC (2005) *Paris Declaration on Aid Effectiveness*, Paris, OECD/DAC.
- OECD/DAC (2006) *DAC Evaluation Quality Standards*, Paris, OECD/DAC.
- OECD/DAC (2010) *Quality Standards for Development Evaluation*, Paris, OECD/DAC.
- Patton, M. and Gornick, J. (2011) "Evaluation of the Phase 2 Evaluation of the Paris Declaration. An Independent review of Strengths, Weaknesses and Lessons", Minnesota, Saint Paul.
- Quartey, P. et al., (2010) "Evaluation of the Implementation of the Paris Declaration on Aid Effectiveness: Phase II Ghana Country Report", *Country Evaluation Report*, Government of Ghana.
- Rahardo, B. et al., (2010) "Joint Evaluation of the Paris Declaration, Phase 2 Final Report Indonesia", *Country Evaluation Report*, Government of Indonesia.
- Scriven, M.S. (1969) "An introduction to meta-evaluation", *Educational Product Report 2* (5): 36-38.
- Stevens, C. J. and Dial, M. (Eds.) (1994) "Preventing the misuse of evaluation", *New Directions for Program Evaluation* no. 64. San Francisco, CA: Jossey-Bass.
- Stufflebeam, (2001) "The Metaevaluation Imperative", *American Journal of Evaluation* 2001 (22):183-209.
- Thornton, P. et al., (2010) "Joint Evaluation of the Implementation of the Paris Declaration, Phase II: Nepal Country Evaluation", *Country Evaluation Report*, Ministry of Finance, Nepal.
- United Nations Evaluation Group (2005) "Standards for Evaluation in the UN System".
- Vaai, K. et al., (2010) "Evaluation of the Paris Declaration on Aid Effectiveness and the Accra Agenda for Action: Samoa Country Report", *Country Evaluation Report*, Government of Samoa.
- VBNK and RBMG (2010) "Phase Two Evaluation of the Paris Declaration: Cambodia Country Study Report Final Version", *Country Evaluation Report*, Royal Government of Cambodia.
- Weiss, C. H. (1998) *Evaluation* (2nd ed). Upper Saddle River, NJ: Prentice Hall.
- Wood, B., Kabell, D., Muwanga, N., Sagasti, F. (2008) *Synthesis Report on the First Phase of the Evaluation of the Implementation of the Paris Declaration*, Copenhagen, Ministry of Foreign Affairs of Denmark.
- Wood, B., Betts, J., Etta, F., Gayfer, J., Kabell, D., Ngwira, N., Sagasti, F., Samaranayake, M. (2011) *The Evaluation of the Paris Declaration, Final Report*, Copenhagen, Danish Institute for International Studies.
- WYG International Limited (2011) "Phase Two Evaluation of the Implementation of the Paris Declaration and Accra Agenda for Action in South Africa: Final Country Evaluation Report", *Country Evaluation Report*, Government of South Africa.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage

## ANNEX I: PARIS DECLARATION INDICATORS: BASELINES, TARGETS AND STATUS

Indicator		2005 baseline	2010 actual	2010 target	status
1	<b>Operational Development Strategies</b> % of countries having a national development strategy rated "A" or "B" on a five-point scale	17%	37% (of 76)	75%	Not met
2a	<b>Reliable public financial management (PFM) systems</b> % of countries moving up at least one measure on the PFM/CPIA scale since 2005	--	38% (of 52)	50%	Not met
2b	<b>Reliable procurement systems</b> % of countries moving up at least one measure on the four-point scale since 2005	--	--	no target	--
3	<b>Aid flows are aligned on national priorities</b> % of aid for the government sector reported on the government's budget	42%	41%	85%	Not met
4	<b>Strengthen capacity by co-ordinated support</b> % of technical co-operation implemented through co-ordinated programmes consistent with national development strategies	48%	57%	50%	Met
5a	<b>Use of country PFM systems % of aid for the government sector using partner countries' PFM systems</b>	40%	48%	55%	Not met
5b	<b>Use of country procurement systems % of aid for the government sector using partner countries' procurement systems</b>	39%	44%	no target	--
6	<b>Strengthen capacity by avoiding parallel PIUs</b> Total number of parallel project implementation units (PIUs)	1817	1 158	565	Not met
7	<b>Aid is more predictable</b> % of aid for the government sector disbursed within the fiscal year for which it was scheduled and recorded in government accounting systems	41%	43%	71%	Not met
8	<b>Aid is untied</b> % of aid that is fully untied	75%	86%	more than 89%	Not met
9	<b>Use of common arrangements or procedures</b> % of aid provided in the context of programme-based approaches	43%	45%	66%	Not met
10a	<b>Joint missions</b> % of donor missions to the field undertaken jointly	18%	19%	40%	Not met
10b	<b>Joint country analytic work</b> % of country analytic work undertaken jointly	42%	43%	66%	Not met
11	<b>Results-oriented frameworks</b> % of countries with transparent and monitorable performance assessment frameworks	7%	20% (of 44)	36%	Not met
12	<b>Mutual accountability</b> % of countries with mutual assessment reviews in place	22%	38%	100%	Not met

## ANNEX II: OECD/DAC EVALUATION QUALITY STANDARDS

### 1.0 Rationale, purpose and objectives of an evaluation

- 1.1 The rationale of the evaluation  
*Describes why and for whom the evaluation is undertaken and why it is undertaken at a particular point in time.*
- 1.2 The purpose of the evaluation  
*The evaluation purpose is in line with the learning and accountability function of evaluations. For example the evaluation's purpose may be to:*
- *Contribute to improving an aid policy, procedure or technique*
  - *Consider a continuation or discontinuation of a project/programme*
  - *Account for aid expenditures to stakeholders and tax payers*
- 1.3 The objective of the evaluation  
*The objectives of the evaluation, specify what the evaluation aims to achieve. For example:*
- *To ascertain results (output, outcome, impact) and assess the effectiveness, efficiency and relevance of a specific development intervention;*
  - *To provide findings, conclusions and recommendations with respect to a specific policy, programme etc.*

### 2.0 Evaluation Scope

- 2.1 Scope of the evaluation  
*The scope of the evaluation is clearly defined by specifying the issues covered, funds actually spent, the time period, types of interventions, geographical coverage, target groups, as well as other elements of the development intervention addressed in the evaluation.*
- 2.2 Intervention logic and findings  
*The evaluation report briefly describes and assesses the intervention logic and distinguishes between findings at the different levels: inputs, activities, outcomes and impacts. The report also provides a brief overall assessment of the intervention logic.*
- 2.3 Evaluation criteria  
*The evaluation report applies the five DAC criteria for evaluating development assistance: relevance, efficiency, effectiveness, impact and sustainability. The criteria applied for the given evaluation are defined in unambiguous terms. If a particular criterion is not applied this is explained in the evaluation report, as are any additional criteria applied.*
- 2.4 Evaluation questions  
*The questions asked, as well as any revisions to the original questions, are documented in the report for readers to be able to assess whether the evaluation team has sufficiently assessed them.*

### 3.0 Context

- 3.1 The development and policy context  
*The evaluation report provides a description of the policy context relevant to the development intervention, the development agency's and partners' policy documents, objectives and strategies. The development context may refer to: regional and national economy and levels of development. The policy context may refer to: Poverty reduction strategies, gender equality, environmental protection and human rights.*
- 3.2 The institutional context  
*The evaluation report provides a description of the institutional environment and stakeholder involvement relevant to the development intervention, so that their influence can be identified and assessed.*
- 3.3 The socio-political context  
*The evaluation report describes the socio-political context within which the intervention takes place, and its influence on the outcome and impact of the development intervention.*
- 3.4 Implementation arrangements  
*The evaluation report describes the organisational arrangements established for implementation of the development intervention, including the roles of donors and partners.*

#### 4.0 Evaluation methodology

- 4.1 Explanation of the methodology used  
*The evaluation report describes and explains the evaluation method and process and discusses validity and reliability. It acknowledges any constraints encountered and their impact on the evaluation, including their impact on the independence of the evaluation. It details the methods and techniques used for data and information collection and processing. The choices are justified and limitations and shortcomings are explained.*
- 4.2 Assessment of results  
*Methods for assessment of results are specified. Attribution and contributing/confounding factors should be addressed. If indicators are used as a basis for results assessment these should be SMART (specific, measurable, attainable, relevant and time bound).*
- 4.3 Relevant stakeholders consulted  
*Relevant stakeholders are involved in the evaluation process to identify issues and provide input for the evaluation. Both donors and partners are consulted. The evaluation report indicates the stakeholders consulted, the criteria for their selection and describes stakeholders' participation. If less than the full range of stakeholders was consulted, the methods and reasons for selection of particular stakeholders are described.*
- 4.4 Sampling  
*The evaluation report explains the selection of any sample. Limitations regarding the representativeness of the evaluation sample are identified.*
- 4.5 Evaluation team  
*The composition of evaluation teams should possess a mix of evaluative skills and thematic knowledge, be gender balanced, and include professionals from the countries or regions concerned.*

#### 5.0 Information Sources

- 5.1 Transparency of information sources  
*The evaluation report describes the sources of information used (documentation, respondents, literature etc.) in sufficient detail, so that the adequacy of the information can be assessed. Complete lists of interviewees and documents consulted are included, to the extent that this does not conflict with the privacy and confidentiality of participants.*
- 5.2 Reliability and accuracy of information sources  
*The evaluation cross-validates and critically assesses the information sources used and the validity of the data using a variety of methods and sources of information.*

#### 6.0 Independence

- 6.1 Independence of evaluators vis-a-vis stakeholders  
*The evaluation report indicates the degree of independence of the evaluators from the policy, operations and management function of the commissioning agent, implementers and beneficiaries. Possible conflicts of interest are addressed openly and honestly.*
- 6.2 Free and open evaluation process  
*The evaluation team is able to work freely and without interference. It is assured of cooperation and access to all relevant information. The evaluation report indicates any obstruction which may have impacted on the process of evaluation.*

#### 7.0 Evaluation ethics

- 7.1 Evaluation conducted in a professional and ethical manner  
*The evaluation process shows sensitivity to gender, beliefs, manners and customs of all stakeholders and is undertaken with integrity and honesty. The rights and welfare of participants in the evaluation are protected. Anonymity and confidentiality of individual informants should be protected when requested and/or as required by law.*
- 7.2 Acknowledgement of disagreements within the evaluation team  
*Evaluation team members should have the opportunity to dissociate themselves from particular judgements and recommendations. Any unresolved differences of opinion within the team should be acknowledged in the report.*

## 8.o Quality assurance

- 8.1 Incorporation of stakeholders' comments  
*Stakeholders are given the opportunity to comment on findings, conclusions, recommendations and lessons learned. The evaluation report reflects these comments and acknowledges any substantive disagreements. In disputes about facts that can be verified, the evaluators should investigate and change the draft where necessary. In the case of opinion or interpretation, stakeholders' comments should be reproduced verbatim, such as in an annex, to the extent that this does not conflict with the rights and welfare of participants.*
- 8.2 Quality control  
*Quality control is exercised throughout the evaluation process. Depending on the evaluation's scope and complexity, quality control is carried out either internally or through an external body, peer review, or reference group. Quality controls adhere to the principle of independence of the evaluator.*

## 9.o Relevance of the evaluation results

- 9.1 Formulation of evaluation findings  
*The evaluation findings are relevant to the object being evaluated and the purpose of the evaluation. The results should follow clearly from the evaluation questions and analysis of data, showing a clear line of evidence to support the conclusions. Any discrepancies between the planned and actual implementation of the object being evaluated are explained.*
- 9.2 Evaluation implemented within the allotted time and budget  
*The evaluation is conducted and results are made available in a timely manner in relation to the purpose of the evaluation. Un-envisaged changes to timeframe and budget are explained in the report, any discrepancies between the planned and actual implementation and products of the evaluation are explained.*
- 9.3 Recommendations and lessons learned  
*Recommendations and lessons learned are relevant, targeted to the intended users and actionable within the responsibilities of the users. Recommendations are actionable proposals and lessons learned are generalizations of conclusions applicable for wider use.*
- 9.4 Use of evaluation  
*Evaluation requires an explicit acknowledgement and response from management regarding intended follow-up to the evaluation results. Management will ensure the systematic dissemination, storage and management of the output from the evaluation to ensure easy accessibility and to maximise the benefits of the evaluation's findings.*

## 10.o Completeness

- 10.1 Evaluation questions answered by conclusions  
*The evaluation report answers all the questions and information needs detailed in the scope of the evaluation. Where this is not possible, reasons and explanations are provided.*
- 10.2 Clarity of analysis  
*The analysis is structured with a logical flow, Data and information are presented, analysed and interpreted systematically. Findings and conclusions are clearly identified and flow logically from the analysis of the data and information. Underlying assumptions are made explicit and taken into account.*
- 10.3 Distinction between conclusions, recommendations and lessons learned  
*Evaluation reports must distinguish clearly between findings, conclusions and recommendations. The evaluation presents conclusions, recommendations and lessons learned separately and with a clear logical distinction between them. Conclusions are substantiated by findings and analysis. Recommendations and lessons learned follow logically from the conclusions.*
- 10.4 Clarity and representativeness of the summary  
*The evaluation report contains an executive summary. The summary provides an overview of the report, highlighting the main conclusions, recommendations and lessons learned.*

### ANNEX III: TALLIED SCORES OF THE 15 COUNTRY EVALUATION REPORTS ON THEIR COMPLIANCE WITH THE EVALUATION FRAMEWORK (AGGREGATED BY SUB-QUESTIONS)

	No. of components	Weak	Partially Satisfactory	Satisfactory	Excellent	Index	Ranking
<b>Core Q1: What are the important factors that have affected the relevance and implementation of the Paris Declaration and its potential effects on aid effectiveness and development results? (The Paris Declaration in context)</b>							
What are the key characteristics of the country that have been most relevant to the implementation of the Paris Declaration? (Ensuring analytical not descriptive treatment)	6	14	13	22	41	3.00	3
What are the most important national and international events that have affected [in the country] the implementation of the Paris Declaration and the Accra Agenda for Action priorities, and how?	1	0	2	7	6	3.27	2
What is the place of aid subject to Paris Declaration principles among all sources of development finance and resources? What have been the trends from early roots to 2005 and since?	4	7	19	19	15	2.70	4
Which are the key actors, in the country and among its development partners, who take major decisions on aid? What influence do the Paris Declaration and Accra Agenda for Action commitments have on them, in relation to their other priorities and incentives?	7	36	12	28	29	2.48	5
To what extent and where have the Paris Declaration principles been implemented?	2	3	4	4	19	3.30	1
<b>Core Q2: To what extent and how has the implementation of the Paris Declaration led to an improvement in the efficiency of aid delivery, the management and use of aid and better partnerships? (Process and intermediate outcomes)</b>							
Country ownership over development	3	8	12	17	8	2.56	3
Building more inclusive and effective partnerships for development	6	6	29	34	21	3.13	1
Delivering and accounting for development results	2	4	7	11	8	2.77	2
<b>Core Q3: Has the implementation of Paris Declaration strengthened the contribution of aid to sustainable development results? How? (Development outcomes)</b>							
Were results in specific sectors enhanced through the application of the Paris Declaration principles?"	2	3	7	7	13	3.17	1
Did the implementation of the Paris Declaration help countries to improve the prioritisation of the needs [beyond income poverty] of the poorest people, including women and girls?	2	6	3	9	12	3.03	3

	No. of components	Weak	Partially Satisfactory	Satisfactory	Excellent	Index	Ranking
Has Paris Declaration implementation led to sustainable increases in institutional capacities and social capital at all levels to respond to development challenges? Why, how and where, and what are the effects?	3	7	9	9	20	2.93	2
How and why has the mix of aid modalities (including general or sector-specific budget support) evolved and what has been learnt on the development results?	2	6	5	6	13	2.87	4
<b>Framework for Overall Conclusions</b>							
What has been the relevance of the Paris Declaration and the ways it has been implemented to the challenges of aid effectiveness?	-	5	1	1	8	2.80	1
To what extent has each of the five principles of the Paris Declaration been observed and implemented, and the Accra Agenda priorities reflected? Why? Have there been conflicts or trade-offs between them?	-	5	1	1	8	2.80	1
What has the Paris Declaration achieved for aid effectiveness and development results? How significant are these contributions? How sustainable? Is there evidence of better ways to make aid more effective and contribute more to development results, for women and men and for those who are excluded?	-	5	3	2	5	2.47	6
What effects has the implementation of the Declaration had on the respective burdens of aid management falling on the partner country and its respective donors, relative to the changing volumes and quality of aid and of the aid partnership itself? Are these effects likely to be transitional or long term?	-	6	1	2	6	2.53	5
What has been the added value of Paris Declaration-style development cooperation compared with the pre- Paris Declaration situation, and seen alongside other drivers of development in the country, other sources of development finance and development cooperation partners beyond those so far endorsing the Declaration?	-	7	2	1	5	2.27	7
What are the key messages for a) national stakeholders, and b) donor countries and agencies?	-	6	0	0	9	2.80	1
What are the key implications for aid effectiveness in the future taking account of new challenges and opportunities (e.g. climate change) and new actors and relationships?	-	6	1	1	7	2.60	4

## ANNEX IV: STRENGTHS OF EVIDENCE FOR INTENDED OUTCOMES UNDER CORE QUESTION TWO.

Intended outcome	Strengths of evidence
<b>Country ownership of development</b>	
I. Stronger national development strategies and operational frameworks:	Good
i. National strategic	
ii. Detailed operational	
II. Increased alignment of aid with partner country:	Adequate
i. Priorities, systems and procedures	
ii. Building of capacity in systems	Good
III. Defined measures and standards of performance and accountability in country systems	Good
<b>Building more inclusive and effective partnerships for development</b>	
IV. Less duplication of efforts and rationalised more cost-effective donor activities	Good
V. Reformed and simplified donor policies and procedures, more collaborative behaviour	Good
VIa. More predictable and multi-year commitments on aid flows	Good
VIb. More shared conditionalities (Accra commitment, para. 25)	Poor
VII. Sufficient delegation of authority and incentives to donors' field staff for effective partnership working	Good
VIII. Sufficient integration of global programmes and initiatives into partner countries broader development agendas	Adequate
IX. Stronger partner countries institutional capacities to plan, manage and implement results-driven national strategies	Adequate
<b>Delivering and accounting for development results</b>	
Xa. Enhanced donors and partner countries' respective accountability to their citizens and parliaments	Good (donors) Adequate (partner countries)
Xb.	
i. Enhanced transparency for development results	Good
ii. Structured arrangements for mutual accountability	Adequate
XI. Less corruption and more transparency; strengthening public support and effective resource mobilisation and allocation	Adequate (donors) Poor (partner countries)



