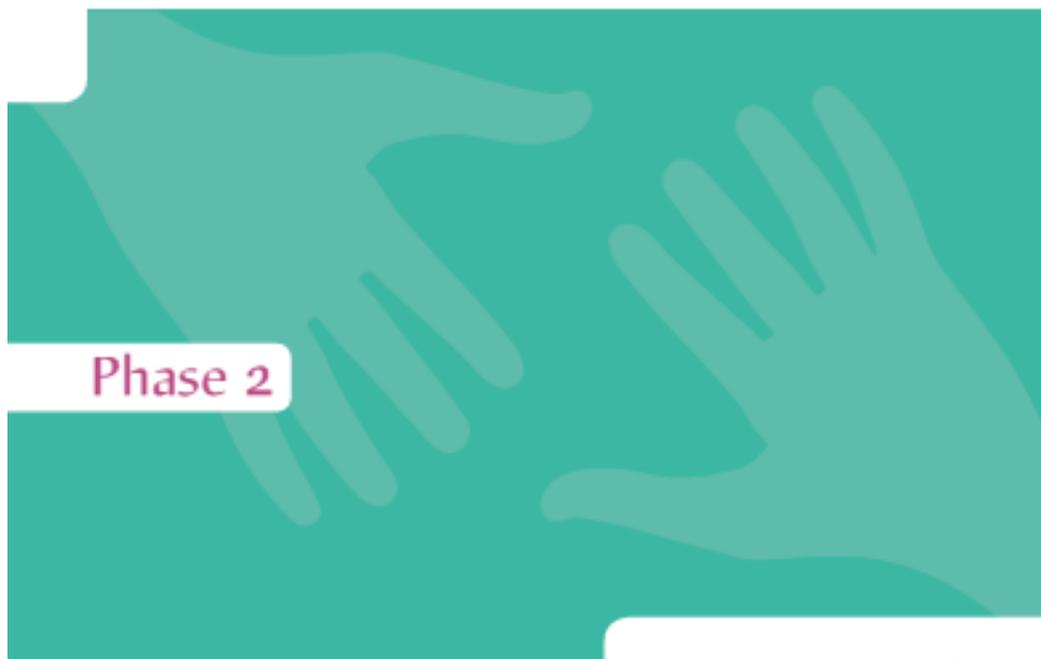


**EVALUATION of the PHASE 2 EVALUATION  
of the PARIS DECLARATION**

**An Independent review of Strengths, Weaknesses, and Lessons**



**Michael Quinn Patton  
Jean Gornick  
Utilization-Focused Evaluation  
Saint Paul, Minnesota, USA**

**July, 2011**

# EVALUATION of the PHASE 2 EVALUATION of the PARIS DECLARATION

## An Independent review of Strengths, Weaknesses, and Lessons

Michael Quinn Patton  
Jean Gornick  
Utilization-Focused Evaluation  
Saint Paul, Minnesota, USA

July, 2011

### Contents

<b>Executive Summary</b> .....	i-iv
Introduction to the Independent Review .....	1
<b>Strengths, Weaknesses, and Lessons</b>	
Major Strengths.....	2
Major Weaknesses .....	7
Lessons .....	10
<b>Evidence for Conclusions and In-depth Analysis</b>	
Gathering the Perspectives of Those Involved .....	16
Country-Level Evaluation Process Issues.....	18
Quality of Evidence .....	23
Quality of Evidence Patterns and Themes .....	27
Final Synthesis Report Data Quality .....	28
Utility, Importance, Quality and Credibility of Evidence .....	32
<b>Conclusion on Quality of Evidence</b> .....	35
Evaluator Independence and Evaluator Collaboration.....	36
Evaluation – Stakeholder Relationships (graphic) .....	37
Survey Data On Collaboration and Independence .....	38
<b>Concluding observations about collaboration and independence</b> .....	41
Time Pressure and Meeting Deadlines .....	44
General vs. Particular: Aggregating and Synthesizing While Taking Context Seriously...	50
Top-Down vs. Bottoms-Up Evaluation Processes (graphic) .....	51
Capacity Issues .....	56
Impacts of Being Involved in the Paris Declaration Evaluation Process.....	60
Use .....	62
Expected Use of the Final Synthesis Report.....	64
The Paris Declaration Evaluation as a Model .....	65
Independent Audit Statement from this Evaluation of the Evaluation.....	66
Appendices (additional data from the surveys) .....	67

## Executive Summary

Given the importance of the *Phase 2 Evaluation of the Paris Declaration*,<sup>1</sup> the Evaluation Management Group commissioned this independent assessment of the evaluation. It has become a standard in major high-stakes evaluations of this kind to commission an independent review to determine whether the evaluation meets generally accepted international standards of quality. That is what this report does.

Our evaluation of the Paris Declaration Evaluation included reviewing data collection instruments, templates, guidance documents, and processes; reviewing the partner country and donor evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group (IRG) where the evidence was examined and the Final report conclusions revised accordingly; engaging IRG participants in a reflective practice session on lessons learned; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. This evaluation of the evaluation includes assessing both the report's findings and the technical appendix that details how findings were generated.

We conclude that the findings, conclusions, and recommendations presented in the *Phase 2 Evaluation of the Paris Declaration* adhere closely and rigorously to the evaluation evidence collected and synthesized. Obtaining high quality evidence and thoughtfully analyzing that evidence was the constant theme of the evaluation. Both strengths and weaknesses in the evaluation are appropriately acknowledged. The comprehensive Technical Annex accurately describes data collection and analysis approaches. Partner country and donor evaluation reports, upon which the Synthesis Evaluation is based, were openly and transparently shared with the International Reference Group to allow peer review and make visible both strengths and limitations in those reports. Partner country reports were screened for adherence to quality standards with particular attention to the strength of evidence to support conclusions reached.

Those countries and donors that undertook this voluntary Phase 2 evaluation of the implementation of the Paris Declaration have engaged in systematic and in-depth reflection and evidence-based processes that make their conclusions and insights worthy of serious attention. The Synthesis Report accurately captures those evidence-based conclusions and insights. The conclusions drawn and recommendations made are appropriate for and derived from the evidence analyzed and synthesized.

This evaluation of the evaluation identifies strengths, weaknesses, and lessons for future such efforts. Each of these are discussed and documented in this report.

### Major Strengths

1. Focusing attention on and inquiring deeply into the Paris Declaration and its effects.
2. Expanding knowledge of and engagement with the Paris Declaration principles.
3. Being utilization-focused throughout.
4. Keeping quality of evidence at the center of the evaluation

---

<sup>1</sup> *Phase 2 Evaluation of the Paris Declaration*

[http://www.oecd.org/document/60/0,3343,en\\_21571361\\_34047972\\_38242748\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/60/0,3343,en_21571361_34047972_38242748_1_1_1_1,00.html)

5. Establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout.
6. Making the evaluation a genuinely and authentically joint endeavor based on and adhering to the principles of the Paris Declaration.
7. Creating a standardized data gathering, assessment, and reporting framework
8. Basing the Phase 2 evaluation on the findings and lessons of Phase 1
9. Leading the evaluation effort based on vision, principles, standards, experience, competence, and commitment.
10. Assuring evaluation of the evaluation

### Major Weaknesses

1. Variations in evidence quality
2. The self-selected sample of partner countries and donors
3. Variation in use of and adherence to the standardized evaluation operational matrix for partner country evaluations
4. Lack of standardized and defined criteria for rating judgments
5. Not all Phase 2 donor studies followed the revised matrix.
6. Variations in length and degree of partner country engagement.
7. Capacity variations within partner countries and greater need for capacity support than anticipated

### Significant Lessons

1. Assuring evaluator independence and genuine stakeholder involvement can be mutually reinforcing in support of enhanced evaluation quality and credibility. Stakeholder involvement ensures the relevance of evaluations and evaluator independence ensures credibility. But stakeholder involvement is sometimes seen as undermining independence while processes to assure independence often limit stakeholder involvement. The Evaluation of the Paris Declaration has demonstrated how to strike the appropriate balance by creating structures and processes that provide meaningful and authentic stakeholder involvement without impinging on evaluator independence. Page 36 presents a graphic depiction of the balance that needs to be struck in evaluation-stakeholder relationships to ensure both high quality stakeholder involvement and evaluator independence that supports credibility of findings and conclusions.
2. The DAC quality standards for evaluation provide relevant, appropriate and useful guidance for conducting a complex, strategic, and principles-focused evaluation. The Final Report includes a detailed analysis of how the DAC Standards informed the evaluation.
3. The analytical, interpretative and reporting framework for assessing progress on major Paris declaration and Accra Action Agenda outcomes -- **direction, pace and distance travelled** -- proved useful in synthesizing and comparing findings, conclusions, and judgments. It is a unique and creative evaluation framework, well-suited to the challenge of synthesizing findings on progress across 21 partner country case studies
4. A comprehensive and complex international evaluation that involves multiple countries in a genuinely *joint evaluation* endeavor will need to anticipate and provide significant in-

country capacity support and development to help overcome inevitable variations in evaluation capacity. Capacity has to be built and supported for some countries throughout the evaluation.

5. An evaluation design that is “good enough,” even though not ideal, will provide credible and useful findings. The question is not whether an evaluation design is ideal, but whether it is adequate, given the constraints and challenges faced in mounting such an enormous effort. Invoking the “good enough rule,” we find that the evaluation *satisfices*. Modest claims and conclusions can be supported by relatively modest data. Strong claims and conclusions require strong supporting evidence. The Phase 2 Evaluation of Paris Declaration draws conclusions and makes recommendations appropriate to the quality of evidence generated and available.
6. Engagement of diverse international stakeholders in a global evaluation process requires careful planning, structure, execution, and facilitation. Those who would convene large and diverse stakeholder groups to advise on evaluation design, data collection, analysis, and reporting would do well to study the structure and processes of the Paris Declaration International Reference Group, especially with regard to the first lesson above, finding the appropriate balance between stakeholder involvement and ensuring evaluator independence. The joint, collaborative, and participatory nature of the evaluation meant that both bottoms-up, collaborative processes and top-down, guidance and coordination processes had to be managed. A graphic depiction of the tensions in top-down versus bottoms-up evaluation processes is presented on page 50.
7. Evaluation findings, conclusions, and recommendations are more accessible, understandable, and potentially useful for informing policy deliberations and decision-making when stated clearly, and jargon is avoided, including but not limited to evaluation jargon.
8. Evaluation findings and conclusions are more likely to invite and inform deliberation when core evaluation questions are framed as descriptive, open-ended inquiries rather than asking for simplistic judgments. It would be premature to render a summative judgment about the Paris Declaration. The evaluation design, the quality of data, and, most importantly, the complexity of development aid, do not support a summative judgment. The lesson here is to pay careful attention to the realistic and plausible purpose an evaluation can fulfill and make sure that the framing of questions are answerable and appropriate.
9. The value of an evaluation lies not just in the findings, but in the impacts that emerge among those who participate in the evaluation as they engage in the inquiry. The evaluation process increased awareness of and attention to the Paris Declaration and aid reform well beyond what would have occurred without the evaluation.
10. The joint nature and full transparency of the Evaluation of the Paris Declaration makes it a model for international evaluation. The evaluation adhered to the Paris Declaration principles even as it evaluated the implementation and results of the Paris Declaration. That is what makes it a model of international evaluation excellence.

11. Use of findings, conclusions, and recommendations can begin immediately when a high-stakes evaluation is appropriately conducted, deadlines are met, and results are disseminated widely and immediately. The Secretariat, the Evaluation Management Group, the International Reference Group, and the Core Evaluation Team attended to use from the beginning and throughout all aspects of the evaluation. It is already clear that the report is garnering widespread international attention and being used in preparations for the 4th High Level Forum on Aid Effectiveness to be held at the end of 2011.
12. Evaluation of an evaluation (meta-evaluation) can, and should be undertaken during the evaluation, not just afterward. Meta-evaluations are often undertaken only after an evaluation is completed, as a post hoc, retrospective activity. This evaluation of the evaluation was commissioned halfway through the evaluation process in time for us to observe the direction, pace, and distance travelled by the evaluation as it concluded. This made it possible to interview key participants, survey those involved, and document how the evaluation was actually conducted and the final report constructed, reviewed, and finalized. The commitment to and timing of the meta-evaluation has been exemplary.

### **An Independent Audit of the Synthesis Evaluation**

The Final Report of the *Phase 2 Evaluation of the Paris Declaration* includes, as part of the preface, our independent audit statement of the Synthesis Evaluation. That statement is at the end of this report on page 65. That statement notes that prior to undertaking this review, we had no prior relationship with any members of the Evaluation Management Group, the Secretariat, or the Core Evaluation Team. We had complete and unfettered access to any and all evaluation documents and data, and to all members of the International Reference Group, the Evaluation Management Group, the Secretariat, and the Core Evaluation Team.

Both the audit statement and this full report on which the audit statement was based conclude that, in our opinion, the Final Synthesis Report can be trusted as independent, evidence-based, and adhering to international standards for quality evaluation. Notwithstanding inevitable limitations inherent in such a complex and comprehensive evaluation initiative, and fully disclosed in the report and discussed herein, the findings of the *Phase 2 Evaluation of the Paris Declaration* can be studied and used as trustworthy and credible.

# **EVALUATION of the EVALUATION of the PARIS DECLARATION**

An Independent Review of Strengths, Weaknesses, and Lessons

Michael Quinn Patton  
Jean Gornick  
Utilization-Focused Evaluation  
Saint Paul, Minnesota, USA

## **Introduction**

It has become a standard in major high-stakes evaluations to commission an independent review to determine whether the evaluation meets generally accepted standards of quality, and, in so doing, to identify strengths, weaknesses, and lessons. The Phase 2 Evaluation of the Paris Declaration certainly qualifies as a high-stakes evaluation. As the Report states:

The Evaluation is important both for accountability and to point the way for future improvements. The underlying stakes are huge: better lives for billions of people (reflected in the approaching Millennium Development Goals for 2015); hundreds of billions of dollars expended; vital international relationships; and growing demands to see results from development aid. (p. xii)

The scope of the Evaluation is immense. It has involved more than 50 studies in 22 partner countries, reviews in 18 donor agencies, as well as several studies on special themes. It has taken place over four years, in two phases between 2007 and 2011. The current report focuses on Phase 2, the period since the Accra High Level Forum held in September, 2008. The evaluation, officially launched in September, 2009, is targeted to the next High Level Forum on aid effectiveness to be held in Busan, Korea, in December, 2011.

Given the importance of the Evaluation of the Paris Declaration, the Management Group commissioned this independent assessment of the evaluation. Prior to undertaking this review, we had no prior relationship with any members of the Management Group or the Core Evaluation Team. We had complete and unfettered access to any and all evaluation documents and data, and to all members of the International Reference Group, the Management group, and the Core Evaluation Team. Our evaluation of the evaluation included reviewing data collection instruments, templates, and processes; reviewing the partner country and donor evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group (IRG) where the evidence was examined and the conclusions refined and sharpened accordingly; engaging IRG participants in a reflective practice lessons learned session; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. This evaluation of the evaluation includes assessing both the report's findings and the technical appendix that details how findings were generated.

We begin with a summary of major strengths, weaknesses and lessons. The remainder of the report provides the evidence for these conclusions.

## STRENGTHS, WEAKNESSES, and LESSONS

Our evaluation of the Paris Declaration Evaluation included reviewing data collection instruments, templates, guidance documents, and processes; reviewing the partner country and donor evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group (IRG) where the evidence was examined and the Final report conclusions revised accordingly; engaging IRG participants in a reflective practice session on lessons learned; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. Drawing on all of these data sources, we open this evaluation of the evaluation with a summary of major strengths, weaknesses, and lessons.

### Major Strengths

11. *Focusing attention on and inquiring deeply into the Paris Declaration and its effects.*
12. *Expanding knowledge of and engagement with the Paris Declaration principles.*
13. *Being utilization-focused throughout.*
14. *Keeping quality of evidence at the center of the evaluation*
15. *Establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout.*
16. *Making the evaluation a genuinely and authentically joint endeavor based on and adhering to the principles of the Paris Declaration.*
17. *Creating a standardized data gathering, assessment, and reporting framework*
18. *Basing the Phase 2 evaluation on the findings and lessons of Phase 1*
19. *Leading the evaluation effort based on vision, principles, standards, experience, competence, and commitment.*
20. *Assuring evaluation of the evaluation*

### Discussion of major strengths

1. *Focusing attention on and inquiring deeply into the Paris Declaration and its effects.*

For two years the evaluation engaged a dedicated, knowledgeable, and diverse group of people in thinking about, gathering data on, and interpreting the effects of the Paris Declaration and Accra Action Agenda. A comprehensive and in-depth evaluation process involves defining terms, clarifying and focusing priority questions, operationalizing key concepts, formalizing the theory of change, establishing boundaries around the inquiry, building partnerships, constructing data collection protocols, building capacity to gather and analyze data, facilitating relations, and focusing attention on what is being evaluated: the implementation of the Paris Declaration and its effects on development aid. The final product, *The Evaluation of the Paris Declaration Phase 2 Final Report*, is the culmination of all this effort and will, appropriately, be the focus of attention going forward. But behind that report, both making it possible and extending its impact, is a large number of people around the world who have been deeply engaged in thinking about these issues and bringing evaluative thinking and methods to bear on critical questions of effectiveness, efficiency, and impact. It seems fair to suggest that the large and diverse group of people engaged in the many activities and distinct studies that, together, make up the whole of the evaluation, have thought as

deeply about these issues as any group anywhere in the world. An evaluation process is not just talking about something. It means studying the thing being evaluated. At its best, as in this case, it involves deep intellectual and rigorous methodological inquiry. Thus, a major strength and accomplishment of this two-year inquiry has been the very process of focusing in-depth attention on and inquiry into the Paris Declaration.

### *2. Expanding knowledge of and engagement with the Paris Declaration principles.*

This second strength follows directly from the first. The first strength highlights the in-depth knowledge and understanding created among those engaged in the evaluation. This second strength highlights the fact that the evaluation process increased awareness of and engagement with the Paris Declaration, as well as issues of aid effectiveness, well beyond what would have occurred without the evaluation. Policy makers, ministers, civil servants, agency directors and staff, NGOs, and others who were interviewed as part of country and donor studies became more aware of and thought more deeply about the Paris Declaration as a result of being interviewed. The evaluation, therefore, served a dissemination function even as the inquiry gathered data about the degree of knowledge about and engagement with the Paris Declaration. Members of national reference groups, country evaluation teams, and donor study groups have all reported that the evaluation contributed to greater knowledge about and, in many cases, greater commitment to and implementation of the Paris Declaration. This has come to be known in the evaluation literature as “process use,” which captures the fact that evaluation inquiries have an impact on the focus of inquiry quite apart from and well before the findings are reported.

### *3. Being utilization-focused throughout.*

The evaluation was designed and implemented to contribute to the forthcoming 4<sup>th</sup> High Level Forum on Aid Effectiveness, which is to meet in Korea at the end of 2011. The delegates to that High Level Forum, and those preparing and supporting the delegates, are the primary intended users of the evaluation. The evaluation questions focus on the priority issues that the High Level Forum is expected to discuss. The timing of the evaluation, including very tight deadlines that were rigorously adhered to and met, were based on when the evaluation report had to be produced to contribute to the deliberations of the High Level Forum. The recommendations are differentiated for and aimed at specific intended users. Three recommendations are aimed at policymakers in partner countries; three are directed specifically to policymakers in donor countries; and five are offered to policymakers in both partner countries and donor countries and agencies. This exemplifies being utilization-focused.

### *4. Keeping quality of evidence at the center of the evaluation*

Every aspect of the evaluation was **evidence-driven**. This is clear in the supporting documents and protocols produced, from the inception report, through the standardized country data collection matrix, to the template for country reports, and the interactions at the International Reference Group meetings. The focus on data quality and evidence stemmed in part from recognition that the evaluation questions were complex, multiple and diverse methods would be needed, data collection challenges would be enormous, timelines would be extremely tight, and triangulation would be essential. While the evaluation matrix required rendering judgments about the extent to which progress was being made on various outcomes, the constant message was that those judgments had to be supported by evidence. In the end, the quality of evidence varied across questions and

countries. The Final Report is open and transparent in acknowledging those variations in quality of evidence and this evaluation of the evaluation discusses those variations at some length. Thus, those weaknesses in the data that are reported and acknowledged are a result of capacity, resource, and time constraints, inherent measurement and methodological challenges, and the complexity and comprehensiveness of the entire evaluation initiative, but not to any lack of attention to the centrality of having high quality and credible evidence as the basis for evaluation conclusions, judgments, and recommendations.

*5. Establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout.*

Independence is essential for credibility, and credibility is essential for use. The Paris Declaration Evaluation was established under the auspices of a free-standing and independent Secretariat. Thus, it was not subservient to or dependent on any existing organization, agency, or bureaucracy. The Secretariat has cooperated with the DAC Working Party on Aid Effectiveness (WP-EFF)<sup>2</sup>, but is independent of that coalition. 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 were all established to assure the evaluation's independence. The processes of data collection, analysis, interpretation, and stakeholder review all maintained and supported the independence of the evaluation. The Core Evaluation Team had final autonomy and ultimate ownership in determining the evaluation findings, conclusions, and recommendations. This report discusses the issue of independence, and the evidence supporting this conclusion, at length.

*6. Making the evaluation a genuinely and authentically joint endeavor based on and adhering to the principles of the Paris Declaration.*

While the governing arrangements were aimed at ensuring the evaluation's independence, they were also geared toward making the evaluation a truly joint exercise. The evaluation process was participatory and consultative among partner countries, donors, and international organization participants. Country ownership by partner countries was made a priority to ensure full participation and engagement. Country representatives participated in designing the synthesis matrix. Quality control was based on peer reviews and international participant feedback. Providing sufficient support to make the evaluation a trilingual exercise -- English, French and Spanish -- was aimed at ensuring full participation in and access to all aspects of the evaluation. In-country evaluation team procurement was the responsibility of and followed partner country processes. Indeed, from the beginning, the evaluation was conceptualized, designed, and implemented to adhere to Paris Declaration principles. This was not always easy and resulted in administrative challenges and, in some cases, delays in implementation, but the evidence we gathered supports the conclusion that the evaluation was a genuine and authentic joint endeavor.

---

<sup>2</sup> Working Party on Aid Effectiveness:

[http://www.oecd.org/document/35/0,3343,en\\_2649\\_3236398\\_43382307\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/35/0,3343,en_2649_3236398_43382307_1_1_1_1,00.html)

### 7. *Creating a standardized data gathering, assessment, and reporting framework.*

The Final Report is a synthesis of 22 partner country evaluations (including one that only participated in Phase 1), 18 donor and multilateral development agency studies<sup>3</sup>, thematic studies, sector studies, supplementary studies, and secondary and supporting data and studies. The terms of reference for the Core Evaluation Team charged them with creating a standardized template to guide the synthesis. Early on, at the first International Reference Group meeting, the metaphor emerged of treating the implementation of Paris Declaration changes as a journey, which led to focusing the synthesis framework on three critical dimensions of that journey for evaluating the Paris Declaration: *direction of travel on each key intended outcome, and the pace and distance travelled so far*. Multiple sources of evidence were gathered to support the judgments rendered, mainly qualitative evidence but also sound quantitative data where available. Aggregating and synthesizing the massive amounts of data from case studies and secondary sources was greatly facilitated and given coherence by this simple but powerful template: **direction of travel, distance, and pace**.

Moreover, the operational matrix created jointly by country participants and the core evaluation team proved especially useful in providing coherence to the evaluation.

- In our survey of national reference group and partner country evaluation team members, 87% rated the standardized matrix as “very” or “somewhat useful.”
- In our survey of International Reference Group members, 96% of donors and international organization representatives rated the standardized matrix as very or somewhat important.

### 8. *Basing the Phase 2 evaluation on the findings and lessons of Phase 1*

A major and important strength of the Phase 2 evaluation is that the effort was not starting from scratch. Although the Phase 1 evaluation was considerably more modest in scope and much narrower in focus than Phase 2, the first phase provided a useful foundation for the second phase evaluation. The Phase 1 evaluation highlighted the centrality of context and emphasized that processes for improving aid effectiveness are political and not just technical. Phase 1 also provided critical methodological lessons that guided the Phase 2 design, for example, the emphasis on sound evidence for comparative analysis via a clear central framework (see strength #7 above). In short, the Phase 1 evaluation offered recommendations that guided the design of Phase 2. The continuity of leadership and key evaluation team members proved to be a huge asset in completing the much more comprehensive and complex Phase 2 evaluation as a joint initiative involving many more partner countries and donors.

---

<sup>3</sup> The World Bank and European Commission were notable by their absence, a fact often commented on, negatively, by those who did participate.

*9. Leading the evaluation effort based on vision, principles, standards, experience, competence, and commitment.*

People matter. A substantial body of research on evaluation credibility and use points to the importance of what has been called “the personal factor,” which points to the interest, commitment, engagement, and leadership of key individuals as essential to conducting high quality evaluations. Much of evaluation focuses on procedures, protocols, measures, methods, and analytical techniques. But none of that happens without people. The Paris Declaration Secretariat brought vision and experienced leadership to this enterprise, without which it would not have happened. Assembling resources, garnering the support of key institutions, using networks to negotiate critical agreements, and finding competent people to manage and implement the evaluation are leadership functions. It was at this level of visionary leadership that key decisions were made and commitments engendered to base the Evaluation of the Paris Declaration on the principles of the Paris Declaration, to adhere throughout to the DAC standards for quality evaluations, and to take on such a complex, controversial, and challenging initiative in the face of many, many naysayers who loudly declared that this could not be done. Having studied for years the factors that distinguish high quality, useful evaluations from those that are neither, we would be remiss to not call attention to the great extent to which the Paris Declaration is an exemplar of the critical role played by committed and effective leadership at all levels of this evaluation. Being privy to some of the background stories of things that had to be negotiated, problems that had to be solved, resources that had to be found, people who had to be engaged, crises that had to be handled, and challenges that had to be met, it is clear that a successful evaluation is about much, much more than systematic procedures, rigorous methods, careful analysis, and thoughtful, evidence-based judgments. The personal factor, exemplified in effective leadership, is at the core of evaluation success.

*10. Assuring evaluation of the evaluation*

It has become a standard in major high-stakes evaluations of this kind to commission an independent review to determine whether the evaluation meets generally accepted international standards for quality. This evaluation of the evaluation is one example of taking that commitment seriously. But the Final Report was also peer reviewed for quality, strategic and policy relevance, and communicative power by distinguished international experts (see p. v of the Final report Preface). The emerging findings of the report were thoroughly reviewed at the 3<sup>rd</sup> International Reference Group meeting in December, 2010, and the draft final report’s findings, conclusions and recommendations were systematically reviewed by 4th International Reference Group meeting in April, 2011. The Evaluation Management Group provided feedback on the evaluation processes and report in playing a quality assurance role throughout. Evaluation of the evaluation for quality enhancement, clarity of communications, and policy relevance was built into the concluding stages of the evaluation. Seriously and systematically assuring evaluation of the evaluation, with diverse stakeholder involvement, has been one of the strengths of the Paris Declaration evaluation process.

This concludes the discussion of major strengths. (See page 2 for a summary of the 10 strengths.) The next page begins the discussion of major weakness.

### **Major Weaknesses**

8. *Variations in evidence quality*
9. *The self-selected sample of partner countries and donors*
10. *Variation in use of and adherence to the standardized evaluation operational matrix for partner country evaluations*
11. *Lack of standardized and defined criteria for rating judgments*
12. *Not all Phase 2 donor studies followed the revised matrix.*
13. *Variations in length and degree of partner country engagement.*
14. *Capacity variations within partner countries and greater need for capacity support than anticipated*

### Discussion of Major Weaknesses

#### 1. *Variations in evidence quality*

One of the strengths noted in the preceding section was *keeping quality of evidence at the center of the evaluation* (# 4). Despite that emphasis, the quality of evidence varies across partner countries, donor studies, and different evaluation questions in the synthesis. The Final Report is open and transparent in acknowledging these variations in quality of evidence. It provides “strength of evidence ratings” by the Core Evaluation Team for 17 major intended outcomes of the Paris Declaration: none are rated very good; 9 are rated good; 6 adequate; and 2 are rated poor (Final Report, pp. 18-19). This evaluation of the evaluation reports strength of evidence ratings from partner country participants and members of the International Reference Group.

#### Strongest evidence ratings

- The country’s ownership over development
- Overall implementation of the principles of the Paris Declaration.
- Improvements in the efficiency of aid delivery.
- Better (more inclusive and effective) partnerships.
- Overall improvements in the management and use of aid.

#### Weakest evidence ratings

- Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.
- Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.
- Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.
- The implications for aid effectiveness in the future.
- Delivering and accounting for development results.
- Evolution in the mix of aid modalities.

## 2. *The self-selected sample of partner countries and donors*

The voluntary nature of participation in the evaluation resulted in 21 partner countries and 7 donors doing Phase 2 case studies. The partner country sample is certainly diverse, from very large and populous countries (e.g., Indonesia and Bangladesh) to quite small countries (e.g., Cook Islands and Samoa), and from higher income countries (e.g., Colombia and the Philippines) to low income countries (e.g., Benin and Malawi), and more stable contexts (e.g., Ghana and Vietnam) to fragile states (Afghanistan). There are 10 African countries, 7 from Asia, and only two from South America. “Two proposed country evaluations (in Kyrgyzstan and Sri Lanka) were cancelled because of political change and uncertainty. Considerable effort was made to expand the number of countries in Latin and Central America but Bolivia and Colombia were the only eventual participants” (Annex 5, p. 198, Final Report). In the end, the sample included a diverse range of partner countries that could be used to extrapolate patterns and trends but is neither representative nor generalizable in a classic statistical sense. The donor sample is even smaller. Particularly noteworthy by their importance *and absence* were The World Bank and European Commission.

It is worth noting that the sample reflects the inevitably political nature of the evaluation and the decision to adhere to Paris Declaration principles in honoring the decision of partner countries about whether to participate in the evaluation process. As one Evaluation Team member commented:

Just as partners signed up to the Paris Declaration on a voluntary basis, so they signed up for the evaluation on the same terms. If the evaluation had been a purely technocratic exercise, the Evaluation Management Group could have sent in highly qualified international teams to select a completely robust sample base. But, in fact, the sample is a reflection of the political nature of process. This is important.

## 3. *Variation in use of and adherence to the standardized evaluation operational matrix for partner country evaluations*

The Final Report acknowledges that “most of the Phase 2 evaluations – the central source of evidence – did not consistently use any standard set of data sources or always apply the comparable rating scales suggested in the Evaluation Matrix. Instead, they selected from and supplemented the range of sources and ratings proposed to capture the particularities of each country situation” (p. 7). In our survey of partner country participants, 20% of evaluation team leaders and national coordinators acknowledged that they struggled to cover the full scope of questions in the matrix. This was particularly true for those partner countries that assembled their evaluation teams late and missed the opportunity to participate in capacity-building workshops on the matrix offered by the core evaluation team.

## 4. *Lack of standardized and defined criteria for rating judgments*

The operational matrix provided scales for rating judgments. For example, each country evaluation team would be asked to assess results using a scale like Very significant change, Quite significant change, Limited change, Very limited change, or No change. Actual scales varied by questions. The most common scale used for rating progress was:

- *Substantial progress*
- *Some progress*
- *Little progress*
- *None*
- *Regression*

Each country team then had to determine its own meaning for the points on the rating scale. Furthermore, teams were instructed to apply their own weighting across one or more indicators when reaching summary judgments on progress for a core question section. This is a reasonable approach to assure context sensitivity by providing standard scale terminology but allowing each team to operationalize and interpret the scales in ways that are meaningful in their context. However, this approach invites criticism from measurement specialists that the aggregation and synthesis of scales that mean different things in different contexts essentially invalidates comparisons and is fundamentally uninterpretable. Thus, from a traditional measurement perspective this approach is a glaring weakness. On the other hand, from an inclusive and social constructivist perspective, this approach is a significant strength. These differing perspectives illustrate how divergent criteria lead to contrasting judgments about what constitutes strengths and weaknesses. *From our perspective*, the lack of standardized and defined criteria for rating judgments was a pragmatically justified approach to dealing with the realities and complexities of this multi-faceted, diverse-contexts, collaborative process.

#### *5. Not all Phase 2 donor studies followed the revised matrix.*

Phase 2 donor studies followed a general template from Phase 1 and were also offered a light additional standardized matrix for their Phase 2 studies including key "mirror questions" for the country matrix, but not all applied the additional matrix. While some partner countries struggled with the standardized partner country operational matrix, most understood its importance to the synthesis, found it useful, and attempted to follow it. In contrast, an adapted template was developed to guide donor studies, but was not applied systematically (p. 5, Final Report). The adapted donor study matrix was developed only after the Phase 2 donor studies had already received the same initial terms of reference as in Phase 1 and some had started work. The Management Group and Evaluation Team discussed and negotiated the framework for donor studies to allow comparability across all donor studies (Phase 1 and Phase 2). The donor studies were used to explain and elucidate patterns in partner country studies, but not all donors updated their Phase 1 studies so the smaller sample of Phase 2 donor participants and variability in how donor studies were conducted contributed to the donor data playing a relatively smaller role in the evaluation synthesis.

#### *6. Variations in length and degree of partner country engagement.*

One-third of the national coordinators we surveyed and over half of the evaluation team leaders reported that their country "struggled with the [national] contracting process to get the evaluation underway in a timely manner." Thus, evaluation teams formed at different times with varying degrees of training and support – and time to get the work done. Those countries that started late had less time to collect high quality data and ended up relying on a few sources and available documents. In one case, the partner country evaluation team was not contracted until November 1, 2010, a mere month before the draft country report was due to the international Core Evaluation Team.

As one Evaluation Team member commented:

This [country-controlled procurement process] was a specific, high risk case of “walking the talk” of Paris Declaration principles. It was not the only cause of delays but added a lot [to delays that did occur] and the resulting wear and tear [in getting the synthesis done]. But the alternatives would have been a centralized “Northern” managed process and/or a very much longer evaluation process, jeopardizing the relevance of results.

*7. Capacity variations within partner countries and greater need for capacity support than anticipated*

Different evaluation teams and team leaders had varying degrees of evaluation experience and expertise. Regional workshops were conducted to enhance the capacity of country teams to use the operational matrix and gather appropriate evidence. But some country-level evaluators needed more capacity development than the regional workshops could provide and, as noted above, not all team leaders made it to a regional workshop, in some cases because of delayed procurement and contracting processes in-country. This placed a greater responsibility on the Core Evaluation Team to provide additional, individualized capacity support, a level of engagement that had not been anticipated in the original terms of reference. While 80% of evaluation team leaders in our survey reported receiving excellent or good support and guidance from the core evaluation team, 20% reported that the guidance was only fair or poor on matters like using the matrix, collecting strong evidence, and writing the country report.

This concludes the discussion of weaknesses. See page 7 for a summary of the seven weaknesses. We turn now to lessons from the experience of evaluating Phase 2 of the Paris Declaration.

## **LESSONS**

*1. Assuring evaluator independence and genuine stakeholder involvement can be mutually reinforcing in support of enhanced evaluation quality and credibility.*

Stakeholder involvement assures the relevance of evaluations and evaluator independence ensures credibility. But stakeholder involvement is sometimes seen as undermining independence while processes to assure independence often limit stakeholder involvement. The Evaluation of the Paris Declaration has demonstrated how to strike the appropriate balance by creating structures and processes that provide meaningful and authentic stakeholder involvement without impinging on evaluator independence. Clearly delimited roles, skilled facilitation, and shared commitment to high quality evidence on all sides are factors that support balance and enhance the potential for use. A degree of ongoing, mutual monitoring between the international and national governance structures and evaluation teams contributed to establishing a contextually appropriate balance between evaluator independence and stakeholder involvement, with early intervention when potential difficulties arose.

Discussion of lessons continued next page.../

*2. The DAC quality standards for evaluation provide relevant, appropriate and useful guidance for conducting a complex, strategic, and principles-focused evaluation.*

The DAC standards were developed to guide evaluation of projects and programs. But the Paris Declaration is not a program. It is a set of principles and political commitments. As the Final report notes:

The object of the Evaluation – an agreed set of principles and commitments to improve aid effectiveness – is not a project or programme, the more normal objects of development evaluation. In a broad sense it is more like a strategy, a domain where evaluation is beginning to be tested, but the Declaration campaign has less-clear boundaries than most strategies. Interesting common elements can also be found in the growing experience in evaluating policy influence.  
(Final report, p. 3)

Despite this ambiguity about exactly what was being evaluated, an ambiguity commented on by many of the partner country and donor study participants, the DAC standards proved highly appropriate and useful. Three-fourths (76%) of country participants reported that the DAC standards were “very useful” or “somewhat useful” to the partner country evaluations. They also proved appropriate and useful for this evaluation of the evaluation. The Final Report includes a detailed analysis of how the DAC Standards informed the evaluation (see pp. 218-221).

*3. The analytical, interpretative and reporting framework for assessing progress on major Paris declaration and Accra Action Agenda outcomes -- **direction, pace and distance travelled** -- proved useful in synthesizing and comparing findings, conclusions, and judgments.*

This framework was created by the Core Evaluation Team in consultation with national coordinators, the Evaluation Management Group, and the International Reference Group. Direction, pace and distance travelled treats development as a journey, including the application of agreed changes in aid programmes. It is a unique and creative evaluation framework, well-suited to the challenge of synthesizing findings on progress across 21 partner country case studies.

*4. A comprehensive and complex international evaluation that involves multiple countries in a genuinely joint evaluation endeavor will need to anticipate and provide significant in-country capacity support and development to help overcome inevitable variations in evaluation capacity.*

The original evaluation terms of reference for the core evaluation team did not include substantial capacity-building and support for individual country evaluation teams. The national procurement processes were expected to yield evaluation teams with sufficient capacity to undertake the partner country evaluations. The Core Evaluation Team expected to provide regional training to evaluation team leaders. However, not all evaluation team leaders were able to participate in the regional workshops. In some cases, delays in procurement of evaluation teams meant that the evaluators had not been selected when the regional workshops were held. The standardized operational matrix was not easy to use without support and training. Other capacity weaknesses surfaced as draft country reports were submitted. Adjustments were made by the Management Group and the Core Evaluation Team to increase capacity support, but this came late, and introduced the possibility that the Core Evaluation Team might be perceived as having too much influence in the completion of

country reports. This concern, about undue influence, expressed by a few at the draft stages of review during the 3<sup>rd</sup> International Reference Group meeting in Indonesia had disappeared by the 4<sup>th</sup> International Reference Group meeting in Copenhagen, when it was clear that the partner countries owned their own evaluations and all involved understood how those evaluations were used for the synthesis. Still, concerns about inadequate capacity remained. Had the capacity needs been better anticipated and planned for, a separate capacity-building and support function might have been created separate from the core evaluation team. At any rate, a lesson from this evaluation is that capacity will have to be built and supported for some countries throughout the evaluation. In that regard we would affirm the core team's lesson about capacity support:

Where new approaches are required, capacity constraints need to be anticipated and addressed. Similarly, support to the technical aspects of individual evaluations is also crucial to ensuring consistency and commonality of approach. Adequate resources for specific technical support are an essential part of the Evaluation design; they should not be seen as 'additional' or to be drawn on 'if/where required'. (Final Report, p. 218)

*5. An evaluation design that is “good enough,” even though not ideal, will provide credible and useful findings.*

Evaluation pioneer Peter H. Rossi formulated the “good enough rule” for evaluation admonishing evaluators to choose the best possible design, taking into account practicality and feasibility. This rule evokes Nobel-prize winning economist Herbert Simon's notion of satisficing (as opposed to optimizing or maximizing) in making decisions. *Satisfice* combines *satisfy* with *suffice* to indicate that adequacy is often the best one can do in real world situations. Modest claims and conclusions can be supported by relatively modest data. Strong claims and conclusions require strong supporting evidence. The Phase 2 Evaluation of Paris Declaration draws conclusions and makes recommendations appropriate to the quality of evidence generated and available.

The previous section of this report noted a few methodological and measurement weaknesses in the Evaluation of the Paris Declaration. The Final Report notes the limitations of the design and data collection. Those who seek to criticize the design and question the findings will find much to complain about and question. It is far from an ideal evaluation. The question, however, is not whether it is ideal, but whether it is adequate, given the constraints and challenges faced in mounting such an enormous effort. Invoking the “good enough rule,” we find that the evaluation satisfices. The evidence for that judgment is presented in this evaluation of the evaluation.

*6. Engagement of diverse international stakeholders in a global evaluation process requires careful planning, structure, execution, and facilitation.*

The International Reference Group (IRG) that advised the evaluation included participants from the 22 partner countries, donors, multilateral organizations, civil society representatives, and, chiefly, members of the OECD/DAC Network on Development Evaluation as well as members of the OECD/DAC Working Party on Aid Effectiveness. The IRG was co-chaired by representatives from Malawi and Sweden and met four times during Phase 2 to provide guidance to the evaluation process. We observed the 3<sup>rd</sup> and 4<sup>th</sup> meetings of the IRG in Indonesia and Copenhagen, respectively. The IRG members reviewed and commented on the evaluation framework and the inception report, emerging findings and the draft Final Report. Those who would convene large and diverse stakeholder groups to advise on evaluation design, data collection, analysis, and reporting

would do well to study the structure and processes of the Paris Declaration IRG, especially with regard to the first lesson above, finding the appropriate balance between stakeholder involvement and ensuring evaluator independence.

This report describes how those involved assess their involvement in the evaluation process and the results of the evaluation. On our survey of partner country participants, among those who felt they knew enough to rate the IRG's capacity to engage in the evaluation, 75% rated the capacity as "high" and 25% as "moderate." None rated the group's capacity as "poor." It is worth highlighting this lesson on careful planning, structure, execution, and facilitation of stakeholder involvement so that those who would undertake such processes appreciate and understand that stakeholder engagement can be difficult, challenging, expensive, controversial, conflict-laden, and undermining, and will be all those things to an extent that harms the evaluation's credibility unless the process is carefully planned and facilitated. The more diverse the group, the greater the risks. The IRG process more than satisfied. It was, in our judgment, exemplary both in design and execution.

*7. Evaluation findings, conclusions, and recommendations are more accessible, understandable, and potentially useful for informing policy deliberations and decision-making when stated clearly, and jargon is avoided, including but not limited to evaluation jargon.*

Throughout the evaluation, the Core evaluation Team was committed to avoiding jargon and acronyms, and the clarity of the Final Report is a tribute to their commitment in this regard. The Final Report provides conclusions and recommendations to inform the deliberations of the 4<sup>th</sup> High Forum on Aid Effectiveness and, in so doing, avoids the formative and summative jargon of evaluation, as well as other off-putting professional and methodological jargon. Much of the final meeting of the International Reference Group was devoted to clarifying and sharpening findings, conclusions, and recommendations to enhance communications and impact. The most significant difference between the review draft of the final synthesis and the Final Report was the clarity and sharpness of the conclusions and recommendations.

*8. Evaluation findings and conclusions are more likely to invite and inform deliberation when core evaluation questions are framed as descriptive, open-ended inquiries rather than asking for simplistic judgments.*

The original terms of reference for the evaluation called for the evaluation to be both formative and summative (see p, 194, Annex of the Final report). Indeed, one of the three primary evaluation questions was originally framed as summative, and that wording remains in the opening of Chapter 4:

“Has the implementation of Paris Declaration strengthened the contribution of aid to sustainable development results?” (Final Report, Chapter 4 opening, p. 43)

This question wording invites a simple Yes/No summative answer, even when it adds the question of how the contribution, if any, occurred. Wisely, in our view, the Executive Summary, and the Evaluation Framework and Matrix ask a more nuanced and open-ended question, one that invites deliberation and balance:

“What contributions have improvements in aid effectiveness made to sustainable development results? (Final report, p. xii).

And later: “What contributions can aid effectiveness reforms *plausibly* be judged to have made to development results?” (Chapter 1, p. 5; and p. 202, technical Annex 5)

This change in wording becomes significant in light of the evaluation’s answer:

“None of the [country] evaluations finds an immediate and direct connection between the Declaration campaign and development results achieved, but there is evidence in a solid majority of the reports that it has made at least some *plausible contributions* to better results through the pathways of the reforms traced in the previous chapter.” (p. 43)

It would be premature to render a summative judgment about the Paris Declaration. The evaluation design, the quality of data, and, most importantly, the complexity of development aid, do not support a summative judgment. The lesson here is to pay careful attention to the realistic and plausible purpose an evaluation can fulfill and make sure that the framing of questions are answerable and appropriate. Don’t raise the stakes to summative judgment when such judgment is not possible, appropriate, nor necessary.

One Core Evaluation Team member, reflecting on this issue, put the matter succinctly:

Be realistic in both the questions and the design about what is feasible, and you’re more likely to get some honest answers.

*9. The value of an evaluation lies not just in the findings, but in the impacts that emerge among those who participate in the evaluation as they engage in the inquiry.*

The second strength listed earlier was that the evaluation expanded knowledge of and engagement with the Paris Declaration principles and commitments. The evaluation process increased awareness of and attention to the Paris Declaration and aid reform well beyond what would have occurred without the evaluation. This added layers of impact beyond the value of the findings generated and affects calculation of the cost-benefit ratio of undertaking an evaluation, including what those involved learned about evaluation. As one donor member of the International Reference Group commented:

“The Paris Declaration evaluation had an educational value in the sense that it demonstrated to the partner countries how an international collaborative evaluation is conducted. We are confident that the participation of a large number of scholars, experts and decision makers from the developing countries has enhanced their understanding of the evaluation process, the challenges which evaluators face, and the nature, strength and limitations of the findings and recommendations.”

The impact of an evaluation process on those involved is much more likely to be substantial and sustainable when the evaluation is a joint effort -- the focus of the next lesson.

*10. The joint nature and full transparency of the Evaluation of the Paris Declaration makes it a model for international evaluation.*

Country participants in the evaluation were asked to rate “the extent to which the way the Paris

Declaration evaluation was conducted constitutes a model for international policy implementation evaluation.” 74% rated the evaluation model as excellent or good, affirming that they have come to see it as an international model with implications beyond this particular exemplar. Among donors and international organization participants in the IRG, 52% rated the evaluation as excellent or good as a model. In particular, our interviews affirm that the evaluation succeeded in its aspiration to be conducted as a “joint evaluation,” including the commitment to conduct the evaluation in accordance with the principles of the Paris declaration. This is what is sometimes called “walking the talk.” The evaluation adhered to the Paris Declaration principles even as it evaluated the implementation and results of the Paris Declaration. That is what makes it a model of international evaluation excellence.

Technical Annex 5 provides details on the joint evaluation and other dimensions of the approach and model. For those who want to know how the evaluation was conducted and what makes it a model, as well as lessons learned by the Core Evaluation Team, Annex 5 is invaluable. Indeed, Annex 5 is, itself, a model of how an evaluation’s technical and methodological report should be written and reported. Too few evaluations include this level of detail, disclosure, transparency, and guidance. All the documents produced as background materials and guidance documents have been posted on the DAC website and made openly available and accessible. That degree of sharing is also a model for how evaluations should be conducted and reported.

*11. Use of findings, conclusions, and recommendations can begin immediately when a high-stakes evaluation is appropriately conducted, deadlines are met, and results are disseminated widely and immediately.*

While this evaluation of the evaluation focuses on auditing how the evaluation was conducted and its adherence to international standards of quality, the ultimate evaluation of the evaluation will be how it is used. Hopefully, a subsequent evaluation of use will occur. But as this is written, about six weeks after the Final Report was released, it is already clear that the report is garnering widespread international attention and being used in preparations for the 4<sup>th</sup> High Level Forum on Aid Effectiveness to be held at the end of 2011. The Secretariat, the Evaluation Management Group, the International Reference Group, and the Core Evaluation Team attended to use from the beginning and throughout all aspects of the evaluation. The third strength listed earlier was that the evaluation has been utilization-focused at every stage. All those involved have planned for use. They didn’t make the all-too-common mistake of waiting until the evaluation was completed to attend to use. Indeed, the final meeting of the International Reference Group in April, 2011, was dominated by concerns about making the report useful: clarifying findings, sharpening conclusions, and ensuring that recommendations were meaningful and actionable. Early feedback from those already using the report appears to affirm that those efforts were successful and worthwhile. Further evaluation will be needed to assess actual use among participants in the 4<sup>th</sup> High Level Forum, but the early indications are the attention to utility is yielding the intended results.

*12. Evaluation of an evaluation (meta-evaluation) can, and should be undertaken during the evaluation, not just afterward.*

We opened this evaluation of the evaluation by noting that it has become a standard in major high-stakes evaluations to commission an independent review to determine whether the evaluation meets generally accepted standards of quality, and, in so doing, to identify strengths, weaknesses, and lessons. But meta-evaluations are often undertaken only after an evaluation is completed, as a post

hoc, retrospective activity. This evaluation of the evaluation was commissioned halfway through the evaluation process in time for us to observe the direction, pace, and distance travelled by the evaluation as it concluded. This made it possible to interview key participants, survey those involved, and document how the evaluation was actually conducted and the final report constructed, reviewed, and finalized. The commitment to and timing of the meta-evaluation, like the evaluation process itself, has been exemplary. All that remains is to assure that ultimate use is documented and evaluated.

Having reviewed major strengths, weaknesses and lessons, we turn to the details of our approach to evaluation of the Evaluation of the Paris Declaration.

### **GATHERING the PERSPECTIVES OF THOSE INVOLVED**

As part of this evaluation of the evaluation, we conducted a survey of national reference group members and evaluation team members in the 21 Phase 2 partner countries. The survey was developed collaboratively with the Evaluation Secretariat and Core Evaluation Team following our participation in the International Reference Group meeting in Indonesia, in December 2010. At that meeting we had the opportunity to interview a number of national coordinators and partner country evaluation team leaders. We also facilitated a reflective practice session with International Reference Group participants in which they identified strengths, weaknesses, and lessons. The results of that exercise are included in this report and formed the basis for developing the survey aimed at all country-level participants. Details of the survey will be reviewed below.

We also administered a survey to members of the International reference Group who were donor and international organization representatives. That survey was administered during the 4th International Reference Group meeting in Copenhagen in late April, 2011. Those results, from 23 donor and international organization members of the International Reference Group meeting will also be presented in this report.

This evaluation of the evaluation does not include data from national and international policymakers who are ultimately the primary intended users of the *Phase 2 Evaluation of the Paris Declaration*. A future study of their judgments about and uses of the findings, conclusions, and recommendations would add an important perspective beyond those of the people directly involved in producing the findings, which is the focus of this report.

#### **Survey of partner country participants**

Members of national reference groups and partner country evaluation teams were surveyed in late March through early April, 2011. The survey was administered online with opportunities for hard copies to be completed by those with limited online access or who simply preferred a hard copy. The Minnesota Center for Survey Research at the University of Minnesota administered the survey and tabulated responses. The survey was available in English, French and Spanish. In addition, country level participants at the International Reference Group meeting in Copenhagen who had not completed the survey were given the opportunity to add their responses. In all cases, survey responses were anonymous. The only background identification question asked was whether the respondent was involved in a National Reference Group or country evaluation team function. The survey went to 209 potential respondents. The overall response rate was 38% (n=80); not everyone answered every question so the number of responses can vary by 1 to 3 per table. Of the 21

National Coordinators, 76% responded; 71% of evaluation team leaders responded. Parts of this analysis will compare the responses of National Coordinators and Evaluation Team Leaders. The survey focused on the partner country-level evaluations.

The survey included both closed-ended questions analyzed statistically and open-ended questions which we content-analyzed to determine patterns and themes, and which provide illuminative quotations included throughout this report. Sections of the survey covered the following topics:

- How well the country evaluations handled some common evaluation issues like explaining the purpose of the evaluation to key stakeholders and using appropriate methods to answer major questions.
- Perceptions and ratings on strength of evidence to support conclusions.
- Factors affecting the use of country evaluations
- Dissemination activities for and uses of the country evaluation reports
- Utility of core documents, guidance notes, input papers, and reference materials provided by the core evaluation team
- Support and quality of guidance received to conduct the country evaluation
- Capacity issues
- Overall perceptions and ratings of various aspects of the evaluation process and results

We began with the survey with questions concerning some common issues that arise in evaluation. Respondents were asked to provide their honest and confidential judgment about how well the country evaluation in which they had been involved addressed each of these issues. The table on the next page presents the aggregate results. That is followed by comparisons of the responses of National Coordinators and Evaluation Team Leaders. We begin with these results because they provide an overview of the kinds of issues involved in conducting the evaluation of the Paris Declaration *at the country level*.

The evaluations conducted in the partner countries provided the bulk of the findings on which the final synthesis report was based. As the survey results on the next page show, the experiences in the partner countries varied, as did the quality of the data gathered and the rigor of the analysis.

One national coordinator wrote in the survey: “The Country evaluation was important in assessing the results of Government policies that were implemented even before the signing of the Paris Declaration on aid effectiveness. It also provided useful insight into new initiatives by Government to strengthen aid coordination and application of more effective results management policies. The evaluation was treated as an important and serious undertaking.”

In contrast, another wrote: “There were difficulties throughout the process and, in the analysis, associating development outcomes to the PD. In our country, the widely held perception and belief is that the results realised over the years are due to a host of so many other factors that the influential or catalytic role of the PD is remote. And we found generally limited knowledge and understanding of the PD amongst stakeholders making the evaluation of its implementation a complex and difficult exercise.”