META-EVALUATION: EVALUATING THE EVALUATION OF THE PARIS DECLARATION

Michael Quinn Patton
Founder and Director, Utilization-Focused Evaluation
Saint Paul, Minnesota, USA

Abstract: 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. This is called a meta-evaluation. Given the historic importance of the Evaluation of the Paris Declaration, the Management Group commissioned a meta-evaluation of the evaluation. The meta-evaluation concluded that the findings, conclusions, and recommendations presented in the Paris Declaration Evaluation adhered closely and rigorously to the evaluation evidence collected and synthesized. The meta-evaluation included an assessment of the evaluation’s strengths, weaknesses, and lessons. This article describes how the meta-evaluation was designed and implemented, the data collected, and the conclusions reached.

Résumé : Il est devenu courant, dans les évaluations comportant des enjeux majeurs, de commander un audit indépendant chargé de vérifier si l’évaluation répond aux critères de qualité généralement acceptés : c’est ce que l’on appelle une méta-évaluation. Compte tenu de l’importance historique de l’Évaluation de la Déclaration de Paris, le groupe de gestion a demandé qu’une telle vérification soit effectuée. Selon la méta-évaluation, les constatations, conclusions, et recommandations présentées dans l’Évaluation étaient rigoureusement fidèles aux preuves collectées et synthétisées. La méta-évaluation inclut également une analyse de ses forces et faiblesses ainsi que des enseignements tirés de l’évaluation. Cet article décrit la conception et la mise en œuvre de la méta-évaluation, de même que les modalités de collecte des données et les conclusions formulées.
Shinkfield, 2007, p. 649). The major addition to the Joint Committee Standards for Evaluation, when revised in 2010, was the addition of “Evaluation Accountability Standards” focused on meta-evaluation. Exhibit 1 presents the new evaluation accountability standards.

**Exhibit 1**

**Evaluation Accountability Standards**

| E1 Evaluation Documentation | Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes. |
| E2 Internal Meta-evaluation | Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes. |
| E3 External Meta-evaluation | Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards. |

*Source: Joint Committee on Standards for Educational Evaluation, 2010; Yarbrough, Shulha, Hopson, & Caruthers, 2010.*

**EVALUATION OF THE EVALUATION OF THE PARIS DECLARATION**

Given the historic importance of the Evaluation of the Paris Declaration (Dabelstein & Patton, 2012b), the Management Group commissioned an 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, the Secretariat, 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 about the evaluation process and partner country evaluations; and interviewing key people involved in and knowledgeable about how the evaluation was conducted. The evaluation of the evaluation included assessing both the evaluation report’s findings and the technical appendix that details how findings were generated (Wood et al., 2011a, 2011b).

The meta-evaluation report (Patton & Gornick, 2011) was published and made available online two weeks after the Final Evaluation report was published. This timing was possible because the meta-
Overall Meta-evaluation Conclusion

The meta-evaluation concluded that the findings, conclusions, and recommendations presented in the *Phase 2 Evaluation of the Paris Declaration* (Wood et al., 2011a) adhered 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 were appropriately acknowledged. The comprehensive Technical Annex (Wood et al., 2011b) accurately described data collection and analysis approaches. Partner country and donor evaluation reports, upon which the Synthesis Evaluation was 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 the voluntary Phase 2 evaluation of the implementation of the Paris Declaration engaged in systematic and in-depth reflection and evidence-based processes that made their conclusions and insights worthy of serious attention. The Synthesis Report accurately captured those evidence-based conclusions and insights. The conclusions drawn and recommendations made were appropriate for and derived from the evidence analyzed and synthesized (Patton & Gornick, 2011, p. i).

Major Evaluation Strengths Identified by the Meta-evaluation

The meta-evaluation identified and discussed 10 major strengths of the Paris Declaration Evaluation.

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.
4. Keeping quality of evidence at the centre of the evaluation.
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 endeavour 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. (Patton & Gornick, 2011, pp. 2–6)

Detailed discussion of these strengths is available in the online report (Patton & Gornick, 2011).

Major Weaknesses Identified by the Meta-evaluation

The meta-evaluation identified and discussed seven weaknesses of the Paris Declaration Evaluation.

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 judgements.
5. Variations in adherence to the revised matrix by some Phase 2 donor studies.
6. Variations in length and degree of partner country engagement.
7. Capacity variations within partner countries and greater need for capacity support than anticipated.

Detailed discussion of these weaknesses is available in the online report (Patton & Gornick, 2011).

AN INDEPENDENT AUDIT OF THE SYNTHESIS EVALUATION

Having examined the strengths and weaknesses of the evaluation process and methods, and extensively reviewed the Final Report of
the *Phase 2 Evaluation of the Paris Declaration*, the meta-evaluation reached an overall judgement about the merit and worth of the Evaluation. In effect, this served the accountability function of the meta-evaluation and took the form of an independent audit statement included as part of the preface in the Final Evaluation (Patton, 2011). Both the audit statement and this full report on which the audit statement was based concluded that, in our opinion, the Final Synthesis Report could 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 therein, the findings of the *Phase 2 Evaluation of the Paris Declaration* can be studied and used as trustworthy and credible. Exhibit 2 presents the independent audit statement and is included here to illustrate what such a statement can look like. Few will have read the actual meta-evaluation. The audit report, however, was included as a preface in the actual evaluation and was therefore likely to get more widespread attention.

**SIGNIFICANT LESSONS IDENTIFIED AND DISCUSSED IN THE META-EVALUATION**

The meta-evaluation was commissioned not solely as an accountability exercise, but also to support learning. In this special issue of the *Canadian Journal of Program Evaluation*, Dabelstein & Patton (2012a) present lessons that were identified by experts in joint evaluation at a workshop in Paris in early 2012 that was the final activity of the Paris Declaration Evaluation. This section reviews 12 lessons that were included in the meta-evaluation.

**Lessons from the Paris Declaration Evaluation**

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. Exhibit 3 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 (Development Assistance Committee, 2010) provide relevant, appropriate, and useful guidance for conducting a complex, strategic, and principles-focused evaluation. The Paris Declaration Evaluation Final Report (Wood et al., 2011a) includes a detailed analysis of how the DAC standards informed the evaluation.

Exhibit 2
The Independent Audit Statement

May, 2011

An Independent Audit of the Evaluation

In our opinion, the findings and conclusions generated adhere closely and rigorously to the evaluation evidence collected. 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 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 Final Report accurately captures those evidence-based conclusions and insights.

In our opinion, the 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, the findings can be studied and used as trustworthy and credible.

Source: Patton, 2011.
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 judgements. 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 endeavour 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 satisfies. 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 from the meta-evaluation of the tensions in top-down versus bottoms-up evaluation processes is presented as Exhibit 4.

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 judgements. It would be premature to render a summative judgement about the Paris Declaration. The evaluation design, the quality of data, and, most importantly, the complexity of development aid do not support a summative judgement. 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.

Exhibit 4
Top-Down vs. Bottoms-Up Evaluation Processes

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. The Evaluation garnered widespread international attention upon publication and was used in preparations for the 4th High Level Forum on Aid Effectiveness 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 was exemplary.

META-EVALUATION OF COUNTRY STUDIES

An additional independent meta-evaluation of the partner country case studies that generated the basic data for the synthesis evaluation report was conducted by researchers at the Institute of Development Policy and Management, University of Antwerp, Antwerp (Songco, Holvoet, & Inberg, 2012). This country cases meta-evaluation, conducted completely independently of the Paris Declaration Evaluation and after both the Final Evaluation Report (Wood et al., 2011) and commissioned meta-evaluation (Patton & Gornick, 2011) were published, opened by noting that “the evaluation of the Paris Declaration (PD) is one of the most important and challenging evaluative undertakings of the past decade in the aid
The scope and importance of this evaluation makes it a particularly suitable subject for a meta-evaluation” (Songco et al., 2012, p. 1). The report summarizes the country cases meta-evaluation as follows:

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. (Songco et al., 2012, p. 1)

Whether to exclude some of the weaker country reports in the final synthesis was a matter of considerable discussion. In the end, the core evaluation team judged that all country reports met minimum standards of evidence for inclusion. I would add that the logic of synthesis in this case is not captured by the proverb that “a chain is only as strong as its weakest link.” Thus, even several weaker country studies can be informative and, I would argue, are informative when they further confirm the overall patterns and themes as, indeed, they do. At the synthesis level it is the overall direction and cross-country patterns that are conclusive.

THE PERSPECTIVES OF THOSE INVOLVED IN THE PARIS DECLARATION EVALUATION

Debates about quality and credibility of evidence come with the territory in evaluation. As noted earlier, one of the weaknesses identified in the commissioned meta-evaluation was the variation in quality of evidence at the country level. But how did those involved in generating the partner country case studies view the evaluation process and quality of evidence?
As part of the evaluation of the evaluation, we conducted a survey of national reference group members and evaluation team members in the 21 Phase 2 partner countries. I drafted the initial survey questions based on observations and interviews at 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 formed the basis for developing the survey aimed at all country-level participants. The survey questions were revised based on feedback from the Evaluation Secretariat and Core Evaluation Team. The survey was independently administered and tabulated by the Minnesota Center for Survey Research, University of Minnesota.

We also administered a survey to members of the International Reference Group (IRG) who were representatives of donor and international organizations. 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, were also included in the meta-evaluation. (See Dabelstein & Kliest, 2012, in this special issue for details about the nature, composition, and role of the International Reference Group of stakeholders and the National Reference Groups.) The survey was administered to the two different groups at different times based on when the survey questions were most relevant to each group. The initial survey of country teams was done before the draft synthesis report was available, so only country reports were available for meta-evaluation review in the country surveys. The donor and International Reference Group survey was done after the draft synthesis report was ready for review and therefore includes questions focused on the synthesis report and process. All participants and respondents understood the timing issues. At the International Reference Group meeting in Copenhagen, both country representatives and donors responded in person to the draft synthesis report.

Survey of Partner Country Participants

Members of national reference groups and partner country evaluation teams were surveyed from late March through early April 2011. The survey was administered online with opportunities for hard cop-
ies 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. Of the 21 National Coordinators, 76% responded; 71% of evaluation team leaders responded. The full meta-evaluation analysis compared 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 that we content-analyzed to determine patterns and themes, and that 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 the survey with questions concerning some common issues that arise in evaluation. Respondents were asked to provide their honest and confidential judgement about how well the country evaluation in which they had been involved addressed each of these issues. Exhibit 5 presents the aggregate results. These results 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 showed, the experiences in the partner countries varied, as did the quality of the data gathered and the rigour of the analysis.

Exhibit 5
Assessments of Country-Level Evaluation Strengths and Weaknesses by Country Participants

<table>
<thead>
<tr>
<th>Participants were asked to respond with reference to the country evaluation in which each was involved (n = 77)</th>
<th>We did this very well (%)</th>
<th>We did this adequately (%)</th>
<th>We struggled with this (%)</th>
<th>We failed at this (%)</th>
<th>Not sure (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Explaining the purpose of the evaluation to key stakeholder</td>
<td>47</td>
<td>44</td>
<td>6</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>2. Managing the contracting process to get the evaluation underway in a timely manner</td>
<td>29</td>
<td>40</td>
<td>25</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3. Covering the full scope of questions in the matrix</td>
<td>28</td>
<td>44</td>
<td>24</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>4. Adding country-specific questions to the matrix</td>
<td>24</td>
<td>51</td>
<td>16</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>5. Finding and using in-country evaluators with competence to do all that the evaluation demanded</td>
<td>28</td>
<td>44</td>
<td>24</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>6. Having a cohesive evaluation team that worked well together</td>
<td>38</td>
<td>40</td>
<td>19</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>7. Coordination between the national coordinator, the National Reference Group, and the evaluation team</td>
<td>46</td>
<td>42</td>
<td>11</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>8. Explaining the context and its implications in the report</td>
<td>30</td>
<td>60</td>
<td>10</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>9. Using appropriate methods to answer major questions</td>
<td>28</td>
<td>54</td>
<td>13</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>10. Getting access to existing data we needed</td>
<td>22</td>
<td>43</td>
<td>33</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11. Getting access to a broad range of stakeholders to ensure diverse and balanced perspectives</td>
<td>34</td>
<td>51</td>
<td>12</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>12. Getting a good response rate for interviews and surveys</td>
<td>20</td>
<td>49</td>
<td>30</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>13. Getting sufficient high-quality evidence to draw strong conclusions</td>
<td>13</td>
<td>53</td>
<td>31</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Triangulating data sources (using multiple methods to cross-validate findings)</td>
<td>22</td>
<td>48</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>15.</td>
<td>Making the sources of data for the evaluation findings transparent</td>
<td>28</td>
<td>61</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>16.</td>
<td>Dealing with conflicts of interest among diverse stakeholders that could affect the evaluation’s credibility</td>
<td>11</td>
<td>60</td>
<td>25</td>
<td>—</td>
</tr>
<tr>
<td>17.</td>
<td>Maintaining the independence of the evaluation</td>
<td>40</td>
<td>45</td>
<td>14</td>
<td>—</td>
</tr>
<tr>
<td>18.</td>
<td>Having enough time to do a good job on the evaluation</td>
<td>14</td>
<td>39</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td>19.</td>
<td>Getting the report done on time</td>
<td>31</td>
<td>38</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>20.</td>
<td>Believing that the evaluation was worth doing</td>
<td>48</td>
<td>46</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>21.</td>
<td>Conducting the evaluation in a way that would be truly useful in the country</td>
<td>32</td>
<td>54</td>
<td>11</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* Survey results from members of National References Groups and country evaluation teams.  

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.
Exhibit 5 shows survey results from those who participated in the Paris Declaration Evaluation at the country level. Some highlights of the survey results are shown below.

**Strengths:** What the country participants report they did very well

1. Believing that the evaluation was worth doing  48%
2. Explaining the purpose of the evaluation to key stakeholders  47%
3. Coordinating between the national coordinator, the National Reference Group, and the evaluation team  46%
4. Maintaining the independence of the evaluation  40%
5. Having a cohesive evaluation team that worked well together  38%

Sample quotes from the open-ended question on strengths:

- “We had a good evaluation team that worked well together. With relevant methodologies and extensive literature on aid in our country, their independent report reflects a comprehensive overview about our aid picture. And I think this is really useful and can be seen as a handbook for those who want to know about aid here.”
- “The evaluation was able to draw engagement from a broad spectrum of stakeholders. It served as a venue to renew interest on the PD commitments, deepen engagement of other stakeholders on aid effectiveness, and create urgency to sustain important reform initiatives especially on use of country systems, managing for results and engagement with CSOs.”
- “The Team and the Office of National Coordinator coordinated the process very well, and a Working Group mechanism (in addition to National Reference Group) effectively involved all stakeholders very well.”

**Weaknesses:** What the country participants report they struggled with or failed at

1. Having enough time to do a good job on the evaluation  46%
2. Getting access to existing data we needed  34%
3. Getting sufficient high quality evidence to draw strong conclusions  33%
4. Getting a good response rate for interviews and surveys 31%
5. Getting the report done on time 30%

Sample quotes from the open-ended question on weaknesses:

- “We did struggle with obtaining information from the various government agencies because of the large volume of information scattered over various ministries. Some of the key people had either moved overseas or no longer worked in key positions but, somehow, those that we approached managed to locate information or directed us to other personnel who were also involved.”
- “The evaluation team struggled most with the complex, technical nature of the evaluation. It is difficult to evaluate or measure whether the PD strengthened the contribution of ODA [Overseas Development Aid] to development results or not because it takes a long time to see such results clearly, and a result can be affected by a lot of factors. And this evaluation had difficulties due to the time constraint.”
- “Because of the time constraint as well as the number and complexity of the questions that needed to be addressed, the evaluation is not as rigorous methodologically as it should have been. A broad expertise covering macroeconomics, aid structure, behavior and dynamics, sector-specific expertise, statistics/regression, etc. are also needed to be able to respond to the complexity of the evaluation. Such kind of expertise could not be available in a team of 3. A lot of logistical requirements were also required and challenging.”

Prominent Patterns and Themes

What emerged in these data were patterns and themes that were repeated throughout the meta-evaluation, triangulated in other survey items, interviews with participants, direct observations of working sessions, and in the direct analysis of country reports by the meta-evaluation team. The country processes worked reasonably well in launching the evaluations, building momentum for the evaluation’s purpose and importance, establishing national reference groups, selecting evaluation teams, and maintaining the evaluation’s independence. The evaluation teams struggled, however, with data collection and analysis, especially given some capacity deficiencies and
the time constraints under which the evaluation operated. We’ll take up these themes and examine them in more depth as we proceed.

QUALITY OF EVIDENCE RATINGS

Exhibit 6 presents ratings by country participants of quality of evidence for various aspects of the Paris Declaration Evaluation. Many factors affect strength of evidence, including resources and timelines for data gathering, response rates to interviews and surveys, existing availability of relevant data, and challenges in interpreting results and drawing conclusions about contributing factors and attribution. The national participants in the Paris Declaration were asked in the meta-evaluation survey to provide their honest and confidential assessment of the strength of evidence presented in their country reports for each of the major evaluation questions listed in Exhibit 6.

**Exhibit 6**
**Quality of Evidence Ratings**

<table>
<thead>
<tr>
<th>Ratings of the strength of evidence presented in the country reports (80 country respondents)</th>
<th>Strong evidence presented in the report (%)</th>
<th>Moderate evidence presented in the report (%)</th>
<th>Minimal evidence presented in the report (%)</th>
<th>No evidence presented in the report (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Overall implementation of the principles of the Paris Declaration</td>
<td>39</td>
<td>54</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>2. Improvements in the efficiency of aid delivery</td>
<td>22</td>
<td>63</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>3. Overall improvements in the management and use of aid</td>
<td>22</td>
<td>60</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>4. The country's ownership over development</td>
<td>57</td>
<td>30</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>5. Better (more inclusive and effective) partnerships</td>
<td>23</td>
<td>60</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>6. Delivering and accounting for development results</td>
<td>20</td>
<td>49</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>7. Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development</td>
<td>13</td>
<td>47</td>
<td>32</td>
<td>6</td>
</tr>
<tr>
<td>8. Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion</td>
<td>18</td>
<td>34</td>
<td>44</td>
<td>5</td>
</tr>
</tbody>
</table>
Respondents clearly differentiated quality of evidence for different evaluation questions. They reported having obtained the strongest evidence for “country’s ownership over development” and “overall implementation of the principles of the Paris Declaration.”

They reported the weakest evidence for improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion; sustainable increases in institutional capacities and social capital at all levels to respond to development challenges; whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development; and delivering and accounting for development results.

The meta-evaluation discusses at some length the finding that quality and strength of evidence varied by country and question. Weaknesses were due to a combination of factors: short timelines, inadequate resources, capacity limitations, and administrative and implementation challenges. As a result, several draft partner country reports were sometimes heavy on opinion and light on data. The core evaluation team feedback about those early drafts, therefore, focused on strengthening the evidence to support conclusions and judgments. The revised partner country reports reflected that feedback and technical assistance.

The constant theme at the IRG meeting was that conclusions and judgements had to be supported by evidence. The IRG process included peer reviews and open sharing of reports. It became clear to all in attendance which country reports were particularly weak. The emphasis on evidence was established and reinforced a shared norm that achieving higher quality was paramount. Core evaluation team members had coaching sessions with those country teams whose reports were weakest and/or requested such additional coaching sessions. The open and transparent peer review process of country reports at the International Reference Group meeting in Indonesia and the guidance provided by the core evaluation team substantially

<table>
<thead>
<tr>
<th></th>
<th>Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges</th>
<th>14</th>
<th>44</th>
<th>37</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>Evolution in the mix of aid modalities</td>
<td>22</td>
<td>52</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>11.</td>
<td>The implications for aid effectiveness in the future</td>
<td>26</td>
<td>44</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

strengthened the evidence presented in the final partner country reports. The revised reports were considerably improved in most cases compared to earlier drafts. In a concluding section of the survey, respondents \((n = 80)\) were asked to rate “the final quality of your country evaluation report.” Nearly four fifths (79%) rated the final quality of their country evaluation reports as excellent (28%) or good (51%). About one fifth (18%) rated the final quality as fair (10%) or poor (8%), while 3% had no opinion.

It is these fair or poor reports that the independent meta-evaluation of country reports would have excluded from the final synthesis report (Songco et al., 2012). The Final Evaluation Report included all country case studies as a matter of transparency and inclusiveness, concluding that all had at least some useful data, but clearly differentiated stronger and weaker evidence. In essence, the Core Evaluation Team recognized variations in data quality across questions and countries. In the Final Report, the Core Evaluation Team was careful throughout to note variations in data quality and strength of evidence. As one core evaluation team member explained, “The evidence base we applied for Synthesis had already been through a filter and findings which were weak [or] poorly evidenced were screened out.” (For a fuller discussion of variations in data quality and how the evaluation dealt with those variations, see Betts & Wood, 2012.)

**QUALITY OF THE FINAL SYSTHESIS EVALUATION REPORT**

At the 4th International Reference Group meeting in Copenhagen, donors and international organization participants and representatives were surveyed and asked to evaluate the quality of evidence in the Final Synthesis Report. While they were completing that survey, International Reference Group members from the partner countries participated in focus groups to clarify and interpret the country survey responses; some who had not completed country surveys also did so at that time.

Thus, partner country participants primarily answered questions about their own country reports. In contrast, the donors and international organization representatives were responding to the near-final draft of the Final Synthesis Report they had received prior to the Copenhagen meeting and which was the focus of review and feedback at the International Reference Group meeting. Exhibit 7 presents the results of the ratings of quality of evidence in the final report.
### Exhibit 7
#### Strength of Evidence in the Final Synthesis Evaluation Report: Donors and International Organization Representatives’ Ratings

<table>
<thead>
<tr>
<th>Ratings of the strength of evidence presented in the Final Synthesis Report (21 respondents)</th>
<th>Strong evidence presented in the report (%)</th>
<th>Moderate evidence presented (%)</th>
<th>Minimal evidence presented in the report (%)</th>
<th>No evidence or not sure (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Overall implementation of the principles of the Paris Declaration</td>
<td>33</td>
<td>52</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>2. Improvements in the efficiency of aid delivery</td>
<td>24</td>
<td>43</td>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td>3. Overall improvements in the management and use of aid</td>
<td>24</td>
<td>62</td>
<td>14</td>
<td>--</td>
</tr>
<tr>
<td>4. The country’s ownership over development</td>
<td>48</td>
<td>38</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>5. Better (more inclusive and effective) partnerships</td>
<td>24</td>
<td>62</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>6. Delivering and accounting for development results</td>
<td>5</td>
<td>52</td>
<td>38</td>
<td>5</td>
</tr>
<tr>
<td>7. Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development</td>
<td>5</td>
<td>32</td>
<td>53</td>
<td>10</td>
</tr>
<tr>
<td>8. Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion</td>
<td>5</td>
<td>19</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>9. Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges</td>
<td>5</td>
<td>44</td>
<td>43</td>
<td>10</td>
</tr>
<tr>
<td>10. Evolution in the mix of aid modalities</td>
<td>10</td>
<td>62</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>11. The implications for aid effectiveness in the future</td>
<td>10</td>
<td>40</td>
<td>40</td>
<td>10</td>
</tr>
</tbody>
</table>


These same IRG participants (donors and international organization representatives) were asked to rate the quality, credibility, and utility of the Final Synthesis Report ($n = 22$), as shown in Exhibit 8.

- 73% rated the quality of evidence as adequate.
- Nearly two thirds (64%) rated the Final Report as having high or adequate credibility.
Exhibit 8
IRG Participants (Donors and International Organization Representatives) Ratings of the Quality, Credibility, and Utility of the Final Synthesis Report ($n = 22$)

<table>
<thead>
<tr>
<th></th>
<th>High (%)</th>
<th>Adequate (%)</th>
<th>Low (%)</th>
<th>Mixed (%)</th>
<th>Not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Quality of evidence in the Final Synthesis Report</td>
<td>—</td>
<td>73%</td>
<td>9%</td>
<td>18%</td>
<td>—</td>
</tr>
<tr>
<td>2. Credibility of the Final Synthesis Report</td>
<td>18%</td>
<td>46%</td>
<td>14%</td>
<td>23%</td>
<td>—</td>
</tr>
<tr>
<td>3. Utility (likely) of the Final Synthesis Report</td>
<td>18%</td>
<td>36%</td>
<td>27%</td>
<td>18%</td>
<td>—</td>
</tr>
</tbody>
</table>

Source: Patton & Gornick, 2011, p. 34.

In summary, quality and strength of evidence in the Paris Declaration Evaluation varies by question, country, and donor. The weaknesses in the evidence are fully acknowledged in the final report. The meta-evaluation concludes:

What is clear, and can be stated unequivocally, is that attaining the highest possible quality of evidence was the driving force of this evaluation. Concerns about strength of evidence permeate the evaluation framing and guidance documents. Support for and feedback by the core evaluation team to country evaluators focused on supporting conclusions and judgements with evidence. The discussions at the International Reference Group meetings were dominated by attention to quality of evidence.

Based on our review of the evidence, we have concluded as follows:

In our opinion, the findings and conclusions generated adhere closely and rigorously to the evaluation evidence collected. 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 limi-
tations 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 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 Final Report accurately captures those evidence-based conclusions and insights.

In our opinion, the 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, the overall findings can be studied and used as trustworthy and credible. (Patton & Gornick, 2011, p. 35)

REFERENCES


Michael Quinn Patton is founder and director of *Utilization-Focused Evaluation*, Saint Paul, Minnesota. He is former president of the American Evaluation Association and recipient of both the Alva and Gunnar Myrdal Award from the Evaluation Research Society for “outstanding contributions to evaluation use and practice” and the Paul F. Lazarsfeld Award for lifetime contributions to evaluation theory from the American Evaluation Association. The Society for Applied Sociology honored him with the Lester F. Ward Award for Outstanding Contributions to Applied Sociology. He is author of six major evaluation books including *Essentials of Utilization-Focused Evaluation* (2012) and *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (2011). He conducted the meta-evaluation of the Paris Declaration Evaluation.