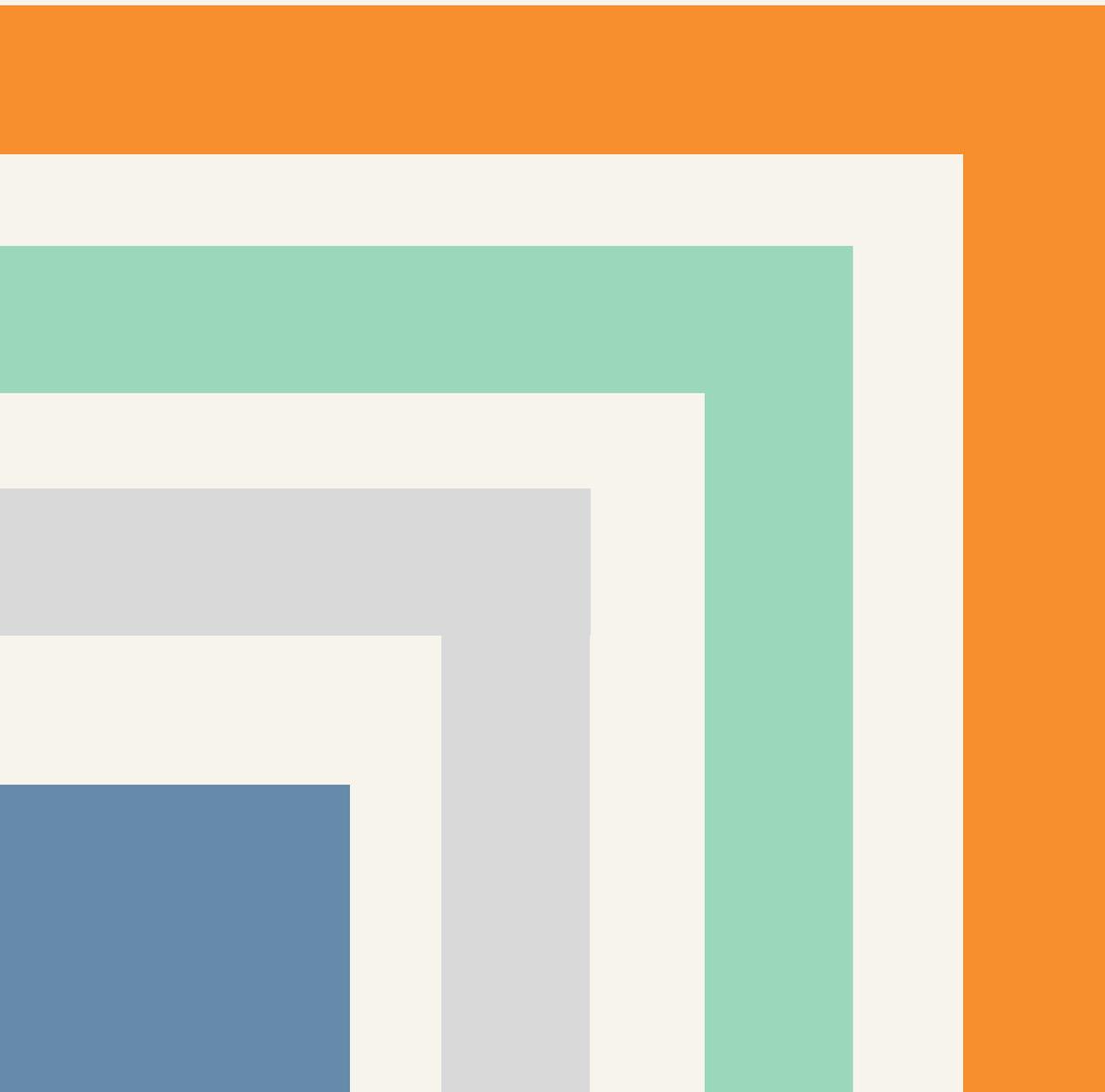




## UNFPA Evaluation Office

---

Assessing the quality of  
developmental evaluations  
at UNFPA



March 2020

Copyright © UNFPA 2020, all rights reserved.

This is an independent publication by the UNFPA Evaluation Office.

Any enquiries about this publication should be addressed to Mr Marco Segone, [segone@unfpa.org](mailto:segone@unfpa.org), and Ms Valeria Carou-Jones, [carou-jones@unfpa.org](mailto:carou-jones@unfpa.org)



[www.unfpa.org/evaluation](http://www.unfpa.org/evaluation)



[unfpa\\_eval](https://twitter.com/unfpa_eval)



[UNFPA Evaluation Office](https://www.youtube.com/UNFPA-Evaluation-Office)

## Table of contents

Foreword by Director, UNFPA Evaluation Office.....	4
Foreword by Michael Quinn Patton .....	5
1. Introduction.....	6
2. Background.....	7
2.1 Evaluations at UNFPA .....	7
2.2 Developmental evaluation .....	8
3. The concept of evaluation quality at UNFPA.....	11
3.1 Principle 1: Utility/utilization-focus.....	12
3.2 Principle 2: Developmental purpose .....	13
3.3 Principle 3: Evaluation rigor .....	14
3.4 Principle 4: Innovation niche .....	14
3.5 Principle 5: Complexity aware .....	15
3.6 Principle 6: Systemic thinking.....	15
3.7 Principle 7: Co-creation .....	16
3.8 Principle 8: Timely feedback.....	16
3.9 Standard 9: Independence and impartiality.....	17
3.10 Standard 10: Credibility .....	17
3.11 Standard 11: Gender equality and human rights .....	18
4. Tensions in ensuring quality of developmental evaluations .....	19
4.1 Major tensions.....	19
4.2 Evaluating the management of tensions.....	21
5. Quality assurance and assessment of developmental evaluations at UNFPA .....	22
6. The approach.....	23
6.1 Developmental evaluation principles and standards.....	23
6.2 Developmental evaluation products and process.....	24
6.3 Rating adherence to minimum practices .....	24
6.4 Data sources and methods.....	25

6.5 The steps for assessing the quality of developmental evaluations..... 25

6.6 Options for scaling up the quality assessment..... 26

Annexes ..... 27

Annex 1: Niches for developmental evaluation ..... 27

Annex 2: Developmental evaluation quality assessment grid for reports ..... 29

Annex 3: Developmental evaluation quality assessment grid for key informant surveys and interviews ..... 30

Annex 4: Developmental evaluation quality assessment grid for interviews of evaluators and UNFPA  
evaluation managers ..... 31

Annex 5: Summary of ratings ..... 32

Annex 6: Evaluation quality assessment report of developmental evaluation of results-based management at  
UNFPA..... 34

## Foreword by Director, UNFPA Evaluation Office

In 2019, UNFPA Evaluation Office managed for the first time, a centralized developmental evaluation of results-based management at UNFPA.<sup>1</sup> The purpose of developmental evaluation is distinct from more traditional formative, summative, and accountability-oriented evaluations in that it aims to inform the creation, development and ongoing adaptation of an initiative or interventions, often in dynamic contexts, that aim to make progress on complex issues. Developmental evaluation is also distinct from more traditional evaluation approaches in terms of a variety of evaluation practices related to roles and relationships, standards and principles.

The UNFPA evaluation policy<sup>2</sup> places a strong emphasis on quality assurance of evaluation processes and products. UNFPA has in place a quality assessment system<sup>3</sup> with several criteria for assessing the quality of final evaluation reports of formative and summative evaluations. However, the uniqueness of the developmental evaluation approach makes the existing quality assessment grid not fit for the purpose for assessing the quality of a developmental evaluation. For this reason, the Evaluation Office embarked on the development of the first ever developmental evaluation-specific quality assessment grid and guidance.

The key dimensions of quality for developmental evaluations are drawn from two major sources: a) the internationally recognized United Nations Evaluation Group's (UNEG) Norms and Standards; and b) the eight principles of effective developmental evaluation. The present guidance and grid combine these into eleven quality principles/standards. To our best knowledge, there are currently no other guidance and grids for assessing the quality of a developmental evaluation process and products in or outside the UN system. Therefore, we hope this tool contributes to further the utilization and quality of developmental evaluations within and beyond the UN system.

I wish to thank Mark Cabaj for the development of this ground breaking tool, Michael Quinn Patton for his valuable insights and guidance, and Valeria Carou-Jones for driving the entire process. Without their engagement, this guidance would have not been possible.

**Marco Segone**

Director, UNFPA Evaluation Office

---

<sup>1</sup> <https://www.unfpa.org/admin-resource/developmental-evaluation-results-based-management-unfpa>

<sup>2</sup> <https://www.unfpa.org/admin-resource/unfpa-evaluation-policy-2019>

<sup>3</sup> <https://www.unfpa.org/admin-resource/evaluation-quality-assurance-and-assessment-tools-and-guidance>

## Foreword by Michael Quinn Patton

Gaps are noticed. Ideas are generated to fill those gaps. New ways of doing things emerge. Those new ways, if useful, become approaches and models. They build on past understandings, apply to current challenges, and point the way to future possibilities. That is how innovation happens, how adaptation unfolds, and how progress occurs. But the ultimate test of utility is always conducted by thoughtful and innovative leaders in ongoing organizations who take the risk of finding out if the theory works in practice, if the idea can be translated into meaningful action. UNFPA has taken that pioneering step in applying developmental evaluation to results-based management.

Experienced colleague and long-time friend Mark Cabaj applied his deep knowledge about developmental evaluation in collaboration with UNFPA Evaluation Office to develop this breakthrough quality assessment framework. This tool opens the door for more developmental evaluation applications in institutional settings worldwide.

The future belongs to those who can adapt in complex dynamic systems. Developmental evaluations help navigate the uncertainties and turbulence the future holds. Take a careful look and see if developmental evaluation and this quality assessment framework will support innovative adaptation in your setting and situation as it has at UNFPA.

### **Michael Quinn Patton**

Founder and Director, Utilization-Focused Evaluation

Michael Quinn Patton is author of *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (Guilford Press, 2011), co-editor of *Developmental Evaluation Exemplars: Principles in Practice* (Guilford Press, 2015), and author of *What is Essential in Developmental Evaluation?* *American Journal of Evaluation* (2016. vol. 37, 2, pp. 250-265).

# 1. Introduction

---

The purpose of this guidance is to clarify the concept of quality in developmental evaluations managed by UNFPA. It is intended to increase the understanding of UNFPA evaluation staff, external evaluation consultants, and those contracted to carry out quality assessment of UNFPA developmental evaluations, of how the principles of developmental evaluation and the UNEG evaluation standards translate into UNFPA evaluation quality assurance and assessment work. This in turn will improve the overall quality of developmental evaluations and support the effective use of evaluative evidence for programming and policy decisions.

## 2. Background

---

### 2.1 Evaluations at UNFPA

The UNFPA evaluation function adheres to UNEG norms and standards for evaluation and the definition of evaluation therein. An evaluation is an assessment, conducted as systematically and impartially as possible, of an activity, project, programme strategy, policy, topic, theme, sector, operational area or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency, impact and sustainability. An evaluation should provide credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations and lessons into the decision-making processes of organizations and stakeholders.<sup>4</sup>

Evaluation at UNFPA serves three main purposes in support of the organization's drive to achieve results:

- Evaluation as a means to demonstrate **accountability** to stakeholders on performance in the achievement of development results, and on invested resources.
- Evaluation as supporting **evidence-based decision-making** in order to achieve sustainable development results.
- Evaluation as contributing important **lessons** on how to accelerate implementation of the Programme of Action of the International Conference on Population and Development and how UNFPA can best support the achievement of the Sustainable Development Goals.

Thus, the purpose of the evaluation function at UNFPA is to ensure institutional accountability and learning through high-quality evaluations that are used to support evidence-based decision-making in order to achieve sustainable development results.

The UNFPA evaluation policy<sup>5</sup> and evaluation function therefore place strong emphasis on quality assurance and assessment of evaluation processes and products. In 2019, UNFPA Evaluation Office rolled out a revised evaluation quality assurance and assessment system (EQAA)<sup>6</sup> that explains the concept of and principles underpinning evaluation quality at UNFPA, and details how this is applied/operationalized through assurance and assessment tools. However, given the complexity of developmental evaluations, the current EQAA system needed to be reframed and adapted for the purpose of developmental evaluations, resulting in the development of this guidance.

---

<sup>4</sup> United Nations Evaluation Group. (2016). Norms and Standards for Evaluation. New York: UNEG, p. 10.

<sup>5</sup> <https://www.unfpa.org/admin-resource/unfpa-evaluation-policy-2019>

<sup>6</sup> UNFPA Evaluation Office. (2019). Evaluation Quality at UNFPA. Principles and their application. Available at [https://www.unfpa.org/sites/default/files/admin-resource/EQAA\\_FULL\\_DESIGN.pdf](https://www.unfpa.org/sites/default/files/admin-resource/EQAA_FULL_DESIGN.pdf)

Evaluation quality assessments for all UNFPA evaluations are posted publically alongside the corresponding evaluation report (and management response) on the Evaluation Database.<sup>7</sup>

## 2.2 Developmental evaluation

The UNFPA has recently begun to employ developmental evaluation in its evaluation activities. In 2019, Evaluation Office undertook its first developmental evaluation on results-based management at UNFPA.<sup>8</sup> The evaluation quality assessment of this developmental evaluation is available in Annex 6.

Michael Quinn Patton states that “Developmental evaluation refers to long-term, partnering relationships between evaluators and those engaged in innovative initiatives and development. Developmental evaluation processes include asking evaluative questions in gathering information to provide feedback and support the developmental decision-making and course corrections along the emergent path. The evaluator is part of a team whose members collaborate to conceptualize, design and test new approaches in a long-term, on-going process of continuous improvement, adaptation, and intentional change. The evaluator’s primary function in the team is to elucidate team discussions with evaluative questions, data and logic, and to facilitate data-based assessments and decision-making in the unfolding and developmental process is of innovation”.<sup>9</sup>

The purpose of developmental evaluation is distinct from more traditional formative, summative, and accountability-oriented evaluations in that it aims to inform the creation and ongoing development of interventions (e.g., strategy, model, programme), often in dynamic contexts, that aim to make progress on complex issues. (See Table 1.)

**Table 1: Different types of evaluations**<sup>10</sup>

Evaluation Purpose	Purpose	Niches
<b>Formative</b>	To improve and refine a strategy, model or programme.	<ul style="list-style-type: none"> <li>Continuously improving an existing initiative</li> <li>Preparing a new initiative for a summative evaluation</li> </ul>
<b>Summative</b>	To judge the merit, value, or worth of a programme strategy, model, or programme.	<ul style="list-style-type: none"> <li>Making decisions about whether to (dis) continue funding an initiative</li> <li>Determining whether to scale an innovation</li> </ul>

<sup>7</sup> [https://web2.unfpa.org/public/about/oversight/evaluations/?\\_ga=2.171862947.2002188476.1582831940-545676458.1573245272](https://web2.unfpa.org/public/about/oversight/evaluations/?_ga=2.171862947.2002188476.1582831940-545676458.1573245272)

<sup>8</sup> <https://www.unfpa.org/admin-resource/developmental-evaluation-results-based-management-unfpa>

<sup>9</sup> Quinn Patton, M. (2008). *Utilization-focused evaluation*. (4th ed). Thousand Oaks, Ca.: Sage Publications, p. 45.

<sup>10</sup> Adapted from Quinn Patton (2008), p. 140.

<b>Accountability</b>	To demonstrate that resources are well used, and the initiative is efficiently achieving outcomes.	<ul style="list-style-type: none"> <li>• Routine or end-of-initiative reporting</li> <li>• Periodic audits</li> <li>• Forensic inquiries</li> </ul>
<b>Developmental<sup>11</sup></b>	To inform the creation, development, and ongoing adaptation of an initiative.	<ul style="list-style-type: none"> <li>• Creating a new initiative</li> <li>• Ongoing adaptation of an initiative</li> <li>• Adapting principles of an effective intervention in new contexts</li> <li>• Cross-scale systems change</li> <li>• Rapid response in crisis situations</li> </ul>

Developmental evaluation is also distinct from more traditional evaluation approaches in terms of a variety of evaluation practices related to roles and relationships, standards, and approach to measurement and methods. (See Table 2.)

**Table 2: Comparing developmental evaluation and traditional evaluation practices<sup>12</sup>**

<b>Features</b>	<b>Traditional formative and summative evaluation</b>	<b>Developmental evaluation</b>
<b>Purpose</b>	Supports improvement, summative tests and accountability	Supports development of innovation and adaptationist, dynamic environments
<b>Roles and relationships</b>	Positioned as an outsider to assure independence and objectivity	Positioned as an internal team function integrated into the process of gathering and interpreting data, framing issues, surfacing and testing model development
<b>Accountability</b>	Focused on external authorities and funders based on explicit and pre ordinate criteria	Centered on the innovators' values and commitment to make a difference
<b>Options</b>	Rigorously options-focused, traditional research and disciplinary standards of quality dominate	Utilization-focused options are chosen in service to development use
<b>Measurement</b>	Measure performance and success against predetermined goals and smart outcomes	Develops measures and tracking mechanisms quickly as outcomes emerge. Measures can change during the evaluation as the process unfolds
<b>Evaluation feedback</b>	Detailed formal reports; validated best practices, generalizable across time and space. Can engender fear of failure	Rapid, real-time feedback. Diverse, user-friendly forms of feedback. Evaluation aims to nurture learning

<sup>11</sup> See Annex 1, for a more detailed description of each niche for developmental evaluation.

<sup>12</sup> Quinn Patton, M. (2006). Evaluation for the Way We Work. *The Non-profit Quarterly*, 13(1): 30. Retrieved from <http://www.scribd.com/doc/8233067/Michael-Quinn-Patton-Developmental-Evaluation-2006>

<b>Complexity and uncertainty</b>	Evaluator tries to control design implementation and evaluation process	Learning to respond to lack of control. Staying in touch with what is unfolding and responding accordingly
<b>Standards</b>	Methodological competence and commitment to rigour; independence; credibility with external authorities and funders; analytical and critical thinking.	Methodological flexibility, eclecticism and adaptability. Systems thinking, creative and critical thinking; high tolerance for ambiguity. Open and agile teamwork and people skills. Able to facilitate rigorous evidence-based perspective.

While the concepts and practices of developmental evaluation have spread rapidly across such fields as evaluation, philanthropy, social innovation, international development, they also have resulted in problems of fidelity, misinterpretations, and misuse of the approach.<sup>13</sup> This requires the UNFPA to lay out what a quality developmental evaluation looks like.

---

<sup>13</sup> Quinn Patton, M., McKegg, K., & Wehipeihana, N. (Eds.). (2016). *Developmental Evaluation Exemplars: Principles in Practices*. New York: Guilford Publications, p. 19.

### 3. The concept of evaluation quality at UNFPA

---

The key dimensions of quality for UNFPA developmental evaluations are drawn from two major sources:

- The **four** internationally recognized evaluation standards drawn from UNEG’s Norms and Standards.<sup>14</sup>
- The **eight** principles of effective developmental evaluation, described in *Developmental Evaluation Exemplars: Principles in Practice*.<sup>15</sup>

The UNFPA Evaluation Office has combined these to create eleven quality principles for developmental evaluation: one that is shared (i.e., utility/utilization-focused), seven that are distinct to developmental evaluation, and two specific to UNEG.

#### Shared principle

1. Utility/utilization-focused evaluation

#### Developmental evaluation principles

2. Developmental purpose
3. Evaluation rigor
4. Innovation niche
5. Complexity perspectives
6. Systems thinking
7. Co-creation
8. Timely feedback

#### UNEG standards

9. Independence and impartiality
10. Credibility
11. Gender equality and human rights

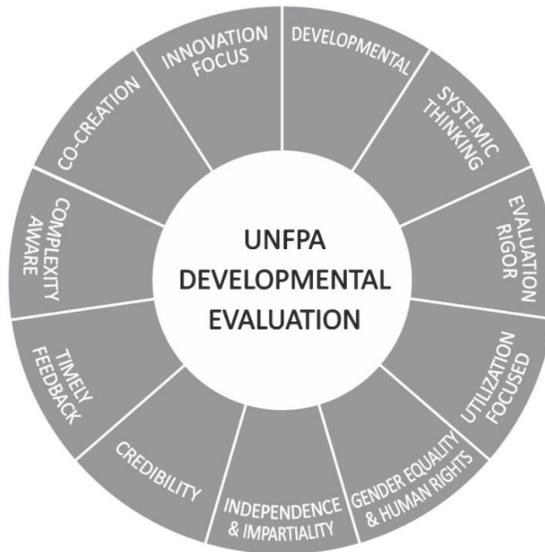
The eleven principles and standards, and minimum quality practices for each, are described ahead.

---

<sup>14</sup> United Nations Population Fund Evaluation Office. (2017, February). Tools and Guidance: Dimensions of evaluation quality at UNFPA. Version 1. New York: UNFPA.

<sup>15</sup> Quinn Patton, McKegg, & Wehipeihana. (2016).

**Figure 1: Eleven quality principles and standards for developmental evaluation**



### 3.1 Principle 1: Utility/utilization-focus

*Focus on intended use by intended user, from beginning to end, facilitating the evaluation process to ensure utility and actual use.*<sup>16</sup>

In order for an evaluation to be useful, it needs to be relevant to the stakeholders for whom it is conducted. It also must be done in a way that encourages a sense of commitment to the process and results: UNFPA evaluation policy states that evaluators should be "... planned and conducted ensuring national ownership and leadership of evaluation processes by rights holders and duty bearers."<sup>17</sup>

The principle of utility and use is thus closely linked to intended use by intended users. Utilization-focused evaluations enhance the potential use of evaluations and provide credible information to support decision-making to inform planning, implementation, and improvements in policies and programmes.<sup>18</sup> Evaluations should be designed and monitored with careful consideration of how the process and product will affect use.

<sup>16</sup> Quinn Patton (2008), p. 299.

<sup>17</sup> United Nations Population Fund Evaluation Office. (2019, January). UNFPA Evaluation Policy. New York: Author, p. 9.

<sup>18</sup> UNFPA Evaluation Policy, p. 6.

Some minimum practices of a **utility/utilization-focused** principle include:

- The evaluation identifies the primary intended users and intended use of the evaluation.
- Evaluators consult with stakeholders throughout the evaluation process to ensure that its objectives are relevant, methods are responsive, and overall process engaging.
- Evaluation communications, both written and oral, are easy to read/understand, concise, and to the point.
- Evaluators take actions to facilitate the use of evaluation findings by primary users in discussions and decision-making processes.

### 3.2 Principle 2: Developmental purpose

*The evaluation aims to illuminate, inform, and support what is being developed, by identifying the nature and patterns of development (innovation, adaptation, systems change), and the implications and consequences of those patterns.<sup>19</sup>*

The aim of developmental evaluation is to develop a new response to a complex challenge. That requires developmental evaluators inform the developments that emerge in that process (e.g., new understandings of the challenge, shift). It also means that developmental evaluation tracks those developments and encourages innovators to ask, “What’s next?”

Some minimum practices of the **developmental** principle of evaluation include:

- Evaluators describe the “base-line” conditions for the initiative/ theme/ system/ programme/ intervention, including a description of the challenge or problem that users/management would like to address; the key ideas and features of their initiative (e.g., logic model, theory of change, strategy); anticipated results and criteria for success; and key considerations that shape how they will proceed (e.g., key questions, resources constraints).
- The evaluation tracks the major developments to emerge in the initiative, including forks in the road, major decision points, significance shifts in context, and results and learnings.
- The evaluation documents the major changes in the initiative’s intent, design, and implementation and expected results, including the rationale, data, and evidence for each major change.
- The evaluation uses evaluation findings to describe the implications for further development of the intervention (rather than conclusions or recommendations), often framed as options, scenarios or pathways.
- The evaluation design evolves to capture new evaluative questions on the part of the initiative’s primary users, triggered by new insights, shifts in context, and the evolution of the initiative.

---

<sup>19</sup> Quinn Patton (2008), p. 291.

### 3.3 Principle 3: Evaluation rigor

*Ask probing evaluation questions, think and engage evaluatively; question assumptions; apply evaluation logic; use appropriate methods; and stay empirically grounded – that is, rigorously gather, interpret and report data.<sup>20</sup>*

Developmental evaluation requires rigorous and evaluative thinking be infused throughout the entire evaluation process, including framing the hypothesis or intervention; developing evaluation questions; choices about how to gather, analyse, and interpret data to answer those questions; and developing conclusions and possible scenarios/ways forward/ implications for further development. It promotes a culture and practice of inquiry that encourages primary users to critically assess and reflect on their work and to question, test, and re-examine their results and conclusions about it, as well as the implications for further development.

Some minimum practices of **evaluation rigor** principle of evaluation include:

- The evaluation clearly lays out the methods employed, including key evaluation questions and the rationale for the evaluation design.
- Methods are situationally responsive to context, including the time and resource constraints, values, cultural and ethical considerations, and other key factors that shape the evaluation design.
- The evaluation identifies the limitations of the methods employed (e.g., bias, data gaps) and how these limitations affect the quality, design and implementation of the evaluation.
- Evaluation findings, statements are empirically based.
- The evaluation design includes processes that allows the evaluation questions and methods to evolve to respond to new questions and influenced as the intervention and evaluation evolve.

### 3.4 Principle 4: Innovation niche

*Elucidate how the change process and results being evaluated involved innovation and adaptation, the niche of developmental evaluation.<sup>21</sup>*

The immediate outcome of a developmental initiative is to find new, better, and/or different solutions to address complex issues. These may be reflected in new programmatic interventions, such as programmes or services, and/or in larger systemic and cultural responses, such as new policies, standards, resources flows, mindsets, or relationships. Developmental evaluation encourages and supports creative thinking and the processes required to develop new approaches as well as the evaluative thinking necessary to inform and assess them.

Some minimum practices of **innovation** principle of developmental evaluation include:

- Developmental evaluators describe the elements of new, innovative, or adaptive practice(s), including their significance, and the extent to which they depart from current practices.
- The evaluation develops and tests quick iterations of some or all parts of the emerging innovation, drawing conclusions about its early results, including an assessment of the extent to which the innovation is likely to contribute to desired change.

---

<sup>20</sup> Quinn Patton (2008), p. 296.

<sup>21</sup> Quinn Patton (2008), p. 301.

- The evaluation surfaces new questions and insights about the nature of the challenge that the group is trying to address, and what does and does not work with the emerging innovation.
- The identification and assessment of the resistance to the innovation that initial stakeholders may experience from their peers and/or by stakeholders in the broader system in which they operate, including data on the source and rationale for that resistance, as well as the implications for further development of the innovation.

### 3.5 Principle 5: Complexity aware

*Understand and interpret development through the lens of complexity and conduct the evaluation accordingly. This means using complexity premises and dynamics to make sense of the problems being addressed: to guide innovation, adaptation, and system change strategies; to interpret what is developed: to adapt the evaluation design as needed; and to analyze emergent finding.<sup>22</sup>*

Developmental initiatives operate within complex environments, characterized by “non-linear interactions and effects, turbulence, and emergence” and stakeholders operating with different values, interests, and positions about what can and should be pursued and done. This unpredictability, limited control, and fluid stakeholder dynamics make the development process emergent and the results uncertain.

Some of the minimum practices of the **complexity** principle include:

- A demonstrated effort to seek out, spot, and make sense of the intended and unintended effects of the initiative, assess their consequence for the innovation, and ensure both are seriously considered in discussions of further development and adaptation of the initiative.
- An emphasis on assessing the “contribution” of stakeholders to observed changes, rather than on proving the attribution of their activities to that change, in the face of complex and dynamic cause-and-effect relationships, including continuously tracking insights about what factors initiative stakeholders can and cannot control and influence.
- Tracking the paradoxes, dilemmas, and stakeholder agreements and disagreements that emerge in the initiative, the framing of those tensions, and identifying how they influence its evolving intent, design and delivery, and outcomes.

### 3.6 Principle 6: Systemic thinking

*Think systemically throughout, being attentive to inter-relationships, perspectives come with boundaries, and other key aspects of the social system and context within which the innovation is being developed and the evaluation is being conducted.*

Most of the complex challenges that stakeholders of UNFPA implemented initiatives/interventions are trying to address are deeply systemic in nature, held in place by larger structures, cultures, and power. Developmental evaluation should enable users to “zoom in” on tracking the emergence, learning, and results of their initiative, as well as “zoom out” to monitor and understand the larger systemic context in which they operate.

---

<sup>22</sup> Quinn Patton (2008), p. 304.

Some of the minimum practices to manifest the **systemic thinking** principle include:

- Mapping and describing the elements of the system(s) in which the initiative operates (e.g., boundaries, relationships, perspectives), as well as the areas of the system(s) it seeks to change (e.g. policies, resource flows, structures).
- Tracking how changes in larger systems and context influence the users' expected results, approach (e.g., theory of change, strategy, design), and perceptions of opportunities or barriers for development and results.
- Intentionally seeking out the diverse perspectives of stakeholders on the characteristics of the systems they are trying to change as well as the nature and significance of systems change triggered by the initiative.

### **3.7 Principle 7: Co-creation**

*Develop the innovation and evaluation together – interwoven, interdependent, iterative, and co-created – such that the developmental evaluation becomes part of the change process.*

In traditional evaluation, evaluators are perceived to be outside the evaluation process. In developmental evaluation, by contrast, evaluation is considered an inside function, where evaluators are part of the team, helping stakeholders to integrate evaluative thinking and processes into an ongoing process of action and reflection. While the commitment to co-creation generates some tensions with the important principle of impartiality and independence of the evaluation, it is critical that evaluative practices become a seamless part of the organic process of development.

Some of the minimum practices of the **co-creation** principle include:

- Evaluation is positioned as an internal team function, with evaluator(s) able and encouraged to participate in design discussions, planning sessions, and decision-making moments.
- Evaluators work closely with stakeholders to integrate evaluative practices and data into the conceptualization, design, and testing of their initiative.
- Evaluators meaningfully engage stakeholders to surface evaluative questions, gather and make sense of data, draw conclusions, surface implications for further development (e.g. options, scenarios, new questions), and facilitate the use of findings to inform decisions.

### **3.8 Principle 8: Timely feedback**

*Time feedback to inform ongoing adaptations as needs, findings, and insights emerge, rather than only at pre-determined times (e.g., quarterly, or at mid-term and end-of-project).*

To be effective, developmental evaluators must constantly track opportunities to surface evaluation questions, test quick iterations, and facilitate learning moments as they emerge in “real time”. They must also be diligent in integrating these insights into the ongoing reflection and decisions of primary users. Developmental evaluation findings have a short “shelf life” and must be developed, delivered, and used in real time in order to be of practical value.

Some of the minimum practices that reflect a commitment to **timely feedback** include:

- A commitment to take mid-term and end-of-initiative reporting seriously for contracting purposes, but with a willingness and administrative capability to adjust the content, size, and scheduling of these reports, should primary users and evaluators feel that such measures could enhance their timeliness, and therefore, their usefulness.
- The use of varied, real-time, feedback formats (e.g., memos, presentations, learning briefs, PowerPoint documents, evaluative exercises) to complement more traditional reporting formats that primary users find useful and appropriate.
- Careful attention to the process of drawing conclusions and making evaluative judgments in rapid-feedback situations in which the collection, analysis, and interpretation of data is unavoidably constrained by limited time and resources.

### 3.9 Standard 9: Independence and impartiality

*Independence and impartiality are fundamental to attain credibility and utility of an evaluation. Independence entails the ability of evaluators to work without undue influence by any party and includes freedom to choose the design of the evaluation and to select suitable methods and tools for data collection and analysis.*

Key elements of impartiality are objectivity, professional integrity, and absence of bias. Both independence and impartiality require transparency and should exist at all stages of the evaluation process, from the planning stage and the formulation of mandate and scope, through to conducting the evaluation and formulating findings and scenarios/ways forward.

Some of the minimum practices that reflect a commitment to **independence and impartiality** include:

- Evaluators have final authority on all decisions related to the design and delivery of an evaluation, as long as they meet or exceed UNEG/UNFPA principles and standards.
- Evaluators are committed to making statements, claims, findings and implications for further development that reflect a balanced analysis and interpretation of the data/evidence.
- Evaluators have “full discretion to discretely submit evaluation reports to the appropriate level of decision-making and that they should report directly to an organizational governing body and/or the executive head” and that they have realistic opportunities to exercise that discretion.<sup>23</sup>
- That all UNFPA evaluation reports are publicly available on the Evaluation Office website.

### 3.10 Standard 10: Credibility

The credibility of the developmental evaluation depends on the ability of the evaluators, primary users and commissioners to adhere to the other nine quality principles. This is particularly true of the principles of utility, rigor (including the commitment to gender equality and adherence to ethical principles), impartiality and independence.

---

<sup>23</sup> UNEG (2016), p. 11.

The credibility of the developmental evaluation also depends on the following attributes:

- The real and perceived competence of the evaluators and evaluation managers/ commissioners to structure and administer evaluation contracts and budgets in a way that supports, rather than hinder, the evaluation.
- The real and perceived evaluation expertise of the evaluation team, as well as, their knowledge of the domain, intervention and context.
- The demonstrated commitment of the evaluation team to help develop the innovation or intervention – rather than passively support or even judge the efforts - through their evaluative questions, real time feedback, and rigorous sense-making.

### **3.11 Standard 11: Gender equality and human rights**

The UNEG Guidance Document on Gender Equality and Human Rights confirms that the “promotion and protection of human rights and gender equality are central principles to the mandate of the UN and all UN agencies.”<sup>24</sup> The document provides guidance and options for integrating these dimensions into the evaluation work of UN agencies. It lays out a comprehensive list of practices to promote this integration in the preparation, design, and implementation of an evaluation, stating the following:

- Integrating gender equality and human rights criteria, evaluation questions, and indicators into an evaluation terms of reference, as well as finalizing criteria for the preferred evaluation team, if appropriate.
- Selecting and employing methods to ensure that the gender equality and human rights aspects of the intervention are identified and analysed during the evaluation process, if appropriate.
- Selecting and employing methods to ensure that the intervention’s gender equality and human rights aspects are identified and analysed during the evaluation process, including in collection and analysis of data, if appropriate.
- Including gender equality and human rights issues in the findings and implications for further development, including how these will affect different stakeholders of the intervention.

---

<sup>24</sup> United Nations Evaluation Group. (2011). Integrating Human Rights and Gender Equality in Evaluations: Towards UNEG Guidance Document. Retrieved from: <file:///C:/Users/Mark/Downloads/HRGE%20Handbook.pdf>

## 4. Tensions in ensuring quality of developmental evaluations

---

A number of tensions are unavoidable in the application of developmental evaluation principles in UNFPA-managed evaluations and will affect the quality of an evaluation. These tensions cannot be (re)solved; rather, they must be approached as design challenges and managed creatively.

### 4.1 Major tensions

There are three major tensions in ensuring quality in UNFPA-managed developmental evaluations. They relate to (a) purpose, (b) relationships, and (c) administration.

#### 4.1.1 Purpose tensions

The major purpose tension is between developmental evaluation as an internal process to support the development and adaptation of an intervention and the more traditional use of evaluation as a mechanism to report and account to a senior or external body.

On the one hand, the aim of developmental evaluation is to provide initiative stakeholders with real-time feedback on the results of their efforts, to generate new learnings about the challenges they are trying to address, to identify and critically reflect on the strengths and weaknesses in their work, and to enable data-informed adaptations to their goals and overall approach. This includes the need for the primary users of the evaluation to co-draw upon the findings, implications and options surfaced in the process to make decision on where and how to further develop the intervention, rather than rely on the recommendations of the evaluation team.

On the other hand, UNFPA's evaluation practices and culture are organized around a more traditional formative and summative evaluation approach, and embrace the principle of publicly reporting evaluation findings and results, typically to senior decision-makers and/or external bodies, pointing to concrete conclusions and recommendations. This traditional orientation can make it difficult for the participants and evaluators in developmental situations to embrace the critical self-reflection central to the success of developmental evaluation.

The design challenge for stakeholders involved in UNFPA developmental evaluations is the following: How can UNFPA developmental evaluations be managed in a way that maintains an internal focus on developing and adapting an intervention while operating in an institutional environment designed for more traditional, publicly reported, externally oriented evaluation?

#### **4.1.2 Tensions in evaluator relationships**

Evaluators involved in developmental contexts must navigate the tension between several key quality principles that relate to the relationships between evaluators and primary users: co-creation, impartiality and independence.

On the one hand, utility-oriented developmental evaluation requires that evaluators co-create the evaluation design with primary users. This means selecting data sources and methods that reflect users' "philosophical and organizational" preferences, employing facilitated processes for shared sense-making of findings, and (often) facilitated or co-development of the implications for the further development of the intervention (e.g. leverage points, options, scenarios, additional questions).<sup>25</sup>

On the other hand, evaluators must be able to operate without undue influence by evaluation stakeholders (e.g., primary users, commissioners, beneficiaries), including in the choice of evaluation methods for data collection and analysis, the discussion of findings, and the preparation of conclusions and way forward.

The design challenge for stakeholders involved in UNFPA developmental evaluations is: How can developmental evaluators meaningfully co-create evaluation designs, and engage evaluation users in other utilization-focused practices, while ensuring their own ongoing independence and impartiality in the process?

#### **4.1.3 Tensions in administration**

The UNFPA Evaluation Office and contracted developmental evaluators must deal with tensions in the procurement and administration of developmental evaluation activities.

On the one hand, in order to be useful, developmental evaluation designs must be able to adapt in order to reflect the initiative stakeholder's evolving goals, questions, and context. If not, "rigor-mortis" sets in, reducing the utility of the design. In these cases, the initiative stakeholders either ignore the evaluation results entirely or, in some cases, undermine the development effort by pressuring stakeholders to a no-longer relevant evaluation design.

On the other hand, UNFPA Evaluations are guided by policies, guidelines and practice for commissioning evaluations that require that evaluation contracts carefully identify evaluation activities and deliverables, key timelines for completion, and disbursement schedules that are organized around the submission and approval of key evaluation outputs or products (e.g. reports).

The design challenge for UNFPA developmental evaluation is the following: How can UNFPA Evaluation Office and evaluators develop a preliminary evaluation design that was complete enough to finalize funding and an evaluation contract, yet be flexible enough to adapt in real time to reflect stakeholders' ever-evolving evaluation needs?

---

<sup>25</sup> Quinn Patton (2006), p. 30.

## **4.2 Evaluating the management of tensions**

A fulsome assessment of the quality of a developmental evaluation commissioned by UNFPA will include an assessment of how well initiative stakeholders – particularly developmental evaluators and UNFPA staff – have managed these tensions from the beginning to end of the evaluation.

This will include asking them to identify (a) tensions that emerged in the evaluation, (b) the practices, tactics or measures employed to manage the tension, and (c) a rating on how well they feel they managed the tension, as well as the rationale for that rating.

## 5. Quality assurance and assessment of developmental evaluations at UNFPA

---

The UNFPA Evaluation Office has developed this new Evaluation Quality Assurance and Assessment Framework to assist the Office and third parties in assessing the quality of developmental evaluations completed for the agency.

The framework has three objectives:

1. To inform developmental evaluators about the key dimensions of developmental evaluation quality and how it will be assessed by the UNFPA Evaluation Office.
2. To provide guidance to persons responsible for evaluating the quality of developmental evaluators commissioned on behalf of the UNFPA Evaluation Office, including:
  - a. Establishing criteria for different dimensions of quality (unsatisfactory, fair, good, very good);
  - b. Ensure consistency amongst quality assessment reviews.
3. To inform the ongoing efforts of the UNFPA Evaluation Office to strengthen the capacity and culture of the UNFPA in regard to developmental evaluation.

This quality review process is designed to be completed by an external independent quality assessor with approximately three days of effort. This will provide UNFPA Evaluation Office staff and developmental evaluation stakeholders with a high-level assessment of the quality of a developmental evaluation. However, the UNFPA Office staff can also scale up the framework to complete a broader and/or deeper assessment if necessary, by interviewing a greater number and variety of primary users.

The developmental evaluation quality assessment framework is a work in progress. UNFPA Evaluation Office will periodically review and upgrade the approach based on its application and on what has been learned from various developmental evaluation initiatives.

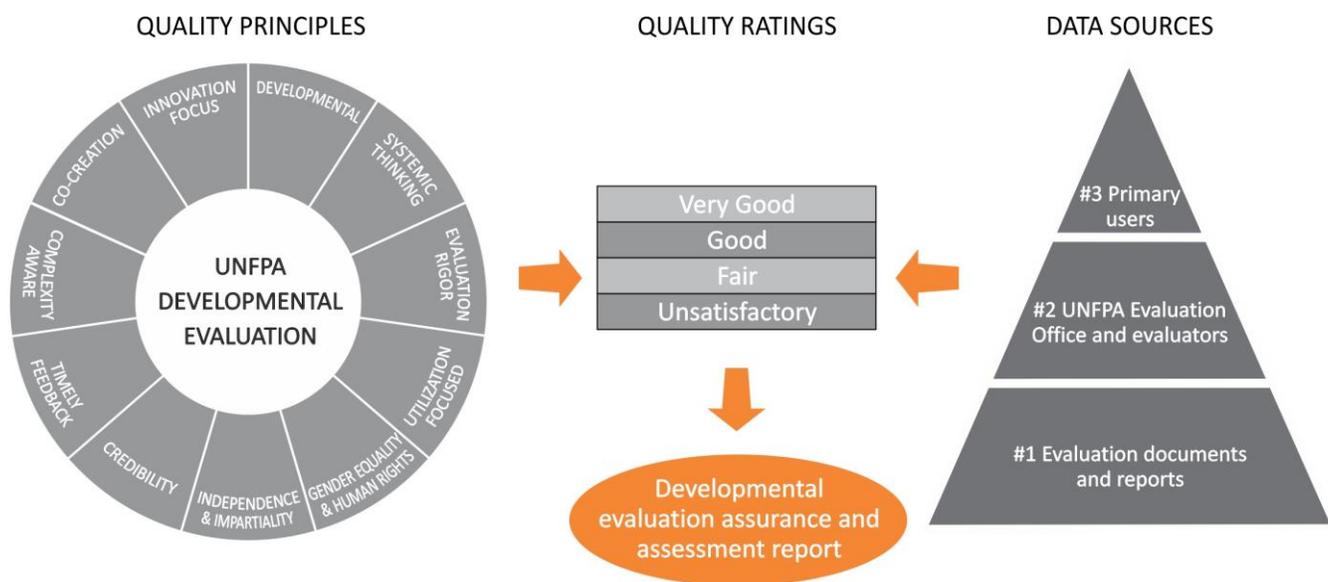
## 6. The approach

The approach to assessing the quality of UNFPA developmental evaluations is organized around six key features:

- A combination of developmental evaluation principles and UNEG evaluation standards as described above
- An emphasis both on (a) evaluation reports (products) and (b) evaluation process
- Rating the evaluation’s adherence to minimum practices for each principle/standard
- Three data sources and methods
- A five-step process
- Options for scaling up

Each of the key features is described in turn below.

**Figure 2: Approach to developmental evaluation quality assurance and assessment**



### 6.1 Developmental evaluation principles and standards

The manifestation of quality principles/standards for developmental evaluation is highly context-sensitive: how effectively they are applied by evaluators and initiative stakeholders depends on a variety of contextual factors, including institutional capacity, length of initiative, the nature of the initiative being evaluated, and stakeholder dynamics, among other factors.

To this extent, the assessment of the quality of UNFPA developmental evaluations will focus on the extent to which the evaluation process and products were able to employ minimum characteristics or practices for each of the eleven quality principles and standards.

### 6.2 Developmental evaluation products and process

In developmental evaluations, the quality of the evaluation is expressed through quality evaluation **products** (e.g., formal reports, real-time learning briefs, PowerPoint presentations, developmental evaluation memos, feedback notes) and **processes** (e.g., relationships, evaluative discussions, co-review of evaluation feedback, data-informed decision-making).

A conventional approach to quality control in UNFPA developmental evaluations that rest entirely on assessing the quality of traditional final reports would have two perverse consequences:

- It would examine only one element – a very modest one at that, sometimes – of the developmental evaluation experience (i.e., a formal report). It thereby would miss important day-to-day evaluation interactions and multiple ways of providing real-time feedback that comprise the minimum of developmental evaluation.
- It would over time incentivize developmental evaluators to place a disproportionate effort on developing “quality control” for the UNFPA Evaluators, while dramatically diminishing the practical value of the developmental evaluation to its primary users.

To ensure a complete and more accurate assessment of the quality of UNFPA developmental evaluations, the UNFPA Evaluation Office Quality Framework covers both **the product and process** of the developmental evaluations under examination.

### 6.3 Rating adherence to minimum practices

The assessment of UNFPA developmental evaluations is be done by rating the extent to which the evaluation products and processes exhibit quality standards for each of the eleven quality principles/standards. The rating system includes four possible ratings:

**Table 3: Rating system**

<b>Very Good</b>	Strong, above average, best practice adherence to standards
<b>Good</b>	Meets a satisfactory, acceptable standard
<b>Adequate</b>	With some weaknesses, still acceptable.
<b>Unsatisfactory</b>	Weak, does not meet minimum standards

## 6.4 Data sources and methods

The quality assessment process of UNFPA developmental evaluations should be done by a third-party external independent assessor using three methods.

**Table 4: Methods for assessing the quality of developmental evaluations**

	Description	Strengths	Limitations	Implication for quality assessment
1	Assessment made based on a review of real-time, and final developmental evaluation reports or products (e.g. memos, presentation decks, etc.)	Requires only modest investment of time and resources	Difficult to capture the quality of the evaluation ‘process’ and contextual factors that shape the developmental evaluation; does not incorporate feedback of evaluation stakeholders	Provides a partial assessment of the quality of the developmental evaluation (i.e. evaluation products) and misses process and context
2	Assessment made based on key informant interviews with select number of primary intended users, evaluators and commissioners	Can explore the quality of the evaluation process based on stakeholder perceptions, as well as contextual factors that shape the evaluation process	Requires a greater investment of time and resources; relies on feedback of small sample size	Provides a more rounded assessment of the quality of the developmental evaluation, both its products and process
3	Assessment made based on interviews with UNFPA staff and evaluation managers	Allows for a review of the management and administrative context for the evaluation	Offers only the perspective of administrators and evaluators	Widens the understanding of the administrative factors that shaped developmental evaluation quality

## 6.5 The steps for assessing the quality of developmental evaluations

Assessing the quality of UNFPA developmental evaluations is a process with three major steps.

### Step 1 - Scoping

Confirm and finalize the developmental evaluation to be assessed, key data sources, and contacts (e.g., evaluation products, list of primary evaluation users, evaluation team and UNFPA staff), deliverables, key milestones, and budget.

## Step 2 – Implementation

Complete the following five tasks:

**Task 1:** Review and rate developmental evaluation products (e.g., reports, presentations) (See annex 2)

**Task 2:** Interview and/or administer user surveys (See annex 3)

**Task 3:** Interview developmental evaluator and UNFPA evaluation managers (See annex 4)

**Task 4:** Prepare the quality assessment report. Compile ratings from all three sources, and summarize the rationale for each rating (See annex 5)

**Task 5:** Submit the quality assessment report

## Step 3: UNFPA Evaluation Office review and sign off

The time required to complete the steps will vary for each evaluation, but typically should be completed within one month.

## 6.6 Options for scaling up the quality assessment

The approach to assessing the quality of UNFPA developmental evaluations outlines *a minimum level of effort* required to obtain *high-level, rounded assessment* of the evaluation. However, UNFPA staff and developmental evaluation stakeholders may choose to scale up the quality assessment to go deeper or broader.

There are (at least) two reasons for deciding to scale up, such as the following:

- To accommodate an initiative that is unusual for its size, complexity, or length, and/or large numbers of evaluation users and evaluators.
- To address the high stakes that the evaluation involves, necessitating a more detailed review of developmental evaluation quality.

There are (at least) five ways to scale up this approach:

1. Expand the number of users, evaluators, and UNFPA evaluation managers interviewed/surveyed.
2. Include follow-up interviews with additional selected survey participants to explore key themes in more depth.
3. Introduce new data sources and methods (e.g., peer reviews).
4. Explore particular aspects of the developmental evaluation in greater depth (e.g., a critical incident, set of decisions, specific outcome).
5. Undertake a process of shared sense-making amongst developmental evaluation stakeholders, rather than one carried out only by the developmental evaluation quality assessor.

The UNFPA Evaluation managers will determine if and how the quality assurance might be scaled up in consultation with the developmental evaluation stakeholders, and reflect this commitment in a larger evaluation budget, longer timeline, and a person or firm to complete the quality assessment.

## Annexes

### Annex 1: Niches for developmental evaluation

Table: The five types of developmental evaluation: variations in the nature of innovation and adaptation

Focus on development	Nature of the innovation	Contribution of developmental evaluation	Key concepts
<b>1. Developing (creating/inventing) a new intervention aimed at addressing significant problems</b>	Creating an original approach, or inventing a new intervention and or innovative programme; the emphasis is on originality within a complex context	Clarifying the elements of the innovation; examining effectiveness, unanticipated consequences, and the emerging dynamics of the innovation	<ul style="list-style-type: none"> <li>• Invention</li> <li>• Original</li> <li>• New</li> <li>• Significantly different and important in identifiable ways</li> </ul>
<b>2. Ongoing adaptive development</b>	Innovatively adapting an existing intervention, approach, or programme to changing conditions, new knowledge, and new clientele	Clarifying the nature of the adaptive innovation: what is carried forward, what is changed, how these interact. The consequences of ongoing innovative adaptation as a way to engage in change through trial and error, double-loop learning	<ul style="list-style-type: none"> <li>• Adapting an existing initiative or programme</li> <li>• Changes that go beyond marginal improvements</li> <li>• Ongoing innovation</li> </ul>
<b>3. Developing greater impact by adapting validated innovation principles and practices to a new context (scaling)</b>	The adaptation of principles-based practices from one context to another, or to a larger context. This means that what has been established in one context is experienced as an innovation in another context	Clarifying and elaborating the ways in which different contexts affect adaptive innovation: the degree, nature, and consequences of adaptive innovation from context to context as ideas and approaches are shared and spread	<ul style="list-style-type: none"> <li>• Scaling, expanding options by context</li> <li>• Adapting principles to a new context</li> <li>• Not replicating a model or recipe</li> </ul>
<b>4. Developing changes in and across systems</b>	Innovation through changed relationships, perspectives, and boundaries within and across systems	Tracking, mapping, and interpreting systems changes both within and across systems. Supporting adaptive innovation responses as	<ul style="list-style-type: none"> <li>• Systems as a focus of change</li> <li>• Complex dynamics in play</li> </ul>

		systems changes become manifest	
<b>5. Developing rapid responses in crisis situations</b>	Building while flying. Rapid adaptive innovations in the face of humanitarian, political, social, and economic crisis	Tracking, documenting, and providing real-time feedback about emerging challenges, urgent needs, flow resources, and aligning interventions in highly turbulent, uncertain conditions	<ul style="list-style-type: none"> <li>• Urgent</li> <li>• Real time feedback</li> <li>• Simultaneous planning, implementation, and evaluation</li> </ul>
<b>Innovation can take the form of initiatives, programmes, projects, policies, collaboration, and interventions. The structure or form is not what makes something a social innovation. Rather, that is determined by the degree and nature of change involved, compared to the existing situation</b>			

## **Annex 2: Developmental evaluation quality assessment grid for reports**

The first data source and method employed in the UNFPA quality assessment process is a third-party review of the final developmental evaluation reports submitted to UNFPA Evaluation Office.

### **1. Introduction**

The UNFPA Evaluation Office has developed an evaluation quality assessment template for developmental evaluations to assist external quality assessors in assessing the quality of developmental evaluation reports and other communication products. The template builds on the quality assessment grid developed for more traditional UNFPA evaluations and is adapted to reflect the unique purpose and features of developmental evaluation.

### **2. Dimensions of quality and minimum practices**

The table below is organized around the eleven quality principles and standards of developmental evaluation at UNFPA. For each dimension of quality, a number of minimum practices have been added to illustrate how the principle should – at a minimum – be applied.

The table includes columns where quality assessors can indicate the extent to which the minimum practices have been applied (poor, adequate, good, very good).



**Organizational Unit:**

**Year of Reports:**

**Title of Evaluation  
Deliverables**

**Overall quality of  
reports:**

**Date of assessment:**

**Overall comments:**

**Assessment Levels**

**Very Good** strong, above average, best practice

**Good** satisfactory, respectable of adequate quality

**Fair** with some weaknesses, still acceptable

**Unsatisfactory** weak, does not meet minimal quality standards

<b>Quality Assessment Criteria</b>	<i>Insert <u>assessment level</u> followed by main <u>comments</u></i>	
<b>I. UTILIZATION FOCUSED</b>	<b>Assessment Level:</b>	
<b><i>1.1 To ensure the evaluation reports are focused on primary intended users.</i></b>	<b>Rating</b>	
1.1.1 The evaluation reports clearly describe the 'primary users' of the evaluation and the 'primary uses' of the evaluation process and findings.		
1.1.2. The evaluation reports describe the major evaluation questions that primary intended users would like to answer in the evaluation.		
<b><i>1.2 To ensure evaluation reports are comprehensive and user-friendly</i></b>	<b>Rating</b>	
1.2.1. The evaluation reports are easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors.		

<p>1.2.2. The evaluation reports are of a reasonable length (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations).</p>		
<p>1.2.3. The evaluation reports are structured in a logical way: there is a clear distinction made between analysis/findings, discussions, and implications for further development.</p>		
<p>1.2.4. The reports include critical background information: e.g. the ToRs; a bibliography; a list of interviewees; an inquiry framework; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process.</p>		
<p><b>1.3 Executive summary</b></p>		<b>Rating</b>
<p>1.3.1. The final DE Report includes an executive summary written as a stand-alone section and presenting the main results of the evaluation.</p>		
<p>1.3.2. The executive summary is clearly structured and includes the major points related to the evaluation as they appear in the DE Report (purpose, methodology, highlights from analysis, and implications for further development).</p>		
<p>1.3.3. The executive summary is reasonably concise (e.g. with a maximum length of 5 pages).</p>		
<p><b>1.4 Evidence of Use</b></p>		<b>Rating</b>
<p>1.4.1. The evaluation reports provide evidence that the evaluation findings and process are being used by the primary users, including one or more of the following three types: (1) conceptual use - leading to new ways of perceiving the challenges, (2) process use - increasing capacity to learn, curiosity, confidence, and (3) instrumental use - using data to inform decisions, new actions or behaviour change.</p>		
<p><b>2. DEVELOPMENTAL PURPOSE</b></p>		<b>Assessment Level:</b>

<b>2.1 Describing Base-Line Conditions</b>		
2.1.1 The development and institutional context of the evaluation are clearly described and the constraints explained.		
2.1.2 The evaluation reports describe the theme/ system/ strategy/ programme/ intervention's rationale, hypothesis, logic and/or theory of change, and assess the adequacy of these.		
<b>2.2 Tracking Developments</b>	<b>Rating</b>	
2.2.1 The evaluation reports document the most significant developments to emerge (i.e. forks in the road, shifts in context, results, etc.) and assess the implications for further development and adaptation.		<b>Good</b>
2.2.2. The evaluation reports document the significant adaptations to the original theme/ system/ programme/ intervention (e.g. rationale, goals, results expectations, hypothesis, logic and/or theory of change) informed by the evaluation feedback, including offering the evidence and/or rationale upon which the adaptations have been made?		
<b>3. EVALUATION RIGOR</b>	<b>Assessment Level:</b>	
<b>3.1 To ensure a rigorous design and methodology</b>	<b>Rating</b>	

3.1.1 The evaluation/inquiry framework - and its link to the primary evaluation questions - are clearly described.		
3.1.2 The tools for data collection and analysis are described and their choice justified.		
3.1.3. The methodological limitations are acknowledged and their effect on the evaluation described.		
<b>3.2 To ensure quality of data and robust data collection processes</b>	<b>Rating</b>	
3.2.1 The evaluation design seeks to triangulate data collected where appropriate and possible.		
3.2.2 The evaluation reports identify and make use of reliable qualitative and quantitative data sources.		
3.2.3 The evaluation reports identify possible limitations (bias, data gaps etc.) in primary and secondary data sources and, if relevant, explain what was done to minimize such issues.		
<b>3.3 To ensure sound analysis and credible findings.</b>	<b>Rating</b>	
3.3.1 The findings are substantiated by evidence.		
3.3.2 The analysis and findings are presented against the evaluation questions.		

<p>3.3.7 The analysis and findings are presented against contextual factors.</p>		
<p><b>3.4 To assess the validity of the implications for further development of the intervention emerging from the findings.</b></p>		<b>Rating</b>
<p>3.4.1 The implications for further development (e.g. new questions, working conclusions, forks in the road, options, leverage points) flow clearly from the findings?</p>		
<p>3.4.2 The implications for further development (e.g. new questions, working conclusions, forks in the road, options, leverage points) are clearly written and framed in a way that informs primary users' choices about the next iteration of the development of their innovation.</p>		
<p>3.4.3 The implications for further development (e.g. new questions, working conclusions, forks in the road, options, leverage points) go beyond the findings and provide a deeper understanding of the underlying issues of the programme/initiative/system being evaluated.</p>		
<p>3.4.4 The implications for further development proposed appear balanced, impartial and convey the evaluators' unbiased perspective.</p>		
<p><b>3.5 To assess the flexibility and adaptability of the evaluation design.</b></p>		<b>Rating</b>
<p>3.5.1 The evaluation reports describe the processes evaluators and intervention stakeholders employ to reflect on and adapt the evaluation design to respond to new evaluation questions and other factors that require changes to the original evaluation design (e.g. users, questions, methods).</p>		

<p>3.5.2 The evaluation reports describe changes to the design that emerge during the evaluation and the rationale for each major change.</p>		
<p><b>4. INNOVATION NICHE</b></p>	<p><b>Assessment Level:</b></p>	
<p>4.1 The evaluation describes the elements of the theme/ system/ programme/ intervention/ strategy/ model or practice being developed, as well as stakeholders' perceptions of the significance of the innovations/ change/ transformation.</p>		
<p>4.2 The evaluation identifies new questions and insights that emerged through the process, including those related to (a) the nature of the challenge stakeholders are trying to address, (b) the relevance, strengths and limitations of the innovation/ change/ transformation, and (c) the context in which they are operating.</p>		
<p>4.3 The evaluation identifies and assesses the sources of resistance to the innovation/ change/ transformation (if any), including the rationale for this resistance and implications for further development and adaptation of the intervention.</p>		
<p><b>5. COMPLEXITY PERSPECTIVES</b></p>	<p><b>Assessment Level:</b></p>	
<p>5.1 The evaluation reports identify, describe and assess the implications of both intended and unintended effects and consequences of the innovation/ change/ transformation in the theme/ system/ strategy/ programme/ intervention.</p>		

<p>5.2 The evaluation reports acknowledge the complex cause-and-effect relationships associated with the intervention and tracked those factors and stakeholders new insights about which factors they could control and/or influence, and assessed the implications for the intervention.</p>		
<p>5.3 The evaluation reports track the tensions, dilemmas, paradoxes, and wicked questions related to the theme/ system/ strategy/ programme/ intervention and their implications for further innovation/ change/ transformation.</p>		
<p><b>6. SYSTEMS THINKING</b></p>	<p><b>Assessment Level:</b></p>	
<p>6.1 The evaluation reports describe the system(s) in which the theme/ system/ strategy/ programme/ intervention is embedded and/or trying to change in its baseline assessment (i.e. its boundaries, main actors, important relationships, key dynamics).</p>		
<p>6.2 The evaluation reports track how changes in the larger systems and context in which the theme/ system/ strategy/ programme/ intervention was unfolding create opportunities and barriers for innovation/ change/ transformation, and influence the theme/ system/ strategy/ programme/ intervention's goals, approach or design.</p>		
<p>6.3 The evaluation reports highlight the diverse perspectives of stakeholders in describing the intervention, the systems in which it operates, and its major developments, including where there was agreement and disagreement across perspectives.</p>		
<p><b>7. CO-CREATION</b></p>	<p><b>Assessment Level:</b></p>	

<p>7.1 The evaluation reports describe how the evaluators/evaluation team were positioned as part of the primary users/stakeholders involved in planning, management and implementation of activities.</p>		
<p>7.2 The evaluation reports describe how they engaged primary intended users in the development of evaluation questions, collecting, analyzing and interpreting data, and surfacing implications for further development and adaptation.</p>		
<p><b>8. TIMELY FEEDBACK</b></p>	<p><b>Assessment Level:</b></p>	
<p>8.1 The evaluation employs a variety of reporting/feedback formats (e.g. memos, PPT presentations, thought pieces, learning briefs) to provide real time evaluative feedback to the primary users.</p>		
<p>8.2 The evaluation reports describe if and how the timing of the reports submission coincided with formal/structured and/or spontaneous moments of reflection and decision-making by the primary intended users.</p>		
<p>8.3 The evaluation reports point out if and how the real time nature of providing feedback influenced the rigor of the evaluation (e.g. Data collection, sampling, analysis, interpretations, leverage points, or recommendations/ ways forward).</p>		
<p><b>11. GENDER EQUALITY &amp; HUMAN RIGHTS</b></p>	<p><b>Assessment Level:</b></p>	<p><b>Fair</b></p>

<p>11.1 The evaluation terms of reference confirm that Gender Equality and Human Rights principles and values should be integrated into the developmental evaluation.</p>					
<p>11.2 The evaluation design includes evaluation questions that specifically issues related to Gender Equality &amp; Human Rights.</p>					
<p>11.3 The evaluation includes gender-responsive and human rights responsive evaluation methodology, data collection and data analysis techniques.</p>					
<p>11.4 The evaluation findings, conclusions and recommendations reflect a Gender Equality and Human Rights perspectives and issues.</p>					
<p><b>Overall Evaluation Quality Assessment</b></p>		<p><b>Assessment Levels (*)</b></p>			
		<p><b>Very good</b></p>	<p><b>Good</b></p>	<p><b>Fair</b></p>	<p><b>Unsatisfactory</b></p>
<p><b>Quality assessment criteria (scoring points*)</b></p>					
<p>1. Utilization-Focused</p>					
<p>2. Developmental Purpose</p>					
<p>3. Evalaution Rigor</p>					
<p>4. Innovation Niche</p>					
<p>5. Complexity perspectives</p>					
<p>6. Systems thinking</p>					
<p>7. Co-creation</p>					
<p>8. Timely Feedback</p>					

11. Gender Equality & Human Rights				
<b>Overall assessment level of evaluation reports</b>	<b>Very good</b> very confident to use	<b>Good</b> confident to use	<b>Fair</b> use with caution	<b>Unsatisfactory</b> not confident to use
<p>If the overall assessment is 'Fair', please explain</p>				
<p><b>How It Can Be Used?</b></p>				
<p></p>				
<p><b>What aspects to be cautious about?</b></p>				
<p></p>				
<p><b>Where relevant, please explain the overall assessment Very Good, Good or Unsatisfactory</b></p>				
<p></p>				
<p><b>Consideration of significant constraints</b></p>				
<p>The quality of this evaluation report has been hampered by exceptionally difficult circumstances:</p>				
<p style="text-align: right;"> <input type="checkbox"/> Yes                      <input type="checkbox"/> No         </p>				
<p>If yes, please explain:</p>				

### **Annex 3: Developmental evaluation quality assessment grid for key informant surveys and interviews**

The second methodology employed in the UNFPA quality assessment process is a combination of a survey and/or key informant interviews with selected primary users of the developmental evaluation.

#### 1. Introduction

The intent of the survey and/or interviews is to get an understanding of how 5-6 (or more) primary users of the developmental evaluation and the evaluators experienced the **process** of developmental evaluation. The number of users can be increased or decreased based on the size and complexity of the initiative under evaluation and the importance of the quality assessment. (See the section on scaling options, section 3.5)

#### 2. The approach

Primary users will be given a description of each principle or standard, as well as a list of practices that illustrate how that principle is typically reflected in an evaluation. They will then be asked to answer two questions:

- How would you rate the extent to which the principle or standard was applied in your evaluation (very good, good, fair, poor)?
- Why did you rate this way?

After the surveys and/or interviews have been completed, the quality assessor will summarize the quantitative ratings in a table, and code the qualitative responses for major and minor themes. The analysis will be synthesized in a short (estimated 3-5 page) report, which will form an appendix to the final Quality Assessment Report.



**Organizational Unit:**  **Year of Evaluation:**

**Evaluation:**

**EQA Process:**  **Date of assessment:**

**Interviewees:**

**Overall comments:**

**Assessment Level**

Very Good	Strong, Above Average	Good	Satisfactory, respectable	Fair	Some weaknesses, still respectable	Unsatisfactory	Weak, does not meet minimum quality standards	Don't Know/Insufficient Data
-----------	-----------------------	------	---------------------------	------	------------------------------------	----------------	---	------------------------------

**PRINCIPLES**

	Interviewee #1	Interviewee #2	Interviewee #3	Interviewee #4	Interviewee #5	
<b>1. UTILIZATION FOCUSED</b>						Overall Assessment Level: <span style="background-color: green; color: white;">Very Good</span>
How would you rate the extent did the evaluation focus on questions that were useful to your work?						
How would you rate the extent to which the evaluation process was user friendly (e.g. options for being involved, ease of interaction)?						
How would you rate the extent to which evaluation documents and reports were easy to read and understand?						
How would you rate the extent to which the evaluation generated findings and discussions were useful in understanding how RBM is unfolding in the UNFPA and ways that RBM can be adapted or improved for greater effectiveness?						
<b>2. DEVELOPMENTAL PURPOSE</b>						Assessment Level: <span style="background-color: #cccccc;">Weak, does not meet minimum quality standards</span>
How would your rate the extent to which the evaluation/evaluators provide data/evidence informed implications for the ongoing development and adaptation of RBM in the UNFPA?						
<b>3. EVALUATION RIGOR</b>						Assessment Level: <span style="background-color: #cccccc;">Weak, does not meet minimum quality standards</span>
How would you rate the extent to which the evaluation/evaluators rigorously collected, and interpreted data and generated findings in way that drew on multiple data sources, triangulated data, and made evaluative statements with a strong link to the Evidence/data?						
<b>4. INNOVATION NICHE</b>						Assessment Level: <span style="background-color: #cccccc;">Weak, does not meet minimum quality standards</span>
How would you rate the extent other which the evaluation identified the significance of the RBM approach to the UNFPA, as well as the implementation challenges and resistance that might have emerged with its adoption?						
<b>5. COMPLEXITY PERSPECTIVES</b>						Assessment Level: <span style="background-color: #cccccc;">Weak, does not meet minimum quality standards</span>

How would you rate the extent to which the evaluation/evaluators surfaced and made sense of the complex array of factors related to the adoption of RBM in the UNFPA, as well as the tensions, unpredictable and unintended consequences that might have emerge?							
<b>6. SYSTEMS THINKING</b>						Assessment Level:	
How would you rate the extent to which the evaluation/evaluators took into account the various formal and informal systems within and beyond the UNFPA (e.g. specific UNFPA units, the larger UN, other development agencies, different countries), that might influence how RBM is perceived, adopted and experienced?							
<b>7. CO-CREATION</b>						Assessment Level:	
How would you rate the extent to which the evaluation/evaluation team co-created and continued to adapt the evaluation questions, design and implementation of the evaluation with the primary users of the evaluation?							
<b>8. TIMELY FEEDBACK</b>						Assessment Level:	
How would you rate the extent to which the evaluation/evaluators facilitated the evaluation discussions and feedback in real time?							
<b>9. INDEPENDENCE &amp; IMPARTIALITY</b>						Assessment Level:	
How would you rate the extent to which the evaluation/evaluators operated with impartiality and independence throughout the developmental evaluation?							
<b>10. CREDIBILITY</b>						Assessment Level:	
How would you rate the expertise, competence and credibility of the evaluation/evaluators?							
<b>OTHER ISSUES</b>							
Is there anything else you would like us to know about your experience with the developmental evaluation and/or its overall quality?							
<b>Quality Assessment Summary</b>							
1. Utilization-Focused							
2. Developmental Purpose							
3. Evaluation Rigor							
4. Innovation Niche							
5. Complexity Perspectives							
6. Systems Thinking							
7. Co-creation							
8. Timely Feedback							
9. Independent & Impartial Approach							
10. Credibility							
<b>Overall assessment level of the evaluation process by primary users</b>							
<b>Additional Comments:</b>							

## **Annex 4: Developmental evaluation quality assessment grid for interviews of evaluators and UNFPA evaluation managers**

### 1. Introduction

The third data source and method employed in the UNFPA quality assessment framework is to carry a follow-up interview with the developmental evaluator(s) and the UNFPA evaluation managers who commissioned and managed the evaluation. The focus of the interview is on the **management** of the evaluation.

### 2. The approach

The focus of the interview is to explore how well both parties felt that they managed the purpose, relationship and administrative tensions in the evaluation. The questions include:

- How would you rate the extent to which the evaluation stakeholders managed the three tensions involved in developmental evaluation?
- Why did you provide that rating? On what data do you base your rating?
- What are some measures or actions that could have been employed to better manage tensions?
- What other tensions emerged? How do you rate how well they were managed? Why?

The emphasis on interviews, rather than a survey, enables the quality assessor to gain deeper insight into the developmental evaluation, including the types of trade-offs that it made and warrant consideration in the final assessment of quality.

For example, the final developmental evaluation report might not have been as comprehensive or as in-depth as desired because initiative stakeholders prioritized real-time evaluation feedback on a series of small-scale experiments early in the initiative. This reduced the resources available for producing the final evaluation report. It is important to understand such trade-offs.

After the interviews are completed, the quality assessor will summarize the quantitative ratings in a table, and code the qualitative responses for major and minor themes. The analysis will be synthesized in a short (5-7 pages) report, which will form an appendix to the final Quality Assessment Report.



Organizational Unit:

Year of Evaluation

Evaluation

EQA Process

Date of assessment:

Interviewees

Overall comments:

**Assessment Levels**

Managed Tension



strong, above average, best practice



satisfactory, respectable



with some weaknesses, still acceptable



major weaknesses, does not meet basic standards

1. TENSION BETWEEN DEVELOPMENTAL & ACCOUNTABILITY PURPOSES	Interviewee 1	Interviewee 2	Interviewee 3	Interviewee 4	Assessment Level:
1.1 To what extent did the evaluation focus on providing internal UNFPA staff with evaluative thinking and feedback to inform the ongoing learning, development and adaptation of RBM while operating in an institutional context where traditional evaluation are often carried out to provide summative judgements and/or hold innovators accountable to external bodies.					
1.2 What affect did this tension have on the design, implementation and usefulness of the evaluation?					
1.3 Do you have suggestions on what could have been done to improve the management of this tension?					
<b>2. TENSION BETWEEN CO-CREATION, IMPARTIALITY AND INDEPENDANCE</b>					Assessment Level:

2.1 To what extent did the evaluators manage to meaningfully co-create evaluation designs, and engage evaluation users in other utilization-focused practices, while ensuring their own ongoing independence and impartiality in the process?					
2.2 What affect did this tension have on the design, implementation and usefulness of the evaluation?					
2.3 Do you have suggestions on what could have been done to improve the management of this tension?					
<b>3. TENSION BETWEEN CONTRACTING FOR TRADITIONAL VERSUS DEVELOPMENTAL EVALUATION CONTRACTING</b>					Assessment Level:
3.1 To what extent did the UNFPA Evaluation Office and Developmental Evaluators develop a preliminary evaluation design that was complete enough to finalize funding and an evaluation contract, yet flexible enough to adapt in real time to reflect stakeholders' ever-evolving evaluation needs?					
3.2 What effect did this tension have on the design, implementation and usefulness of the evaluation?					
3.3 Do you have suggestions on what could have been done to improve the management of this tension?					
<b>4. OTHER TENSIONS</b>					Assessment Level:
4.1 Are there other tensions that emerged during the evaluation that required management?					
4.2 How well were these tensions managed?					
4.3 What affect did this tension have on the design, implementation and usefulness of the evaluation?					
4.4 Do you have suggestions on what could have been done to improvement the management of these tensions?					
<b>Quality Assessment Summary</b>					
1. Tension Between Developmental & Accountability Purposes					
2. Tension Between Co-creation, Impartiality & Independence					
3. Tension Between Traditional Versus Developmental Evaluation Contracting					
<b>Overall assessment level of the tensions management</b>					

## Annex 5: Summary of ratings

The quality assessment report will include the following elements:

### Introduction

A brief description of the purpose of the report, its audience, and content.

### Background

A description of the developmental evaluation being assessed, links to key documents, and approach employed in the quality assessment process (e.g., data sources, methods, timeline), including limitations that should be considered when reviewing the results.

### Summary of findings

A summary of the quality ratings from all three data sources, as well as the triangulated highlights from qualitative feedback for each rating.

**Table: Summary of ratings<sup>26</sup>**

Principle	Data sources/methods			Average ratings (optional)	Assessor summary of key themes
	Method 1: Evaluation reports	Method 2: Key informant interviews	Method 3: Interview of evaluators and UNFPA evaluation managers		
Utilization-focused					
Developmental purpose					
Evaluation rigour					
Innovation niche					
Complexity aware					
System thinking					
Co-creation					
Timely feedback					
Credibility					

<sup>26</sup> Unsatisfactory, Fair, Good, Very Good

Independence and impartiality					
-------------------------------	--	--	--	--	--

**Discussion of findings (Optional)**

Elaborate on the findings, including areas of agreement and divergence, and offer conclusions about the overall quality of the developmental evaluation.

**Reflections on quality assessment process**

The quality assessor’s reflections on the strengths and limitations of the quality assessment framework and process, and suggestions for improving them in future iterations.

**Appendices**

A summary of the quality ratings from all three data sources, as well as the triangulated highlights from qualitative feedback for each rating.

**Annex 6: Evaluation quality assessment report of developmental evaluation of results-based management at UNFPA**

Please see ahead

# Developmental Evaluation of Results-Based Management at UNFPA

A Quality Assessment

December 20, 2019



Mark Cabaj

Here to There  
CONSULTING INC

Contents

- 1. Purpose ..... 2
- 2. Period ..... 2
- 3. Approach ..... 2
- 4. Quality Assessment Team ..... 2
- 5. Summary of the Findings ..... 3
- 6. Detailed Findings..... 4
  - 6.1 Adherence to Quality Principles ..... 4
    - 6.1.1 Developmental Purpose..... 4
    - 6.1.2 Utility-Focus ..... 5
    - 6.1.3 Co-Creation Stance ..... 6
    - 6.1.4 Innovation-Focus..... 6
    - 6.1.5 Systemic Thinking..... 7
    - 6.1.6 Complexity-Aware..... 7
    - 6.1.7 Rigorous Process ..... 8
    - 6.1.8 Timely Feedback ..... 9
    - 6.1.9 Credibility ..... 9
    - 6.1.10 Impartiality & Independence ..... 10
    - 6.1.11 Gender Equality and Human Rights ..... 10

# The Developmental Evaluation of the UNFPA's Results-Based Management: A Quality Assessment

## 1. Purpose

This document summarizes the results of an evaluation quality assessment (EQA) of the developmental evaluation (DE) of the UNFPA's experience by Jordi Del Bas and Josep M. Coll.

The aim of the evaluation was to examine the root causes of persistent challenges in the use of results-based management (RBM) in the UNFPA and to surface evidence for solutions, possible scenarios, and courses of action in the area of RBM.

The audience for this EQA report are all stakeholders interested in using the DE of results-based management at UNFPA.

## 2. Period

The EQA is a retrospective assessment. It covers the period from June 2018 when the Terms of Reference for the report were released, to December 2019, with the review of evaluation documents and reports, and the completion of interviews with primary users, developmental evaluators, and UNFPA staff.

## 3. Approach

The EQA was carried out using the EQA framework for UNFPA Developmental Evaluations. The framework is organized around the 11 principles of DE Quality established by the UNFPA Evaluation Office.

The assessment is based on three data sources and techniques:

External Reporting	<ul style="list-style-type: none"><li>• A review of the major evaluation documents in the assessment, including Terms of Reference, the Inception Report, a variety of PowerPoint presentations, Feedback Notes, and the final DE report.</li></ul>
	<ul style="list-style-type: none"><li>• Interviews with five primary users of the evaluation.</li></ul>
Internal Learning	<ul style="list-style-type: none"><li>• Interviews with the two Developmental Evaluators and two members of the UNFPA Evaluation Office responsible for overseeing the evaluation, to explore issues related to the process of designing and managing of developmental evaluations in the UNFPA.</li></ul>

This is a summary EQA report. A more detailed analysis is contained in the DE-EQA Excel spreadsheets annexed to this report. Quality Assessment Team

Mark Cabaj, President of Here to There Consulting Inc., developed the EQA framework and carried out the EQA review. Dr. Michael Quinn Patton provided advice in the development of the EQA framework and throughout the EQA process.

#### 4. Summary of the Findings

The DE of results-based management at UNFPA was assessed as **Very Good**, as largely met and exceeded the principles of quality laid out in the agency’s EQA framework.

Quality Principles		Ratings			
		Very Good strong, well above minimum standards	Good satisfactory, respectable related to minimum standards	Fair some weaknesses, but still acceptable	Unsatisfactory weak, does not meet minimum standards
1	Developmental Purpose				
2	Utility Focus				
3	Co-Creation Stance				
4	Innovation Focus				
5	Systemic Focus				
6	Complexity Aware				
7	Rigorous Process				
8	Timely Feedback				
9	Credibility				
10	Impartiality & Independence				
11	Gender Equality & Human Rights				

The overall very good quality of the developmental evaluation of results-based management at UNFPA is a significant achievement for three reasons:

- This was the first developmental evaluation undertaken in the UNPFA.
- The work of RBM is complex and the breadth of engagement with RBM stakeholders in the evaluation was significant.
- The evaluation team and UNFPA Evaluation staff were aware of the principles that defined quality in DE, but began the evaluation process without any knowledge of the EQA framework that would be used to assess the quality of the work<sup>1</sup>.

In addition to confirming that the primary users, commissioners, and UNFPA should have confidence in the evaluation process and findings, the EQA process generated useful insights into

<sup>1</sup> At the time this evaluation was launched, no EQA for DE was existing. Therefore, UNFPA Evaluation Office decided to hire Mark Cabaj and Michael Quinn Patton (in a role as technical advisor) to develop the first ever EQA framework to assess quality of DE. The EQA framework for DE was finalized in December 2019, when the draft report of the DE of results-based management at UNFPA was already produced.

how the agency’s future DEs might be strengthened and how the quality of these DEs might be assessed.

## 5. Detailed Findings

### 6.1 Adherence to Quality Principles

This section describes the extent to which the evaluation adhered to each of the 11 quality principles for UNPFA-managed DEs. It includes a rating (i.e., poor, satisfactory, good or very good) as well as a rationale for that rating, based on the review of DE products and/or reviews with primary users, developmental evaluators, and members of the UNFPA Evaluation Office.

#### 6.1.1 Developmental Purpose

<p>Very Good: strong, well above minimum standards.</p>	<p>Good: satisfactory, respectable related to minimum standards.</p>	<p>Fair: some weaknesses, but still acceptable.</p>	<p>Unsatisfactory: weak, does not meet minimum standards.</p>
---	--	---	---

The principle of developmental focus was strongly adhered to throughout the evaluation. The strengths were as follows:

- The developmental purpose of the evaluation was clearly stated in all the major evaluation products, beginning with the Terms of Reference, and confirmed in interviews and workshops with primary users, developmental evaluators, and UNFPA Evaluation Staff.
- The evaluation focused on the tensions that prior evaluations and the scoping exercise identified as being the major issues that influence the continued development, adaptation and usefulness of RBM in the agency.
- The path forward for further development and adaptation of RBM described in the DE report is presented as points of leverage, entry points, and guiding principles, the primary way that DE seeks to inform the next iteration of innovations.

Two challenges pertain in terms of adherence to the principle of developmental purpose, both of which lay outside the control of the evaluation team, and illustrate the challenge of carrying out a developmental evaluation in an agency accustomed to more traditional formative and summative evaluations:

- Several interviewees reported that UNFPA staff are accustomed to working with traditional evaluations that typically offer primary users concrete conclusions and recommendations for future action. They noted that some of their colleagues found the emphasis on laying out options, leverage points, and principles for moving forward to be “overwhelming,” “too abstract,” or “not detailed enough.”
- The purpose of DE is mainly to generate evidence to provide real-time feedback to senior management with the aim of informing ongoing learning and improvements. The fact that,

as any UNFPA centralized evaluations, a report for the UNFPA board was to be developed and presented, created a spirit of “accounting to governing bodies.” Therefore, the evaluators and UNFPA Evaluation Office invested additional time and effort to create a report more normally suited to traditional summative evaluations, rather than one that aims to be more real time, deep reflective, learning oriented and detailed.

### 6.1.2 Utilitization-Focused

<p>Very Good: strong, well above minimum standards.</p>	<p>Good: satisfactory, respectable related to minimum standards.</p>	<p>Fair: some weaknesses, but still acceptable.</p>	<p>Unsatisfactory: weak, does not meet minimum standards.</p>
---	--	---	---

The principle of utility was strongly adhered to throughout the evaluation. The strengths included:

- The primary uses (i.e., “To inform the implementation of the UNFPA RBM policy and development of the RBM strategy”) and users (Policy and Strategy Division) of the evaluation were clearly identified in the Evaluation Scope of Work, Inception Report, and final DE report. The evaluation design evolved to accommodate an expanding and diverse group of users in the countries and region and senior UNFPA staff.
- The evaluation documents were easy to read and primary users report that evaluators’ verbal communications were very good.
- The UNFPA evaluation staff and primary users reported that the evaluators were attentive to their questions and open for spontaneous discussions.
- The utility of the evaluation was substantially enhanced by the evaluators’ willingness and ability to continue to expand the number and variety of primary users to include technical staff and different organizational units. Their interest in participating emerged as the evaluation began to generate timely and relevant feedback.
- The final DE report describes 10 ways in which the evaluator perceived that the evaluation was “useful” to UNFPA staff. In addition, several interviewees identified concrete ways in which they have adapted their thinking or practices in response to the DE.

The interviews with primary users surfaced one limitation related to the utility of the evaluation:

- The decision not to target findings, leverage points, and options to specific stakeholder groups within the UNFPA (e.g., country, regional, or specific units within HQ). This meant that the feedback was, in the words of two interviewees, “a bit general” and “too broad.”

It is important to note that this limitation is due, in part, to two matters. First, the decision to move beyond the Policy and Strategy Division as the primary user, and engage staff in 11 different organization units and a variety of countries and regions: this made it difficult to customize findings to so many different sub-groups. Second, the perception that Evaluation Reference Group and primary users did not ask for such a breakdown in the scoping exercise, or final report.

### 6.1.3 Co-Creation Stance

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The adherence to the principle of co-creation in the evaluation was very good. The strengths were as follows:

- A strong commitment and range of activities to engage the Evaluation Reference Group and primary users in the scoping process and development of the Inception Report.
- The consistent practice of testing, elaborating and adapting early findings with the Reference Group and different units within the UNFPA.
- The willingness to integrate the DE activities into various existing events and meetings at UNFPA Headquarters and regional offices and events.

The EQA process did not uncover any significant limitations related to the evaluation’s adherence to the co-creation principle.

### 6.1.4 Innovation-Focus

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The adherence to the principle of innovation in the evaluation was very good. The strengths included:

- The Inception Report summarizes the history of the RBM “innovation” in UNFPA, the findings of past evaluations and its adoption and adaptation in the agency, and a summary of how it has been continuously developed and adapted over time.
- The central focus on the cultural and structural issues and tensions underlying the adoption, implementation, and resistance to RBM, and how they influence results.
- The emphasis on surfacing points of leverage, entry points, and guiding principles to inform the next iteration of RBM development and adaptation.

The EQA process did not uncover any significant limitations related to the evaluation’s adherence to the innovation principle.

### 6.1.5 Systemic Thinking

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The adherence to the principle of systemic thinking in the evaluation was very good. The strengths included:

- The fulsome description of the root causes – such as structural features and culture – that influence the manner in which RBM is being employed in UNFPA.
- The acknowledgment and analysis of exogenous factors beyond the UNFPA that influence the agency’s adoption and use of RBM.
- The strong and sustained emphasis on engaging a wide range of UNFPA staff from across the organization to ensure a diversity of perspectives in the evaluation.
- The employment of various “systems thinking” metaphors and techniques, such as iceberg framework, archetypes, and stakeholder mapping.
- The focus on organizing implications for further adaptation of the RBM around “points of leverage.”

The EQA process did not uncover any significant limitations related to the evaluation’s adherence to the systemic thinking principle.

### 6.1.6 Complexity-Aware

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The adherence to the principle of complexity awareness in the evaluation was very good. The strengths were as follows:

- An evaluation design organized around understanding and exploring the various tensions involved in employing RBM in the UNFPA, a central feature of a complexity-aware evaluation and the foundation for the inquiry framework.
- The identification of various unintended consequences of employing RBM in the agency.
- The acknowledgment of the difficulty in establishing cause-and-effect relationships between the various factors shaping RBM policy and implementation. Nevertheless, an attempt was made at least to map the dynamic and interdependent relationship between them.

- The evaluation’s emphasis on exploring all the behaviours and results emerging from the RBM experience, including unintended consequences.

The EQA process did not uncover any limitations related to the evaluation’s adherence to this principle.

#### 6.1.7 Rigorous Process

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The evaluation’s adherence to the principle of rigorous process was very good. The strengths were as follows:

- The evaluation questions, methods and data sources are clearly laid out in the Inception Report and summarized in the final DE report.
- The evaluation design includes multiple methods, data sources, and analytical techniques, which allow for triangulation.
- The evaluators’ consistent engagement with the Evaluation Reference Group and primary users to discuss, test, and revise their understanding of evaluative discussions, feedback, and findings.
- The Inception Report identifies the risks related to the evaluation, while the Methodological Note outlines how the risks were managed through the process and the limitations that emerged related to data collection, how they were managed, and the effect on evaluation findings.
- The evaluation findings and descriptions of implications for further development are logically based on data emerging from the various inquiries.

The one limitation in the evaluation rigor is the lack of a consistent breakdown of evaluation data, analysis and, implications for development by different categories or sub-groups (e.g. regions, demographics, countries, agency unit. While this occurred in (1) some of the evaluation inquiries (e.g., the survey of UNFPA staff broke findings down by gender and organizational role), (2) somewhat in the implications for development (e.g., measuring and capturing value, RBM information systems), and (3) in many of the face to face interactions, it is not a consistent feature in the evaluation.

It is important to that in the Methodological Note, the evaluators report that these breakdowns were not requested by the Evaluation Reference Group or users, nor is it clear from the EQA review that the uneven breakdown of evaluation findings by sub-group limits the overall quality of the evaluation. For this reasons, the rating of evaluation rigor remains very good.

### 6.1.8 Timely Feedback

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The evaluation’s adherence to the principle of timeliness was good. The strengths included:

- The willingness and ability of the evaluators and UNFPA staff to participate in variety of regional and HQ meetings and events in “real time.”
- The use of diverse formats to provide constant evaluation feedback, including PowerPoint presentations, Feedback Notes, and presentations and meetings.
- The use of different types of feedback, including consolidated feedback on identified themes (e.g. Information Systems), requested feedback, where the evaluation team responded to agency staff request for input on key issues, and emergent feedback, where the evaluation provided ‘on-the-go’ pieces of information that the evaluation team felt were relevant to the UNFPA’s work.
- The number and variety of spontaneous “sense-making” and feedback sessions during interviews and meetings.

One factor that limited the timeliness of the evaluation process and findings was beyond the control of the evaluators and instead structured into the evaluation itself: the challenge of informing the long-term process of developing and adapting the RBM in the UNFPA within a relatively short cycle of evaluation. This means that the work of facilitating the use of findings into the many different units of the UNFPA, as well as the organization overall, fall outside of the timeline for this cycle of developmental evaluation.

### 6.1.9 Credibility

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The credibility of the evaluation team and process was very good. The strengths included:

- The professionalism and enthusiasm of the evaluation team.
- The evaluators’ broader experience and knowledge of organizational development, change management, and RBM.
- The overall manner in which the team designed and managed the evaluation, including their commitment to co-creation, their willingness to evolve and adapt the evaluation, their

interpersonal skills and commitment to listening, and the rigor with which they carried out the evaluation.

The EQA surfaced only one minor limitation in the review. Two interviewees reported that they did not feel the evaluation team members were leading experts in the RBM field, on top of the most recent developments in areas such as benchmarking. However, both interviewees reported that they felt that this was a minor concern and rated the team’s credibility as very good .

#### 6.1.10 Impartiality & Independence

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The impartiality and independence of the evaluators and evaluation was very good. The strengths identified in interviews with primary users included:

- The strong commitment to gathering a wide diversity of perspectives to answer the evaluation’s core questions.
- The emphasis on linking evaluative statements to the data collected in the evaluation.
- The evaluators’ willingness to explore and debate different interpretations of data, as well as the implications for further development (e.g., points of leverage, entry points, and guiding principles), treating this diversity of perspectives as data, which was summarized and presented in a balanced manner.
- The evaluators’ own reports that they did not feel pressure from the UNFPA Evaluation Office, Reference Group, or primary users to adjust their final oral or written communications in any way, a sentiment confirmed by the five primary users interviewed in the EQA process.

The EQA process did not uncover any limitations related to the evaluation’s adherence to this principle.

#### 6.1.11 Gender Equality and Human Rights

Very Good: strong, well above minimum standards.	Good: satisfactory, respectable related to minimum standards.	Fair: some weaknesses, but still acceptable.	Unsatisfactory: weak, does not meet minimum standards.
--	--	--	--

The evaluation’s adherence to the principle of integrating gender equality and human rights is fair. The strengths included:

- The general direction to follow the UN Evaluation Guidelines in the Inception Report, which include a principle to integrate gender quality and human rights issues.

- The evaluation used a utilization-focused approach, which promotes intended use by intended users, implying a strong focus on participation of users throughout the evaluation process. Users from all levels of the organization participated as well as most UNFPA business units as well as all functions. All stakeholders were given the opportunity for meaningful participation.
- The evaluation design was flexible to easily adapt to changes in user priorities and contexts.
- A combination of data collection methods (mixed-methods approach) was utilized to gather and analyse information to offer different perspectives and promote participation of different stakeholders. This approach also served to validate findings.
- An extra emphasis on engaging people in interviews and surveys in a way that encourage gender and youth perspectives, including ensuring space for open-ended responses by interviewees/surveyed, and active engagement with Tangerine network, a group of young professionals at the UNFPA
- An initial scan of how the UNFPA's current data systems incorporate issues related to gender equality, and discussion of how to integrate relevant questions and methods into the design at the regional workshop.

There are four limitations in how gender equality and human rights were integrated into the evaluation:

- The evaluation design did not include questions related to the role that gender has played in the design and implementation of RBM.
- The evaluation did not include questions relating to the gender equality and human rights aspects of RBM.
- The evaluation did not explicitly employ gender-informed methods for gathering and analyzing data.
- The various reports do not provide data, findings, or implications using the lenses of gender equality.

# **Annex**



**Organizational Unit:** UNFPA Evaluation Office

**Year of Reports:** 2018-2019

**Evaluation Deliverables Reviewed**

The review included the developmental evaluation process and deliverables. There are 10 documents reviewed: (1) Evaluation of UNFPA Approaches to Results-Based Management: Terms of Reference. January 2018; (2) Developmental Evaluation of Results-based Management of UNFPA: Inception Report. November 2019. (3) Creative Tension 1: RBM Conceptual Framework. Developmental Evaluation of Results-based Management at UNFPA. Jordi del Bas & Josep M. Coll. Feedback Note - June 2019; (4) Creative Tension 1: RBM Policy Update & Development. Developmental Evaluation of Results-based Management at UNFPA. Jordi Del Bas & Josep M. Coll. Feedback Note - June 2019 ; (5) Creative Tension 2: Collective Versus Individual Accountability. Developmental Evaluation of Results-Based Management at UNFPA. Jordi del Bas & Josep M. Coll. Feedback Note - July 2019; (6) Organizational Culture & Use of Results. Developmental Evaluation of Results-Based Management at UNFPA. Josep M. Coll & Jordi del Bas. Feedback Note - June 2019; (7) Creative Tension 4: RBM Capacity to Manage for Results: Developmental Evaluation for Results-Based Management at UNPFA. Developmental Evaluation of Results-Based Management at UNFPA. Josep M. Coll & Jordi del Bas. Feedback Note - July 2019; (8) Creative Tension 5: RBM Information Systems: Developmental Evaluation of Results-Based Management at UNFPA. Josep M. Coll & Jordi del Bas. Feedback Note - July 2019; (9) Developmental Evaluation of Results-Based Management at UNFPA: Final Account Report. Josep M. Coll & Jordi del Bas. November 2019; Developmental Evaluation of Results-Based Management at UNFPA: Methodological Note. Josep M. Coll & Jordi del Bas. December 2019. In addition to these documents and reports, the quality assessment also reviewed a half-dozen PowerPoint presentations delivered during the evaluation.

**Overall quality of reports:**

**Very Good**

**Date of assessment:** 23 December 2019

**Overall comments:**

The areas of strength in the DE include: utilization-focused, developmental purpose, innovation-focused, co-creation stance, complexity-awareness. systemic thinking and evaluation rigor. These aspects were rated very good. The limitation in the evaluation documents include uneven adherence to integration of Gender Equality and Human Rights into the evaluation. The evaluation reports' adherence to this rating is fair. The overall assessment level of evaluation reports is 'very good': very confident to use.

Assessment Levels	<div style="background-color: #00aaff; color: white; padding: 5px; text-align: center;"><b>Very Good</b></div>	strong, above average, best practice	<div style="background-color: #90d060; padding: 5px; text-align: center;"><b>Good</b></div>	satisfactory, respectable of adequate quality	<div style="background-color: #ffcc00; padding: 5px; text-align: center;"><b>Fair</b></div>	with some weaknesses, still acceptable	<div style="background-color: #ff0000; color: white; padding: 5px; text-align: center;"><b>Unsatisfactory</b></div>	weak, does not meet minimal quality standards
<b>Quality Assessment Criteria</b>	<i>Insert <u>assessment level</u> followed by main <u>comments</u></i>							
<b>I. UTILIZATION FOCUSED</b>	<b>Assessment Level:</b>						<b>Very Good</b>	
<b>I.1 To ensure the evaluation reports are focused on primary intended users.</b>	<b>Rating</b>						<b>Very Good</b>	
I.1.1 The evaluation reports clearly describe the 'primary users' of the evaluation and their 'primary uses' of the evaluation process and findings.	<b>Very Good</b>	(1) The Terms of Reference for the evaluation identifies the primary users of the assessment as UNFPA management for the purposes of "inform[in] the implementation of the UNFPA RBM policy and development of the RBM strategy."; (2) The Inception report confirms that the focus is on management in UNFPA, and includes a comprehensive list of managers as part of the Technical Team; (3) The Feedback Notes do not include the primary users of the report, but this does not appear a serious gap given the clarity offered in the TOR and Inception report; (4) The Final Account Report refers to primary users, and their different objectives, but does not describe them clearly; (5) the Methodological Notes provides a summary of the emerging nature of the evaluation, where technical staff and senior and Executive Office persons also became engaged in the process and findings, thereby becoming 'users', and includes a description of the II business units in UNFPA that were involved and the use by every unit of the evaluation findings.						
I.1.2. The evaluation reports describe the major evaluation questions that primary intended users would like to answer in the evaluation.	<b>Very Good</b>	The inquiry framework includes a clear set of questions to guide the overall developmental evaluation. These questions provide the organizing framework for all subsequent evaluation activities and communications and remained constant throughout the evaluation.						
<b>I.2 To ensure evaluation reports are comprehensive and user-friendly</b>	<b>Rating</b>						<b>Very Good</b>	

<p>1.2.1. The evaluation reports are easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors.</p>	<p>Very Good</p>	<p>The individual reports are well written, concise and accessible to the general reader. Some readers may struggle with some of the ideas and concepts in the documents, but this may be due more to different learning styles of evaluation users (e.g. some read, others process information through interaction), rather than the quality of the communications and the documents. The use of diagrams and various other visual techniques (e.g. tables, diagrams) adds to the readability of the documents.</p>
<p>1.2.2. The evaluation reports are of a reasonable length (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations).</p>	<p>Very Good</p>	<p>The reports are of reasonable length, ranging from 17 pages for feedback notes on specific items and 51 pages for the entire DE Report.</p>
<p>1.2.3. The evaluation reports are structured in a logical way: there is clear distinction made between analysis/findings, discussions, and implications for further development.</p>	<p>Very Good</p>	<p>The report structure is clear and logically presented, with the Feedback Notes on tensions helpfully and clearly organized around priority questions and "takeaways" of key insights and findings.</p>
<p>1.2.4. The reports include critical background information: e.g. the Terms of Reference; a bibliography; a list of interviewees; an inquiry framework; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process.</p>	<p>Very Good</p>	<p>The Inception Report provides a complete description of the methods to be used to answer the questions. The DE Report provides a summary of the approach employed while the Methodological Note provides a more thorough review and reflection on the methods employed.</p>
<p><b>1.3 Executive summary</b></p>	<p style="text-align: center;"><b>Rating</b></p>	
<p>1.3.1. The final DE Report includes an executive summary written as a stand-alone section and presents the main results of the evaluation.</p>	<p>Very Good</p>	<p>The final DE Report includes a stand-alone executive summary that provides a background to the evaluation, its approach, and major findings and implications for further development and adaptation of RBM in the UNFPA.</p>
<p>1.3.2. The executive summary is clearly structured and includes the major points related to the evaluation as they appear in the DE Report (purpose, methodology, highlights from analysis, and implications for further development).</p>	<p>Very Good</p>	<p>The executive summary provides a background to the evaluation, a description of the methodological approach, and the major findings and types of 'evaluation use' that emerged.</p>
<p>1.3.3. The executive summary is reasonably concise (e.g. with a maximum length of 5 pages).</p>	<p>Very Good</p>	<p>The evaluation summary is precisely five pages.</p>
<p><b>1.4 Evidence of Use</b></p>	<p style="text-align: center;"><b>Rating</b></p>	

<p>1.4.1. The evaluation reports provide evidence that the evaluation findings and process are being used by the primary users, including one or more of the following three types: (1) conceptual use - leading to new ways of perceiving the challenges, (2) process use - increasing capacity to learn, curiosity, confidence, and (3) instrumental use - using data to inform decisions, new actions or behaviour change.</p>	<p>Very Good</p>	<p>The Final DE Account describes 10 'reactions, effects and changes' that the evaluation team perceive as having emerged through the assessment, which are primarily conceptual uses (new ways of seeing things) as well as process uses (e.g. greater number of persons engaged in the process). It does not include a summary of instrumental uses for three reasons: (1) a question about whether DEs - rather than the UNFPA leadership - should be tracking the use of DE feedback, (2) the logistical and methodological difficulties of collecting data on use across a large organization like the UNFPA, and (3) the reality that many primary users may adjust their actions, behaviours and decisions after the relatively short period of the evaluation.</p>
<p><b>2. CO-CREATION STANCE</b></p>	<p><b>Assessment Level:</b> <span style="float: right;"><b>Very Good</b></span></p>	
<p>2.1 The evaluation reports describe how the evaluators/evaluation team were positioned as part of the primary users/stakeholders involved in planning, management and implementation activities.</p>	<p>Very Good</p>	<p>The Inception Report outlines how the evaluators engaged evaluation stakeholders - Evaluation Reference Group and Evaluation Working Group - in the scoping phase to understand the purpose/questions of the evaluation, as well as the Inception phase, to develop and validate the evaluation design, including the 'five tensions' framework, and related evaluation methods and activities; these processes were described once again in the Final Account. The report also included descriptions of how the evaluation team participated in ongoing RBM activities, including three face to face RBM workshops in Cairo, Johannesburg and Kiev.</p>
<p>2.2 The evaluation reports describe how they engaged primary intended users in the development of evaluation questions, collecting, analyzing and interpreting data, and surfacing implications for further development and adaptation.</p>	<p>Very Good</p>	<p>The Inception Report describes the participatory process with which the evaluation team identified and developed the conceptual framework for tensions, the priority questions, and the methods for the evaluation. The Methodological Note describes how the feedback processes and products used to describe the emergent responses to these questions were shared, discussed, elaborated and upgraded in real time with a broad range of primary users.</p>
<p><b>3. DEVELOPMENTAL PURPOSE</b></p>	<p><b>Assessment Level:</b> <span style="float: right;"><b>Very Good</b></span></p>	
<p><b>3.1 Describing Base-Line Conditions</b></p>	<p><b>Rating</b> <span style="float: right;"><b>Very Good</b></span></p>	
<p>3.1.1 The development and institutional context of the evaluation clearly described and the constraints explained.</p>	<p>Very Good</p>	<p>All of the evaluation reports - from the TOR, Inception Report, Feedback Notes and DE Report - provide a comprehensive, clear description of base-line conditions of the rationale and history of RBM in UNFPA, including the rationale for developmental evaluation at this time in the institution's efforts to employ RBM, and reference to the findings from past evaluations. The background in the DE Report is exceptionally comprehensive.</p>

<p>3.1.2 The evaluation reports describe the theme/ system/ strategy/ programme/ intervention's rationale, hypothesis, logic and/or theory of change, and assess the adequacy of these.</p>	<p>Very Good</p>	<p>The Inception Report contains a clear history of the adoption and milestones of RBM in the UNFPA, a series of broad statements about its intended outcomes, and a review of its underlying assumptions and challenges throughout the evaluation reports. It does not include a conventional description and assessment of the underlying UNFPA RBM theory of change. However, this should not be considered a fundamental limitation of the evaluation as the evaluation itself is the next in a series of evaluations of a well known RBM implementation in the agency.</p>
<p><b>3.2 Tracking Developments</b></p>	<p style="text-align: center;"><b>Rating</b></p>	
<p>3.2.1 The evaluation reports document the most significant developments to emerge (i.e. forks in the road, shifts in context, results, etc.) and assess the implications for further development and adaptation.</p>	<p>Good</p>	<p>The Final DE report provides a clear account of (1) developmental evaluation interventions (pages 31-35), (2) engagement with stakeholders (pages 35-36), (3) a summary of the effects of the assessment of the DE on the insights and awareness as perceived by the evaluation team (pages 37-38), (4) a five stage history of the development of the RBM practice within the UNFPA (pages 23-29), but does not include (5) an account of more recent developments within the UNFPA approach to RBM, either those that occurred organically and/or those influenced by the DE. This does not seem to be a limitation in the context of the RBM DE because of the comparatively short time period for the evaluation and the case for a second cycle of evaluation, identified in the limitations section of the Methodological Note</p>
<p>3.2.2. The evaluation reports document the significant adaptations to the original theme/ system/ programme/ intervention (e.g. rationale, goals, results expectations, hypothesis, logic and/or theory of change) informed by the evaluation feedback, including offering the evidence and/or rationale upon which the adaptations have been made?</p>	<p>Not Applicable</p>	<p>This practice is not applicable in this iteration of the developmental evaluation of the RBM: the time frame for the evaluation was too short to track the major developments that emerged due to the evaluation findings. Tracking such developments and adaptations will require a second cycle of developmental evaluation.</p>
<p><b>4. EVALUATION RIGOUR</b></p>		
<p><b>4.1 To ensure a rigorous design and methodology</b></p>	<p style="text-align: center;"><b>Rating</b></p>	
<p>4.1.1 The evaluation/inquiry framework - and its link to the primary evaluation questions - are clearly described.</p>	<p>Very Good</p>	<p>The inquiry framework is organized around the creative tensions that emerged out of the inception phase and are clearly described, including the priority questions, in the Inception Report, Feedback Notes and DE Report.</p>

4.1.2 The tools for data collection and analysis are described and their choice justified.	Very Good	The tools for the data collection and analysis are described fully in the Inception Report and in the DE Report. The Methodological Note provides the rationale for each of the techniques in the design.
4.1.3. The methodological limitations are acknowledged and their effect on the evaluation described.	Very Good	The Inception Report and methods section in the DE Report describe the risks to the evaluation, while the Methodological Note explores the extent to which these anticipated surfaced during the evaluation, as well as other additional limitations that emerged during the evaluation, and how they influenced the quality, design and/or implementation of the evaluation.
<b>4.2 To ensure quality of data and robust data collection processes</b>	<b>Rating</b>	
4.2.1 The evaluation design seeks to triangulate data collected where appropriate and possible.	Very Good	The evaluation employs multiple data sources throughout the assessment (e.g. files, interviews, focus groups, surveys, participant observation), drawn from a wide range of UNFPA staff, to answer priority questions. There is evidence of triangulation of data between sources and methods in many of the Feedback Notes.
4.2.2 The evaluation reports identify and make use of reliable qualitative and quantitative data sources.	Very Good	The evaluation employs largely qualitative data - which is highly appropriate given the topic and inquiry framework - as well as quantitative data from on-line surveys and secondary data/evidence from other institutions employing RBM and studies on the topic.
4.2.3 The evaluation reports identify possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues.	Very Good	The Methodological Note describes two instances where there were limitations in the data sources: (1) The first round of primary data collection and analysis occurred in late 2018, when many UNFPA staff were busy with year end management activities, (2) the interaction and feedback from staff in West and Central Africa Region was lower than in other regions due to the schedules of technical and managerial staff. These limitations were offset by extending the time period used to gather data and employing interviews, rather than surveys, with regional staff. The Methodological Note points out that there was no limitations in accessing UNFPA secondary data.
<b>4.3 To ensure sound analysis and credible findings.</b>	<b>Rating</b>	
4.3.1 The findings substantiated by evidence.	Very Good	The findings are clearly organized and described, most often (though not always) with reference to the evaluation data used to support them in the various Feedback Notes.
4.3.2 The analysis and findings presented against the evaluation questions.	Very Good	The evaluation findings are organized around the priority questions: the link is very clear.
4.3.7 The analysis and findings presented against contextual factors.	Very Good	The analysis includes frequent and consistent references to contextual factors that shape the implementation of RBM in UNFPA in the Inception Report, Feedback Notes and DE Report.

<b>4.4 To assess the validity of the implications for further development of the intervention emerging from the findings.</b>	<b>Rating</b>		<b>Very Good</b>
4.4.1 The implications for further development (e.g. new questions, working conclusions, forks in the road, options, leverage points) flow clearly from the findings?	Very Good	This is a difficult criteria to assess given the complexity of the issue and number of variables; yet, in general, the implications flow logically and clearly from the findings. This is particularly true in the Final DE Report where the rationale for each of the proposed leverage points is clearly described.	
4.4.2 The implications for further development (e.g. new questions, working conclusions, forks in the road, options, leverage points) are clearly written and framed in a way that informs primary users' choices about the next iteration of the development of their innovation.	Very Good	The evaluations typically offered options and scenarios for UNFPA and primary users to consider in the next iteration of development. These are well written and offered insights into options the UNFPA might consider.	
4.4.3 The implications for further development (e.g. new questions, working conclusions, forks in the road, options, leverage points) go beyond the findings and provide a deeper understanding of the underlying issues of the programme/initiative/system being evaluated.	Very Good	The findings and options are framed in a way that suggests a deep understanding of the issues related to integrating RBM into UNFPA systems; the reference to the experiences of other organizations and developmental institutions with RBM adds to the credibility.	
4.4.4 The implications for further development proposed appear balanced, impartial and convey the evaluators' unbiased perspective.	Very Good	The proposed implications for development, particularly the six leverage areas offered in the Final DE Report appear to be primarily informed by the analysis of feedback, by UNFPA stakeholders, or previous evaluation reports and documentation. The implications appear very balanced and impartial.	
<b>4.5 To assess the flexibility and adaptability of the evaluation design.</b>	<b>Rating</b>		<b>Very Good</b>
4.5.1 The evaluation reports describe the processes evaluators and intervention stakeholders employ to reflect on and adapt the evaluation design to respond to new evaluation questions and other key factors that may require such adjustments (e.g. new users, evolving questions, opportunities for new methods).	Good	The Inception Report and Methodological Note do not describe the management processes that were put in place to help identify, assess and decide if and when adjustments in the evaluation design were required. The changes in the evaluation design described below, which refer to the expansion in the number and variety of stakeholders involved in the process, emerged organically in the course of the day to day implementation of the evaluation. The lack of such established processes do not appear to have limited the evolution or quality of the evaluation, though they did contribute to some of the tensions related to contract management. Given that this practice was not clearly articulated in the Terms of Reference, nor in a DE Quality Framework before the evaluation began, this should not be considered a major limitation. However, it should be integrated into future developmental evaluations.	

<p>4.5.2 The evaluation reports describe changes to the design that emerge during the evaluation and the rationale for each major change.</p>	<p>Very Good</p>	<p>While the major questions and general inquiry framework remained largely the same throughout the evaluation, several elements of the design evolved: (1) the broader array of data sources to answer those questions, including technical staff (2) the wider-than-originally-anticipated opportunities for engagement for UNFPA staff, (3) the number, variety and content of feedback mechanisms, and (4) the list of primary users (i.e. expanding to 11 organizational units). These evolutions are described in the Methodological Note.</p>
<p><b>5. INNOVATION-FOCUSED</b></p>	<p><b>Assessment Level:</b></p>	
<p>5.1 The evaluation describes the elements of the theme/ system/ programme/ intervention/ strategy/ model or practice being developed, as well as stakeholders' perceptions of the significance of the innovations/ change/ transformation.</p>	<p>Very Good</p>	<p>The rationale and logic of transforming the implementation of RBM in UNFPA was clearly laid out in all the major evaluation documents reviewed as part of the background to the evaluation, with the most detail being offered on the specific RBM practices employed in UNFPA described in the Terms of Reference. The 'tension' framework that emerged out of the Inception process and described in the Report sharpened the focus of the evaluation to "the most pressing aspects affecting the development of RBM to a next stage [in UNFPA operations]" (page 28), which are key to the RBM's next iteration of adaptation in UNFPA.</p>
<p>5.2 The evaluation identifies new questions and insights that emerged through the process, including those related to (a) the nature of the challenge stakeholders are trying to address, (b) the relevance, strengths and limitations of the innovation/ change/ transformation, and (c) the context in which they are operating.</p>	<p>Very Good</p>	<p>The Creative Tension Feedback Notes provided comprehensive, clear and targeted feedback on the strengths/limitations/tensions of the RBM approach, surfaced and tested assumptions of the overall RBM approach and how it is unfolding/being implemented/manifested in the UNFPA system(s). It is impressively thorough and insightful.</p>
<p>5.3 The evaluation identifies and assesses the sources of resistance to the innovation/ change/ transformation (if any), including the rationale for this resistance and implications for further development and adaptation of the intervention.</p>	<p>Very Good</p>	<p>The evaluation reports do not directly nor consistently refer to the resistance that might emerge in response to RBM, but rather focus on the various tensions that might create or be behind such resistance, including those surfaced in earlier evaluation exercises. This is thoroughly covered in the Feedback Notes and DE Report.</p>
<p><b>6. COMPLEXITY AWARE</b></p>	<p><b>Assessment Level:</b></p>	

<p>6.1 The evaluation reports identify, describe and assess the implications of both intended and unintended effects and consequences of the innovation/ change/ transformation in the theme/ system/ strategy/ programme/ intervention.</p>	<p>Very Good</p>	<p>The Final DE Report (page 26) reviews the unintended consequences of RBM practices in developmental contexts in general (e.g. How the focus on RBM as an accountability mechanism weakens the focus on learning, and in UNFPA in particular (e.g. the cases of some units not aligning with UNFPA policies on RBM).</p>
<p>6.2 The evaluation reports acknowledge the complex cause-and-effect relationships associated with the intervention and tracked those factors and stakeholders new insights about which factors they could control and/or influence, and assessed the implications for the intervention.</p>	<p>Very Good</p>	<p>The challenge of estimating an intervention's level of contribution - rather than attribution - is identified as a structural challenge of RBM in the Inception Report (page 6), more deeply explored in Feedback Note on Collective Versus Individual Accountability, and explored again in the Final DE Report (page 25).</p>
<p>6.3 The evaluation reports track the tensions, dilemmas, paradoxes, and wicked questions related to the theme/ system/ strategy/ programme/ intervention and their implications for further innovation/ change/ transformation.</p>	<p>Very Good</p>	<p>The entire inquiry framework around which the evaluation is organized focuses on the tensions associated with implementing RBM within the UNFPA, as well as their implications for further development and/or adaptation of the approach.</p>
<p><b>7. SYSTEMIC THINKING</b></p>	<p style="text-align: center;"><b>Assessment Level:</b></p>	
<p>7.1 The evaluation reports describe the system(s) in which the theme/ system/ strategy/ programme/ intervention is embedded and/or trying to change in its baseline assessment (i.e. its boundaries, main actors, important relationships, key dynamics).</p>	<p>Very Good</p>	<p>The Inception Report describes how the evaluation is centrally informed by complexity-systemic thinking, employs multiple systems thinking frameworks (e.g. ice-berg metaphor, leverage points, etc.), confirms the inter-relatedness of factors and issues, and identifies these influences throughout the Inception Report, Feedback Notes, and Final DE Report. It further acknowledges the role of cultural and structural factors within UNFPA on the agency's RBM experience, as well the influence of 'exogenous' factors within the larger UN and international development context. The entire evaluation is deeply embedded in a systems approach.</p>
<p>7.2 The evaluation reports track how changes in the larger systems and context in which the theme/ system/ strategy/ programme/ intervention was unfolding create opportunities and barriers for innovation/ change/ transformation, and influence the theme/ system/ strategy/ programme/ intervention's goals, approach or design.</p>	<p>Very Good</p>	<p>The evaluation documents provide a history of the evolution of RBM in the UNFPA, and consistently touches on how factors in the larger context in which the UNFPA was operating (e.g. UN, other international development organizations, etc.) influenced its use in the UNFPA.</p>

<p>7.3 The evaluation reports highlight the diverse perspectives of stakeholders in describing the intervention, the systems in which it operates, and its major developments, including where there was agreement and disagreement across perspectives.</p>	<p>Good</p>	<p>The evaluation reports provide the perspectives of a variety of development agencies, thought leaders, and front-line staff &amp; senior management on the priority tensions and questions at the heart of the assessment. It also points out the diverse understanding of different staff and organizational units in their understanding of RBM and how it is meant to be applied within their spheres of influence. However, these perspectives are only periodically broken down by demographic, organizational unit, or other types of sub-groupings (e.g. geographic).</p>
<p><b>8. TIMELY FEEDBACK</b></p>	<p><b>Assessment Level:</b></p>	
<p>8.1 The evaluation employs a variety of reporting/feedback formats (e.g. memos, PPT presentations, thought pieces, learning briefs) to provide real time evaluative feedback to the primary users.</p>	<p>Very Good</p>	<p>The evaluation combined a variety of feedback formats, including formal reports (e.g. Feedback Notes, Final DE Report), other written documents (e.g. PPTs, feedback on documents), and oral feedback during workshops, ad-hoc and formal meetings. The number and diversity of feedback mechanisms is impressive.</p>
<p>8.2 The evaluation reports describe if and how the timing of the reports submission coincided with formal/structured and/or spontaneous moments of reflection and decision-making by the primary intended users.</p>	<p>Very Good</p>	<p>The Final DE Account confirms that the feedback was provided in three ways: (1) Consolidated, scheduled, feedback on key tensions (6 notes), (2) real time, Requested Feedback, providing on-demand requests for DE feedback (4 instances) and (3) emergent feedback provided on-the-go as the Team felt relevant in 'real time' (multiple instances).</p>
<p>8.3 The evaluation reports point out if and how the real time nature of providing feedback influenced the rigor of the evaluation (e.g. Data collection, sampling, analysis, interpretations, leverage points, or recommendations/ ways forward).</p>	<p>Very Good</p>	<p>The Inception Report and Final DE Report described the limitation of undertaking a point-in-time Developmental Evaluation of a longer-term intervention of RBM implementation in UNFPA, including a risk that the timeframe for the assessment was not long enough to ensure its findings would be used in the longer term development and adaptation of RBM. The Methodological Note surfaced an additional limitation related to the 'sequencing' of evaluation activities, including the how the emergence of new developments within UNFPA (e.g. the launch of UNSDCF Guidance and the subsequent reactions it generated) made the some data either dated or incomplete.</p>
<p><b>9. GENDER EQUALITY &amp; HUMAN RIGHTS</b></p>	<p><b>Assessment Level:</b></p>	
<p>9.1 The evaluation terms of reference confirm that Gender Equality and Human Rights principles and values should be integrated into the developmental evaluation.</p>	<p>Good</p>	<p>The Terms of Reference for the developmental evaluation confirm that the evaluation should follow the UNEG Norms and Standards for Evaluation and abide by UNEG Ethical Guidelines and Code of Conduct. While the Gender Equality and Human Rights is identified as a principle in section 16.c, the Terms of Reference does not directly refer to the principle.</p>

9.2 The evaluation design includes evaluation questions that specifically address issues related to Gender Equality & Human Rights.	Unsatisfactory	The Inception Report does not include specific questions related to Gender Equality and Human Rights.		
9.3 The evaluation includes gender-responsive and human rights responsive evaluation methodology, data collection and data analysis techniques.	Good	The inception report and methodological note describe the ways in which the evaluation integrated a Gender and Human Rights approach into the evaluation, namely: (1) The evaluation used a utilization-focused approach which promotes intended use by intended users, implying a strong focus on participation of users throughout the evaluation process. Following this approach during the evaluation: a. primary users we identified; b. a working group of users was establishes; c. the focus and methods was determined with the users; d. insights, findings, conclusions, leverage points and ways forward were co-created with users. Users from all levels of the organization participated as well as most UNFPA business units as well as all functions. All stakeholders were given the opportunity for meaningful participation. (2) The evaluation design was flexible to easily adapt to changes in user priorities and contexts. (3) A combination of data collection methods (mixed-methods approach) was utilized to gather and analyse information to offer different perspectives and promote participation of different stakeholders. This approach also served to validate findings. (4) Exploring how these areas have been already incorporated into the UNFPA data on RBM; (5) An extra emphasis on engaging people in interviews and surveys in a way that encourage gender and youth perspectives, including ensuring space for open-ended responses by interviewees/surveyed, and active engagement with Tangerine network, a group of young professionals at the UNFPA; and (6) A targeted exploration of how young professionals view and perceive issues related to culture and capacity for RBM at LINEPA		
9.4 The evaluation findings, conclusions and recommendations reflect a Gender Equality and Human Rights perspectives and issues.	Unsatisfactory	The Feedback Notes and DE Account do not integrate or reflect a specific Gender Equality or Human Rights analysis.		
<b>Overall Evaluation Quality Assessment</b>		<b>Assessment Levels (*)</b>		
		<b>Very good</b>	<b>Good</b>	<b>Fair</b>
<b>Quality assessment criteria (scoring points*)</b>				
1. Utilization-Focused	<b>Very Good</b>			
2. Developmental Purpose	<b>Very Good</b>			
3. Co-Creation Stance	<b>Very Good</b>			

4. Rigorous Methods	Very Good			
5. Innovation Focused	Very Good			
6. Complexity Aware	Very Good			
7. Systemic Thinking	Very Good			
8. Timely Feedback	Very Good			
9. Gender Equality & Human Rights			Fair	
<b>Overall assessment level of evaluation reports</b>	Very good very confident to use	Good confident to use	Fair use with caution	Unsatisfactory not confident to use

- (\*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column).
- (b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').
- (c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

**How It Can Be Used?**

The primary use of the evaluation was to provide managers and board of UNFPA with feedback on their ongoing efforts to implement RBM in UNFPA. There are (at least) two other ways the findings could be used. First, the findings may be useful by other development organizations employing RBM methods and techniques: the evaluator team productively draws lessons and insights on the experience of non-UNFPA agencies to inform their efforts and this evaluation expands the evidence based in the field on this topic. Next, the inquiry framework was organized around "tensions", each with their own set of questions and feedback loops (e.g. development notes), which may be a useful evaluation and learning device for other organizations and evaluation teams assessing complex change initiatives.

**What aspects to be cautious about?**

The developmental evaluation was completed without formal guidance on how the quality of evaluation documents and reports would be interpreted and judged.

**Where relevant, please explain the overall assessment Very Good, Good or Unsatisfactory**

See Comments in the introduction. This is a Very Good Developmental Evaluation that sought to provide feedback on the complex task of moving towards the next stage of RBM approach into UNFPA. It unfolded in a relatively quick manner, the parameters relatively clear, the design (i.e. inquiry framework, methods) inventive, and the implementation and feedback, real time, comprehensive, and continuous. The quality of the process, as illustrated in the written documents, is particularly notable given that the evaluation team, commissioners and users did not begin the evaluation with a formal Developmental evaluation's Quality Assurance Framework to guide their efforts.

**Consideration of significant constraints**

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes

No

If yes, please explain:

**Organizational Unit:** UNFPA Evaluation Office **Year of Evaluation:** 2019

**Evaluation:** Developmental Evaluation of Results Based Management in the UNFPA

**EQA Process:** Interviews with Primary Users **Date of assessment:** 30 December 2019

**Interviewees:** The EQA process interviewed five 'primary users' engaged in the developmental evaluation. They were selected to represent different organization units, roles and levels of engagement.

**Overall comments:** The five persons interviewed in the EQA represent different primary users of the developmental evaluation at the UNFPA. They offer diverse perspectives shaped by their different roles and organizational unit of UNFPA HQ. Their reflections reinforce the generally positive review of how the evaluation documents adhere to DE principles, yet also point out some weaknesses in the assessment. These tend to be due less to the quality of the work done by the evaluators, and more related to (1) the 'system-wide' nature of the evaluation, which made it difficult to ensure the process and findings were customized to reflect the unique needs of different organizational units and (2) the UNFPA staff expectation that developmental evaluations should operate like traditional evaluations and generate very specific recommendations for improvement and development, rather than options and pathways to consider when deciding how to develop an intervention further. These issues will be addressed directly in the summary EQA Report.

**Assessment Level:** Very Good (Strong, Above Average), Good (Satisfactory, respectable), Fair (Some weaknesses, still respectable), Unsatisfactory (Weak, does not meet minimum quality standards), Don't Know, Insufficient Data or Uncertain

**PRINCIPLES**

	Interviewee #1	Interviewee #2	Interviewee #3	Interviewee #4	Interviewee #5	
<b>1. UTILIZATION FOCUSED</b>						Overall Assessment Level: <b>Very Good</b>
How would you rate the extent to which the evaluation focused on questions that were useful to your work?	Not Rated	Very Good	Very Good	Very Good	Very Good	All interviewees reported that the evaluation design included questions were directly related to change management and IT systems, as well as overall strategy and policy.
How would you rate the extent to which the evaluation process was user friendly (e.g. options for being involved, ease of interaction)?	Very Good	Very Good	Very Good	Good	Good	Three interviewees reported that the process for engagement was well laid out for everyone and evaluators and UNFPA made many opportunities available to be involved, and communicated regularly. Interviewees referred to the process as "open", "engaging" and "easy to access". Two interviewees pointed out the challenge of trying to engage so many UNFPA staff person located around the world.
How would you rate the extent to which evaluation documents and reports were easy to read and understand and the verbal communications were effective?	Very Good	The documents were well written and easy to navigate. The verbal communications were clear and professional. Two interviewees pointed out that the volume of the information was difficult to keep up to for some users and/or their staff teams.				
How would you rate the extent to which the evaluation generated findings and discussions were useful in understanding how RBM is unfolding in the UNFPA and ways that RBM can be adapted or improved for greater effectiveness?	Very Good	Fair	Fair	Very Good	Very Good	The findings of the evaluation were well organized. The analysis offered some important insights into the causes of RBM implementation challenges in UNFPA, with one person noting that "we had a hunch that these were the issues, perhaps even knew them, but the evaluation confirmed it and legitimized our hunches so we can better address them in the future". All five interviewees reported that the implications for moving forward (e.g. points of leverage, entry points, principles) were useful, but three reported that they and/or their colleagues had hoped for more specific recommendations on how to proceed, either because they were already somewhat aware of the reasons underlying the implementation challenges and/or they were accustomed to receiving more detailed recommendations from other evaluations.
<b>2. DEVELOPMENTAL PURPOSE</b>						Assessment Level: <b>Fair</b>
How would your rate the extent to which the evaluation/evaluators provide data/evidence informed implications for the ongoing development and adaptation of RBM in the UNFPA?	Fair	Fair	Fair	Good	Very Good	Interviewees reported that the evaluation did a good assessment of the factors and tensions related to uneven usefulness of RBM in UNFPA but also a desire for more elaborated and UNFPA unit-specific options for moving forward. Two interviewees reported "feeling overwhelmed" with the feedback, and uncertain what to do with it, while another said that while they felt the implications were "very good", his/her team members wanted more detailed recommendations.
<b>3. CO-CREATION STANCE</b>						Assessment Level: <b>Good</b>
How would you rate the extent to which the evaluation/evaluation team co-created and continued to adapt the evaluation questions, design and implementation with the primary users of the evaluation?	Very Good	Fair	Good	Good	Very Good	All interviewees reported that evaluators checked in several times throughout the process, were responsive and attentive to questions and feedback for adjustments in the evaluation approach. One reported that the evaluators spent significant time in analyzing "our contexts" in order to better understand the RBM experience at the UNFPA. Another noted "Almost every time [the evaluators] identified something new, they shared it, engaged people to get feedback and generate ownership. I appreciate this approach". Two interviewees reported that the UNFPA Evaluation Office made many attempts to engage people but peoples' busy schedules made it difficult or the level of expectation required was "very demanding".
<b>4. INNOVATION FOCUSED</b>						Assessment Level: <b>Very Good</b>
How would you rate the extent to which the evaluation identified the significance of the RBM approach to the UNFPA, as well as the implementation challenges and resistance that might have emerged with its adoption?	Very Good	Not Rated	Not Rated	Good	Very Good	Interviewees reported that the RBM has an 'innovation' that has been around for a while, and that previous evaluations and reviews have surface key issues related to its adoption and implementation. The unique contribution of this evaluation was in trying to go deeper on some of the cultural and structural challenges in its adaptation and ongoing improvement. The inquiry framework organized around tensions, and the urge to uncover root causes, was helpful in "going deeper" on these challenges and resistance. One interviewee called the analysis, "right on the money".
<b>5. SYSTEMIC THINKING</b>						Assessment Level: <b>Very Good</b>

How would you rate the extent to which the evaluation/evaluators took into account the various formal and informal systems within and beyond the UNFPA (e.g. specific UNFPA units, the larger UN, other development agencies, different countries), that might influence how RBM is perceived, adopted and experienced in the agency?	Good	Very Good	Very Good	Very Good	Very Good	Interviewees reported that the evaluation team was very aware of systems in which RBM operated, specifically the culture, standards and structures of UNFPA. Two interviews pointed out the uneven alignment between RBM approach and systems related to Information Technology, Human Resources, etc. A third was highly appreciative of learning more of the comparative experience of other private, public and non-profit institutions with RBM.
<b>6. COMPLEXITY-AWARE</b>						Assessment Level: <b>Very Good</b>
How would you rate the extent to which the evaluation/evaluators surfaced and made sense of the complex array of factors related to the adoption of RBM in the UNFPA, as well as the tensions, unpredictable and unintended consequences that might have emerged?	Very Good	Very Good	Very Good	Good	Very Good	Interviewees reported the the evaluation team was "more aware of some of the complexities than some people in HQ" because of the broad based discussion across the UNFPA. There were three positive comments provided to support the high rating: one interviewee pointed out that the distinction between management in controlled and uncontrolled environments was a useful complexity-informed insight; another identified the usefulness of the stakeholder mapping - and different perspectives that emerged from the exercise - involved in RBM; a third felt that the evaluation nicely surfaced the tendency for some staff to be "binary" in their perception of RBM (e.g. "We are either focused on RBM or we are a learning organization") and offered people a way to help people think of them as tensions instead. One interviewee reported dissatisfaction with the approach the evaluation employed for assessing the risks of RBM, who felt it "needed more work".
<b>7. RIGOROUS PROCESS</b>						Assessment Level: <b>Very Good</b>
How would you rate the extent to which the evaluation/evaluators rigorously collected, and interpreted data and generated findings in way that drew on multiple data sources, triangulated data, and made evaluative statements with a strong link to the evidence/data?	Very Good	Very Good	Very Good	Very Good	Good	Interviewees reported being impressed with the number and variety of internal and external people engaged in the evaluation, the use of evidence based on other RBM experiences, and that the evaluation findings were "accurate, balanced, meticulous and transparent." One interviewee pointed out that the "unique" approach the evaluators used to "create knowledge" in the evaluation, using frameworks, that were meant to encourage critical thinking skills and deeper understanding of key issues was embraced by some staff but was considered too "abstract" by others. The interviewee who rated the evaluation rigor as "good" mentioned it was too easy in some cases to attribute quotes or sentiments in the evaluation findings, and felt that this "primary data" could have been more "buried" in the analysis process.
<b>8. TIMELY FEEDBACK</b>						Assessment Level: <b>Very Good</b>
How would you rate the extent to which the evaluation/evaluators facilitated the evaluation discussions and feedback in real time?	Very Good	Unsure	Unsure	Very Good	Good	The ratings were good, though the interviewees's experience with the real time nature varied. One interviewee reported having had several interactions, but would have liked to have had more interactions, and organized around their own Unit-specific timelines and decision-making moments. One person pointed out that the evaluation was particularly timely because it generated findings that were useful for its Division given its existing efforts to update its plan or approach. A third interviewee reported that whenever there was a development, the evaluators shared it widely, and whenever they required some feedback or interaction, the evaluators were able to respond. One interviewee argued that the evaluation was not as real time as it could have been, because they were already working on the issues surfaced in the assessment and felt like their were waiting for the evaluation to finish before they were allowed to proceed: some of their team was "exhausted" by the time the final evaluation report appeared.
<b>9. IMPARTIALITY &amp; INDEPENDENCE</b>						Assessment Level: <b>Very Good</b>
How would you rate the extent to which the evaluation/evaluators operated with impartiality and independence throughout the developmental evaluation?	Very Good	Interviewees reported that the evaluators offered balanced feedback without "getting people riled up" because it was based in evidence and data. Another noted, "They were excellent listeners – they wanted to understand things well. They were trying to form an independent and solid opinion. I don't have an example of impartiality, but the overall approach on collecting data and information contributes to impartiality. They were very data-driven. They wanted to show the diversity of perspectives rather than give one voice more than others. I have confidence in the impartiality of the findings." A third interviewee noted, there was "pushback" on key ideas, like adaptive management, and the evaluators "stood behind it", yet also reported that the team ended up framing things in a more appreciative manner. A final interviewee noted that the evaluators "were dealing with a lot of highly opinionated people, but they used those opinions as data or evidence: I never saw them succumb to any pressure to say certain things".				
<b>10. CREDIBILITY</b>						Assessment Level: <b>Very Good</b>
How would you rate the expertise, competence and credibility of the evaluation/evaluators?	Very Good	Very Good	Very Good	Good	Good	Three interviewees provided a very good rating. One noted, "The evaluators were good: they knew what they were talking about". Another stated, "They were beyond very good. They were extremely professional and knowledgeable about the substance of things. Very good communication, explaining things, attention to details, personable and engaging. They were also very enthusiastic, energetic and passionate about the work." Two interviewees reported that they felt that the evaluation team was very skillful, diligent, professional and offered a different perspective, but had hoped for evaluators with more lead-edge RBM experience on the team: they were careful to point out that this did not substantially undermine the quality of the evaluation.
<b>11. OTHER ISSUES</b>						
Is there anything else you would like us to know about your experience with the developmental evaluation and/or its overall quality?	Yes	Yes	Yes	No	Yes	(1) I was pleasantly surprised with how useful this approach can be. Great interactions, "learned a lot from them, sparked new insights." I have already made adaptations in my approach; (2) All I can say is the overall experience was positive. Perhaps Developmental Evaluations could be done again, with something less tough than RBM, even something more exciting. The feedback I heard from other sources about the experience and consultants was also positive. The recommendation to adopt adaptive leadership will help with the receptivity for developmental evaluation: it will make its application bigger and wider and even more impactful; (3) I hope developmental evaluation will create some space for learning in the UNFPA as a greater constant.
<b>Quality Assessment Summary</b>						
1. Utilization-Focused	Very Good					Relevant questions, many opportunities to engage, clear communications.
2. Developmental Purpose	Fair					Findings on analysis of successes and challenges clear and substantiated, findings offered future direction: some desire for fewer, more detailed, targeted recommendations.
3. Co-Creation Stance	Good					Participatory, engaging process co-creation, inquiry and sense-making: some interviewees felt too busy to be involved and/or overwhelmed.
4. Rigorous Process	Very Good					High level of confidence in the breadth and diversity, interpretation and evidence-based findings.
5. Innovation Focused	Very Good					Strong sense that the evaluation offered insights into RBM experience, tensions and resistance.
6. Systemic Thinking	Very Good					Evaluation questions, inquiry, analysis and findings strongly embedded in an awareness and analysis of systemic factors.
7. Complexity Aware	Very Good					New insights into the diverse perspectives of multiple UNFPA stakeholders, the implications of managing in less and more controlled environments, and appreciation of unintended consequences.
8. Timely Feedback	Very Good					An appreciation for the ongoing and spontaneous feedback and interactions: yet, sense that process and findings did not line up with the many and diverse 'decision-windows' of individual organizational units.
9. Independence & Impartiality	Very Good					High regard for the attentiveness of evaluators to user questions and feedback, and willingness to skillfully communicate findings in a balanced manner.
10. Credibility	Very Good					High regard for the professionalism, knowledge and skills of the evaluation team, though sense that team were not primarily RBM experts.
<b>Overall assessment level of the evaluation process by primary users</b>	Very Good					





**Ensuring rights and choices for all**

United Nations Population Fund  
Evaluation Office  
605 Third Avenue  
New York, NY 10158 USA



[evaluation.office@unfpa.org](mailto:evaluation.office@unfpa.org)



[unfpa.org/evaluation](http://unfpa.org/evaluation)



[unfpa\\_eval](https://twitter.com/unfpa_eval)



[UNFPA Evaluation Office](https://www.youtube.com/UNFPA-Evaluation-Office)