



4FRONT

META-EVALUATION AND –ANALYSIS OF MFA INNOVATION PROGRAMMES

Report on findings and conclusions, 7.6.2016

CONTENT

1. Background and context
2. Meta-evaluated and analysed programmes
3. Meta-evaluation – methodology, findings and conclusions and identified good practices
4. Meta-analysis – methodology, findings and conclusions, lessons
5. Summa summarum

1. Background and context

This report presents the methodology, findings and conclusions of the Work package 2.1 of ‘Developmental Evaluation of BEAM’ concerning *the Meta-evaluation and –analysis of MFA Innovation Programmes*.

The initial findings were presented in a Workshop held 10 May, 2016. Notes from the discussions have been included in this report.

The work has been conducted during January – May 2016 by a consortium led by 4FRONT. The responsible evaluators were Kimmo Halme, Kristiina Lähde, Juho Uusitalo and Petri Uusikylä.

PURPOSE AND OBJECTIVES (TOR)

- The purpose and objectives of the combined meta-evaluation and analysis were to:
 - Learn lessons of methodologies applied in evaluating innovation projects. These lessons will be used in BEAM-evaluation by the evaluation team and evaluation management to strengthen the developmental evaluation process.
 - Summarise the results of MFA supported innovation projects and understand reasons for successes and failures based on information included in the selected evaluation reports. This information can be used by BEAM programme management and the MFA in future cooperation programmes to promote innovation for improved quality of cooperation and better results.

META-EVALUATION VS META-ANALYSIS

- **Meta-evaluation** refers here to an evaluation of evaluations.
- **Meta-analysis** refers here to a synthesis of existing programme evaluation studies in a given area. Summarising the current knowledge about a particular type of programme.

'Analytical synthesis of evaluation findings, outcomes and lessons'

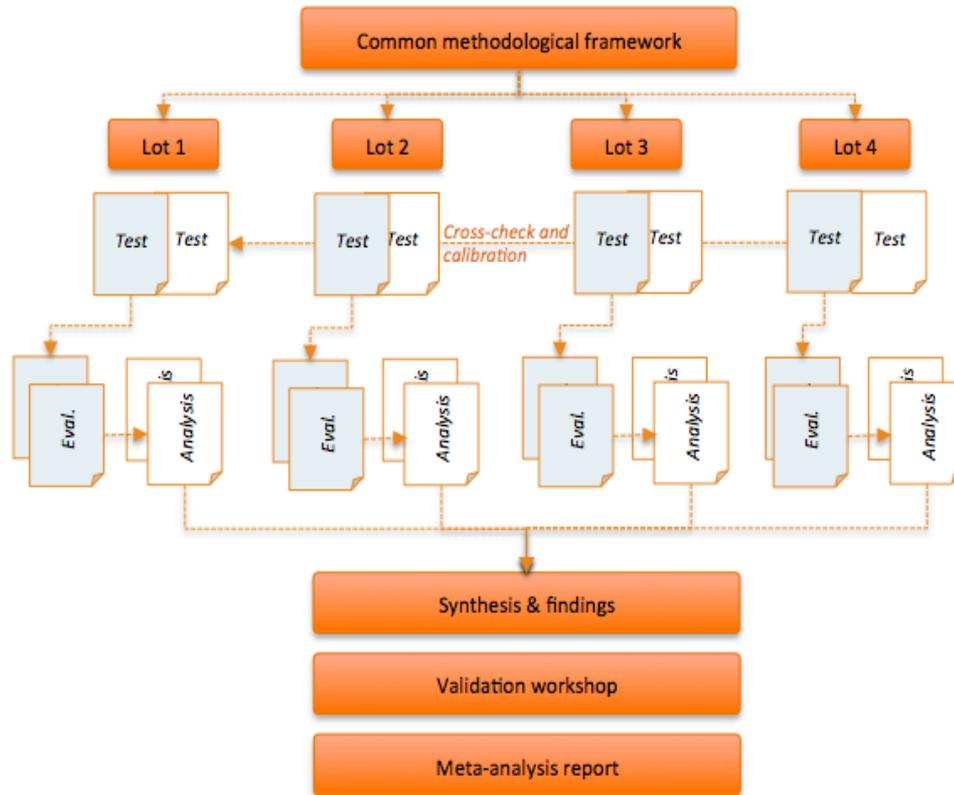
BACKGROUND OF THE ASSIGNMENT

- Ministry for Foreign Affairs (EVA-11) commissions meta-evaluations approximately every two years. Previous meta-evaluations were conducted in 1996, 2007, 2009, 2012 and 2015.
- Meta-evaluations are planned to inform decisions on Finland's policy for development co-operation.
- Meta-evaluations also inform the capacity development products and services provided by EVA-11, including revisions to the guidance provided to MFA staff and partners through evaluation manuals and training.
- This is the first MFA commissioned meta-evaluation focusing on one single "sector" – innovation. All the earlier ones have been cross-cutting meta-evaluations.

QUESTIONS TO BE ANALYSED (TOR)

1. What is the quality of MFA's innovation projects based on the OECD/DAC evaluation standards?
2. What evaluation designs and methods work best in evaluating innovation projects?
3. What can be said about the quality of Finnish innovation projects based on the reliable evaluation reports by each OECD/DAC criteria?
4. What are the main success stories, good practices and challenges as well as the reasons behind these emerging from the innovation project evaluation reports?

WORK PROCESS



2. Meta-evaluated and analysed programmes

EVALUATED PROGRAMMES

- Covers 12 evaluation reports in the field of Science, Technology and Innovation
- Five evaluation reports were included in previous meta-evaluations commissioned by the MFA
- Meta-analyses of EEP-Central America, and AEA + MFS in the Andean region, have been made based on their English language executive summaries. Meta-evaluation has not been attempted.

Programme	MFA meta-evaluation	Meta-evaluated	Meta-analysed
1. AEA+MFS (MTR)			✓
2. ALICT (MTR)		✓	
3. BioFISA	✓	✓	✓
4. CSBKE (MTR)		✓	✓
5. EEP –Central America (MTR)			✓
6.-7. EEP S&EA + Mekong (MTR)	✓	✓	✓
8. IIP Vietnam		✓	✓
9. SAFIPA		✓	✓
10. SAIS	✓	✓	✓
11. STIFIMO	✓	✓	✓
12. TANZICT	✓	✓	✓

3. Meta-evaluation of innovation programmes

*- methodology, findings and conclusions
and identified good practices*

META-EVALUATED INNOVATION PROGRAMMES

- Meta-evaluation covered 10 programmes.
- In order to ensure equal working process those already previously meta-evaluated programmes were included as well.
- EEP-Central America and AEA + MFS in the Andean region were excluded as they only have executive summaries in English.

Programme	Meta-evaluated	Meta-analysed	MFA meta-evaluation
1. AEA+MFS (MTR)		✓	
2. ALICT (MTR)	✓		
3. BioFISA	✓	✓	✓
4. CSBKE (MTR)	✓	✓	
5. EEP –Central America (MTR)		✓	
6-7. EEP S&EA + Mekong (MTR)	✓	✓	✓
8. IIP Vietnam	✓	✓	
9. SAFIPA	✓	✓	
10. SAIS	✓	✓	✓
11. STIFIMO	✓	✓	✓
12. TANZICT	✓	✓	✓

METHODOLOGY OF META-EVALUATION

- Formative assessment of the evaluation reports based on the OECD/DAC and EU standards, focusing on the following areas.

**Answers to questions on 4. sources of information and 5. methodology were merged here, as they are inseparable in practice. Impossible to answer separately.*

1. Executive summary
2. Context
3. Intervention logic
4. Sources of information*
5. Methodology
6. Analysis
7. Answers to ToR evaluation questions
8. Limitations
9. Difference of opinion
10. Stakeholders comments

Meta-evaluation doesn't reveal much about the quality of the evaluation, it only states whether required questions have been covered and sections included

Q1. EXECUTIVE SUMMARY

Ratings are shown in random order in each slide.

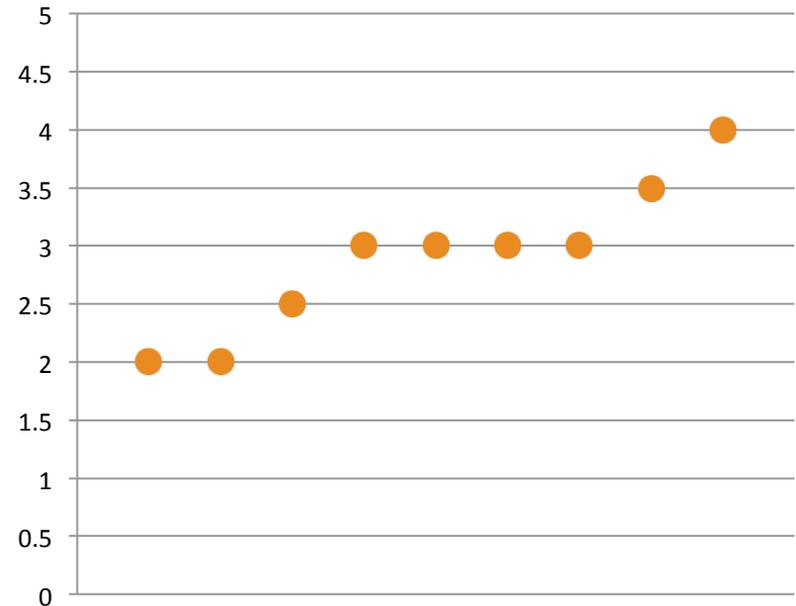
Report quality checklist:

- Contains a clear and representative executive summary of the report?
- Summarises the main findings, conclusions, recommendations in a summary table?
- Presents overall lessons learned?

Rating scale:

1 = criteria mostly not fulfilled or totally absent
2 = criteria partially fulfilled
3 = criteria mostly fulfilled
4 = criteria entirely fulfilled
5 = criteria entirely fulfilled in a clear and original way
n/a = not addressed (shown as 0)

Ratings of executive summaries



FINDINGS & CONCLUSIONS

- Executive summaries, when they exist, are usually well written. This is not the weak point of evaluations.
- Requirement for a summary table was introduced later and is therefore not part of all ES
- Typical weaknesses are too generic descriptions

WORKSHOP DISCUSSION

- MFA does not yet have guidelines on the length of the executive summary.
- It was suggested that the recommended number of pages could be set for 2, in addition to the summary table.

Q2. CONTEXT

Report quality checklist:

- Describes the context of the development programme, and possible changes in it?
- Describes/assesses the influence of the context on programme performance?

Rating scale:

1 = criteria mostly not fulfilled or totally absent

2 = criteria partially fulfilled

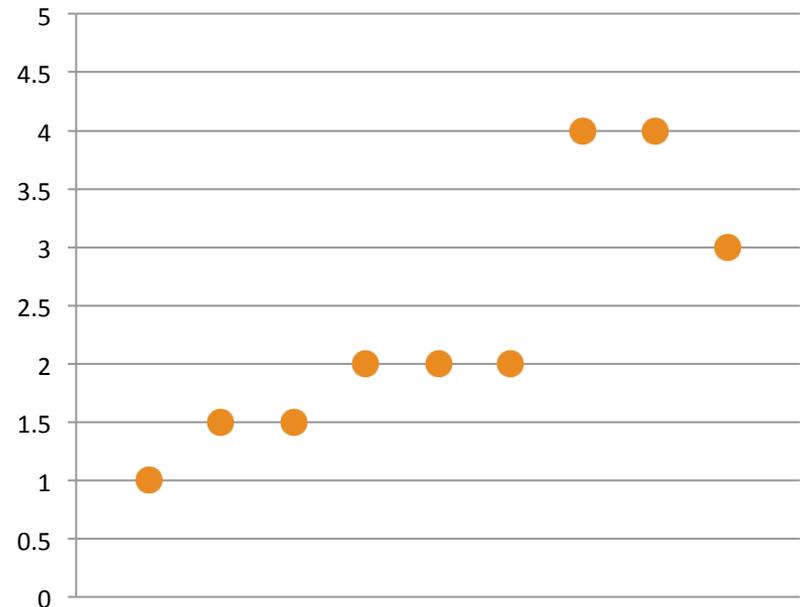
3 = criteria mostly fulfilled

4 = criteria entirely fulfilled

5 = criteria entirely fulfilled in a clear and original way

n/a = not addressed (shown as 0)

Ratings of context



FINDINGS & CONCLUSIONS

- Big variation here. Most evaluations do not put effort on explaining the context, even if they mention that context has changed!
- This is one of the weak points of the evaluations.
- Typical weakness is a too short or generic description of the context.
- One reason for this may be that the overall context is not properly analysed / documented / described in the programme in the first instance!

WORKSHOP DISCUSSION

- Weak contextual analysis in the report doesn't necessarily mean that evaluators' understanding of the context would be weak - it just hasn't been addressed in a report.
- Weaknesses in the contextual descriptions may also reflect the tendency to focus more on outputs and activities. Stronger emphasis should be placed on contextual description and analysis already in inception reports.
- Understanding the context is important with regard to all aspects of the programme design and evaluation.
- Descriptions of the context are often qualitative, but it would be useful to include also quantitative data / analyses as well.

IDENTIFIED GOOD PRACTICE #1

A good practice concerning:

- Description of programme context in the evaluation

Name of referenced project & report:

- Mid-term evaluation of the Southern Africa Innovation Support Programme (SAIS), April 2014, EPRD

What and why?

- The programme context i.e. the STI systems in four countries participating SAIS programme is presented in Annex 7. This chapter is well-written and gives good overall picture of the institutional features and baselines in Namibia, Botswana, Mozambique and Zambia.
- Indicators used and statistics applied are valid based on most recent data available.
- Internet links give reader a good opportunity to seek for additional information.

Q3. INTERVENTION LOGIC

Report quality checklist:

- Describes and assesses the intervention logic (e.g. in the form of a logical framework) or theory?
- Describes and assesses the underlying assumptions and factors affecting the success of the programme?
- Takes into account the evolution of the programme?

Rating scale:

1 = criteria mostly not fulfilled or totally absent

2 = criteria partially fulfilled

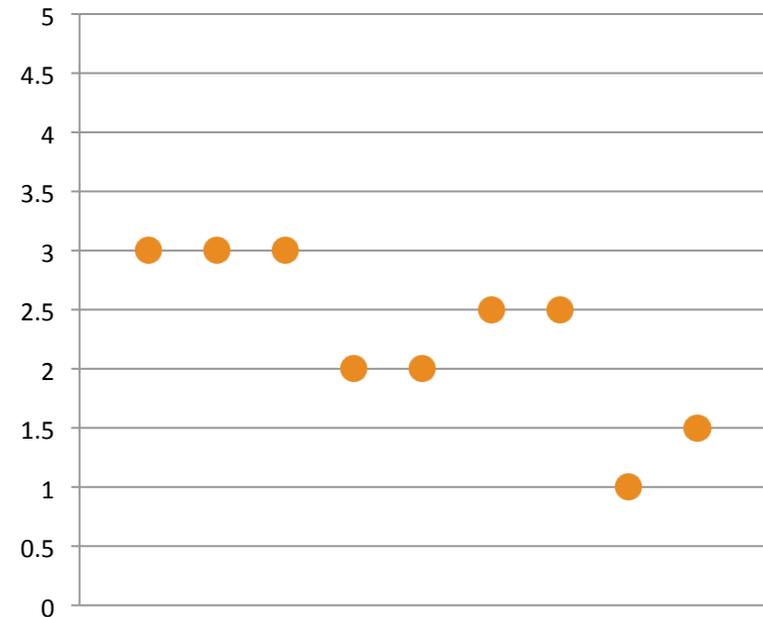
3 = criteria mostly fulfilled

4 = criteria entirely fulfilled

5 = criteria entirely fulfilled in a clear and original way

n/a = not addressed (shown as 0)

Ratings of intervention logic



FINDINGS & CONCLUSIONS

- Big variations, but generally this is described in the evaluation (as it forms the logic of the evaluation, too).
- Underlying assumptions often missing.
- Evolution of the intervention logic properly considered??
- Typical weaknesses are too generic or insufficiently explained intervention logic.
- One obvious reason for this may be that the intervention logic has not been sufficiently clear in the programme plan either!?

WORKSHOP DISCUSSION

- Evolution within innovation programmes are common (by nature) and should be taken into account.
- Although it is important for innovation programmes to maintain certain level of flexibility, a constantly changing intervention logic makes it difficult to validate the chain from planned actions to realised outcomes.
- When changes are made, a systematic documentation including explanations on why certain changes have been made, is necessary. If this is missing, the evaluation should point it out.
- When looking at the intervention logic and its changes the evaluator must always observe whose interest are being served.

Q4. METHODOLOGY AND SOURCES OF INFORMATION

Report quality checklist:

- Annexed to the report explains and justifies the evaluation methodology and its application, including techniques used for data collection and analysis?
- Describes the sources of information (documents, interviews, other) used so that the adequacy of the information can be assessed?
- Explains the selection of case studies or any samples?
- Cross-validates the information sources?
- Critically assesses the validity and reliability of the data?
- Explains limitations and shortcomings, risks and potential biases associated with the evaluation method?
- Lists used methods for data collection and analysis?

METHODOLOGY AND SOURCES OF INFORMATION

Typical evaluation methods (examples listed by the evaluation team):

- Document analysis: e.g. consulted documents, statistics, contextual analysis
- Operational analysis: e.g. basic facts / budgets
- Quantitative methods: e.g. surveys, statistical analyses
- Qualitative methods: e.g. interviews (per category), focus group discussions
- Interactive methods: e.g. stakeholder workshops, validation sessions
- Use of evaluation matrix/quality assurance

Rating scale:

1 = criteria mostly not fulfilled or totally absent

2 = criteria partially fulfilled

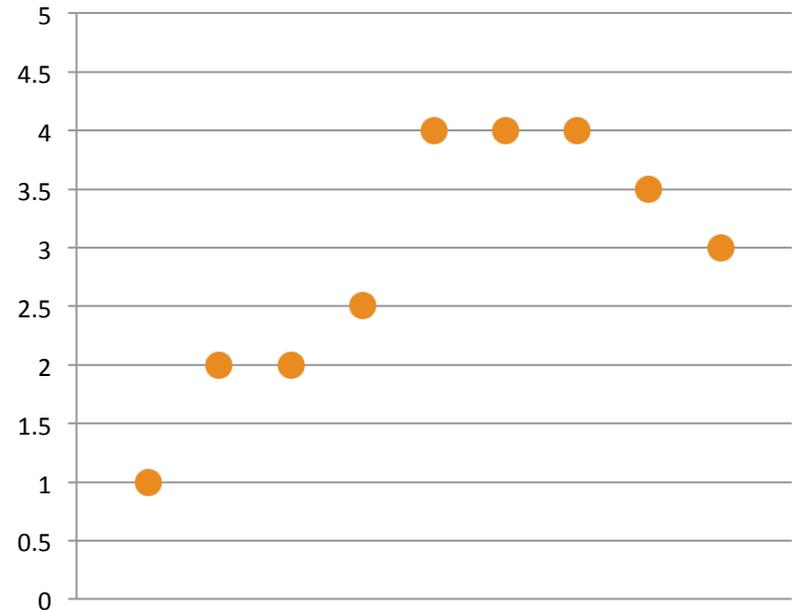
3 = criteria mostly fulfilled

4 = criteria entirely fulfilled

5 = criteria entirely fulfilled in a clear and original way

n/a = not addressed (shown as 0)

Ratings of methodology



FINDINGS & CONCLUSIONS

- Big variations, but generally the methodology is properly described
- Typical weaknesses are that there is not enough details and facts for the reader to judge the appropriateness / validity
- The reasoning for choice of methods usually not explained (triangulation etc)
- Professional use of a set of appropriate research and analysis methodologies is the backbone of all evaluations. It is our interpretation that this question should be further elaborated, more specific and carry higher weight amongst the meta-evaluation questions.

WORKSHOP DISCUSSION

- The evaluation design should reflect ToR questions and the needs of each particular evaluation.
- More attention should be paid on the good choice & sufficient versatility of analysis methods. Evaluation should explain the choice and appropriateness of methods.
- Data collection methods used in evaluations are often traditional (interviews, questionnaires and document review). New methods should be encouraged.

IDENTIFIED GOOD PRACTICE #2

A good practice concerning:

- Methodology and sources of information

Name of referenced project & report:

- Independent mid-term evaluation of the CSBKE Programme (InfoDev's programme financed by bilateral MFA-WB Trust Fund), September 2013

What and why?

- Methodology is well-explained.
- Sources of information are described / listed.
- Limitations to methodology are discussed and mitigation strategy given.

IDENTIFIED GOOD PRACTICE #3

A good practice concerning:

- Methodology and sources of information

Name of referenced project & report:

- Mid-term review of Tanzania Information Society and ICT Sector Development Project (TANZICT), April 2013

What and why?

- Very good qualitative touch and detailed notes from interviews.
- Sources of information well-described
- Through presentation of feedback and information from interviews.

Q5. ANALYSIS

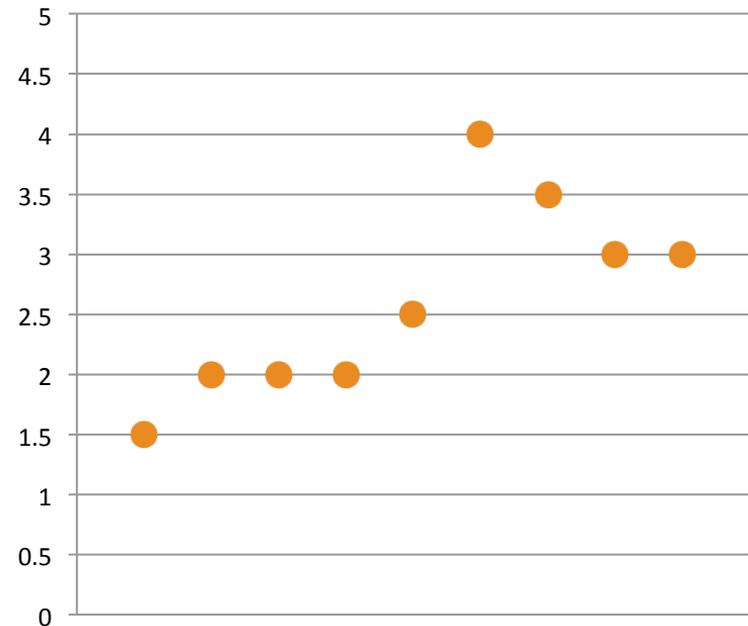
Report quality checklist:

- Presents clear analysis covering findings, conclusions, recommendations and lessons separately and with a clear logical distinction between them.
- Makes explicit the assumptions that underlie the analysis.

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of analysis



FINDINGS & CONCLUSIONS

- Big variations
- Good practices could be shared here (how to nicely present analysis...)
- Typical weakness is that analyses are not systematically done and presented – making sure all relevant aspects are properly considered
- One reason behind could be that some evaluations tend to use a lot of qualitative (non-structured) information and its analysis is more demanding.

WORKSHOP DISCUSSION

- As much as possible, evaluation should consist of both qualitative and quantitative analyses and results should be based on transparent information and logical thinking process.
- The use of direct citations to complement the text is much welcomed to illustrate stakeholders views.

IDENTIFIED GOOD PRACTICE #4

A good practice concerning:

- Analysis, findings, conclusions, recommendations and underlying assumptions

Name of referenced project & report:

- Mid-term evaluation of the Co-operation in Science, Technology and Innovation between Finland and Mozambique (STIFIMO), July 2013

What and why?

- Findings and conclusions presented clearly.
- Tables were easy to read and argumentation logic easy to follow.

IDENTIFIED GOOD PRACTICE #5

A good practice concerning:

- Analysis, findings, conclusions, recommendations and underlying assumptions

Name of referenced project & report:

- Independent mid-term evaluation of the CSBKE Programme (InfoDev's programme financed by bilateral MFA-WB Trust Fund), September 2013

What and why?

- Findings, conclusions, lessons and recommendations are presented separately

Q6. ANSWERS TO TOR EVALUATION QUESTIONS

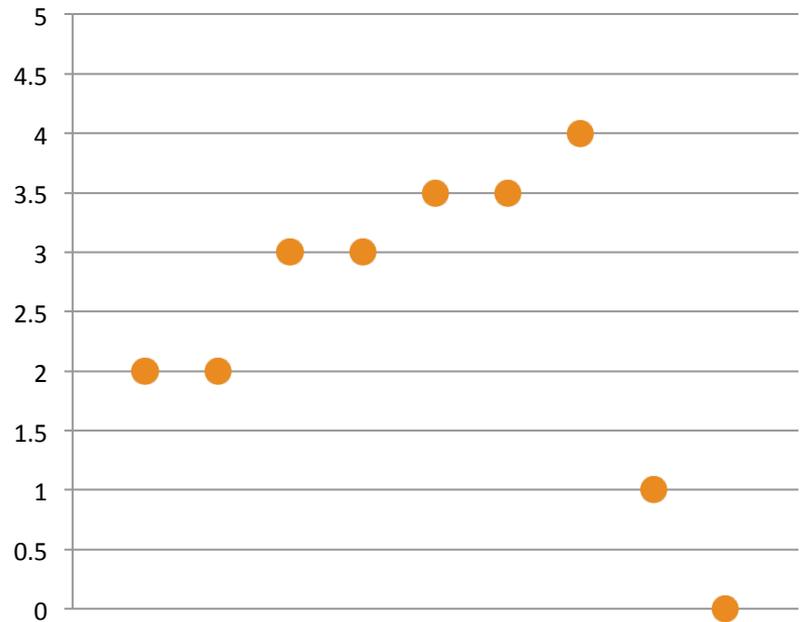
Report quality checklist:

- answers all the questions detailed in the TOR for the evaluation
- covers the requested period of time, and the target groups and socio-geographical areas linked to the programme
- if not, justifications are given

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of answers to ToR evaluation questions



FINDINGS & CONCLUSIONS

- Surprisingly, very big variations
- This is/should be often ensured by the client, too.
- Typical weakness is that the ToR text is not included in the evaluation report, so it cannot be verified
- Some evaluations are part of a larger evaluation

IDENTIFIED GOOD PRACTICE #6

A good practice concerning:

- Answering the evaluation questions detailed in the ToR

What and why?

- All questions answered and ToR annexed in the evaluation

Name of referenced project & report:

- Mid-term review of the Energy and Environment Partnership Programme (EEP) with Southern and East Africa and with Mekong

Q7. LIMITATIONS

Report quality checklist:

- explains any limitations in process, methodology or data, and discusses validity and reliability
- indicates any obstruction of a free and open evaluation process which may have influenced the findings
- explains any discrepancies between the planned and actual implementation and products of the evaluation

Rating scale:

1 = criteria mostly not fulfilled or totally absent

2 = criteria partially fulfilled

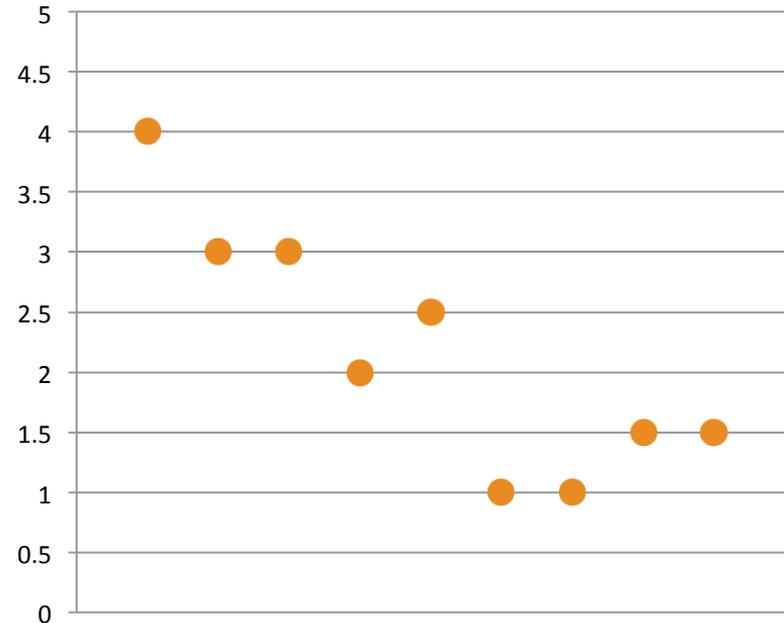
3 = criteria mostly fulfilled

4 = criteria entirely fulfilled

5 = criteria entirely fulfilled in a clear and original way

n/a = not addressed (shown as 0)

Ratings of limitations



FINDINGS & CONCLUSIONS

- This is important question. Some evaluations have obvious limitations (in terms of coverage / suitability of their methods + reliability of results), while only some present this openly.
- Should be included in all reports
- Not only limitations to be mentioned, but to discuss how these have been addressed or mitigated and what are the implications!

IDENTIFIED GOOD PRACTICE #7

A good practice concerning:

- limitations in process, methodology or data, and discussion of validity and reliability

What and why?

- Limitations and mitigation strategies are presented in a clear table format

Name of referenced project & report:

- Independent mid-term evaluation of the CSBKE Programme (InfoDev's programme financed by bilateral MFA-WB Trust Fund), September 2013

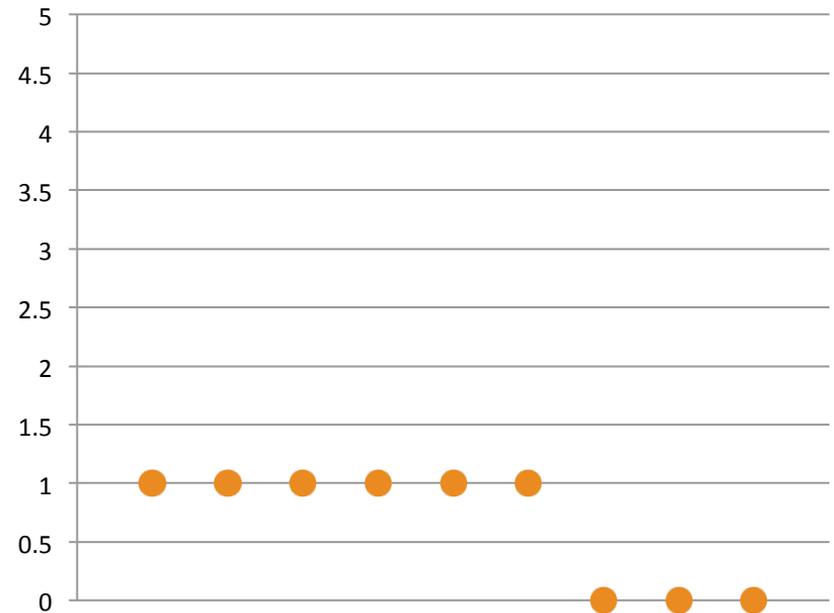
Q8. DIFFERENCES OF OPINION

- Report quality checklist:
 - acknowledges unresolved differences of opinion within the evaluation team

Rating scale:

1 = criteria mostly not fulfilled or totally absent
2 = criteria partially fulfilled
3 = criteria mostly fulfilled
4 = criteria entirely fulfilled
5 = criteria entirely fulfilled in a clear and original way
n/a = not addressed (shown as 0)

Ratings of differences of opinion



FINDINGS & CONCLUSIONS

- Differences on opinion seems to be a very rare case. Most evaluation teams tend to come to a consensus of opinion.
- Differences amongst evaluators or stakeholders (how openly reported)?
- Only to be reported when exists vs requirement?

Q9. STAKEHOLDERS COMMENTS

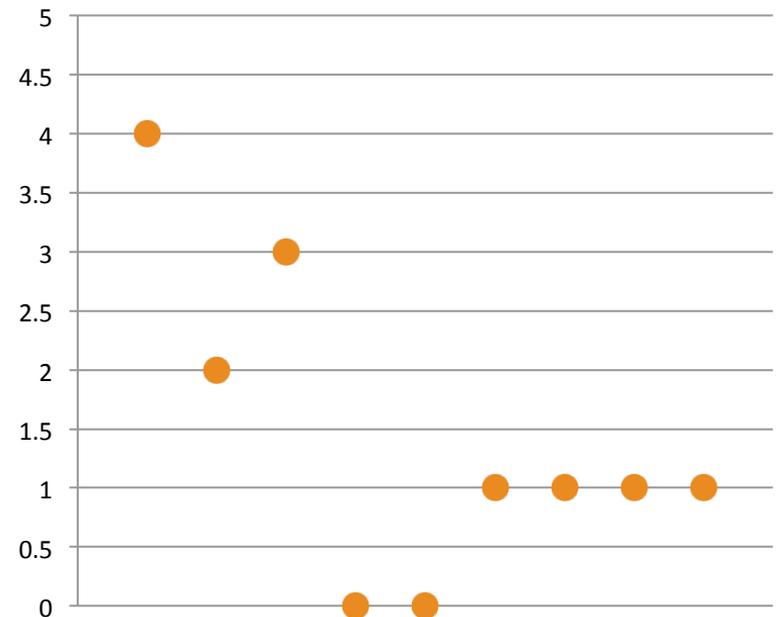
Report quality checklist:

- Reflects stakeholders' comments on the report and acknowledges any substantive disagreements

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of stakeholders comments



FINDINGS & CONCLUSIONS

- Most evaluations do not report direct feedback of stakeholders
- Not all evaluations methodologically ensure stakeholders are closely involved in the evaluation process (debriefings / workshops)
- Direct feedback / comments etc from stakeholders increase the openness of the evaluation process and the credibility of results.
- This should be a good practice to be recommended (vs requirement?)

IDENTIFIED GOOD PRACTICE #8

A good practice concerning:

- Stakeholder comments on the evaluation report

Name of referenced project & report:

- Independent mid-term evaluation of the CSBKE Programme (InfoDev's programme financed by bilateral MFA-WB Trust Fund), September 2013

What and why?

- The method chapter describes the feedback process to the draft report, how the comments were collected and what additional documents were reviewed to inform the final report.

IDENTIFIED GOOD PRACTICE #9

A good practice concerning:

- Stakeholder comments on the evaluation report

Name of referenced project & report:

- Mid-term review of Tanzania Information Society and ICT Sector Development Project (TANZICT), April 2013

What and why?

- Very comprehensive and systematic synthesis of the comments by stakeholders.
- Note: these were not comments to the report, but results of the evaluation interviews.

GENERAL / OVERALL OBSERVATIONS

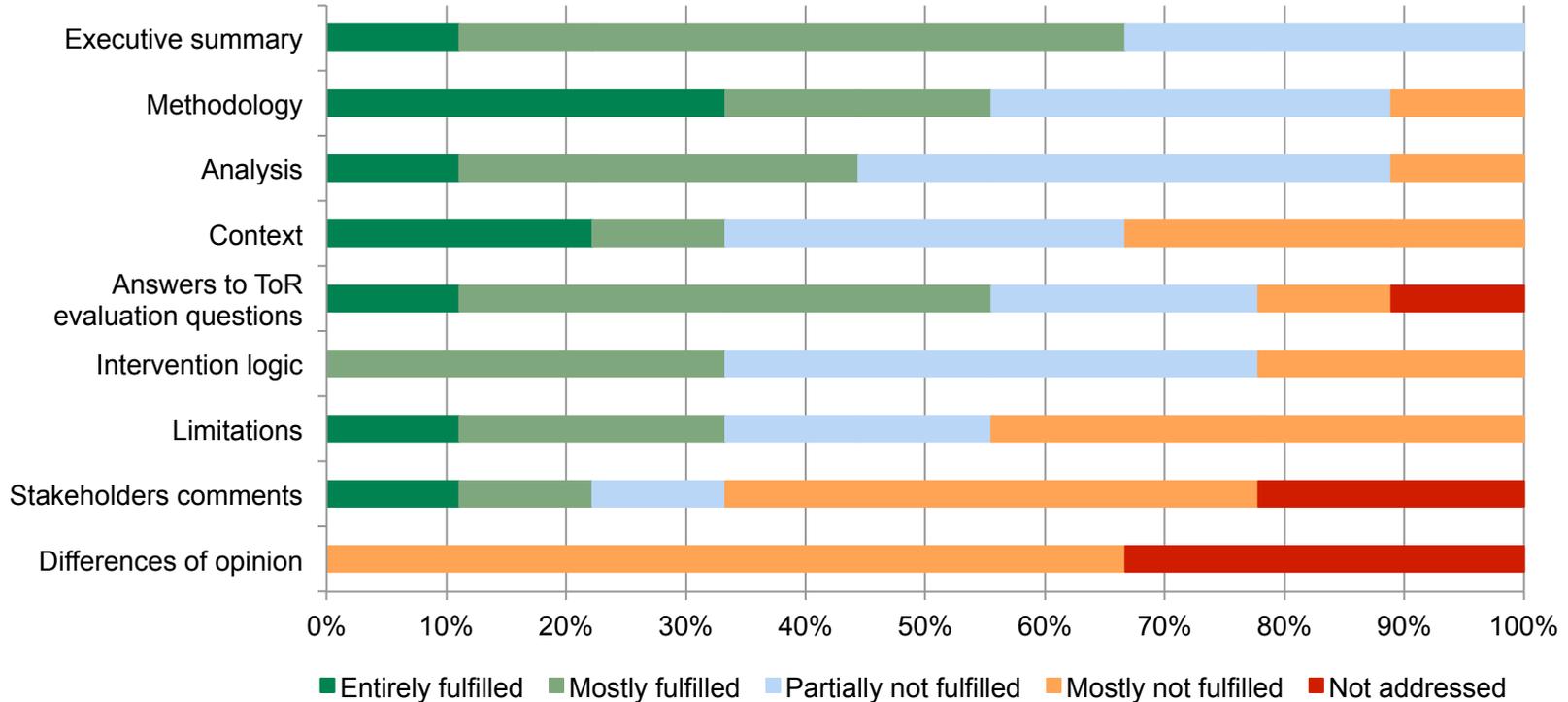
- The evaluation reports (and probably the evaluations as well) vary greatly in terms of their technical quality
- The meta-evaluation focuses mainly on formal compliance of the report – not on the quality, substance or reasoning. That should be included.
- More emphasis should be paid on sharing of good evaluation practices, with regard to:
 - How to intelligently use methods and sources of information
 - How to address difficulties (lack of data, etc)
 - How to (nicely and systematically) present evaluation results
 - What makes an excellent evaluation vs a technical analysis?

WORKSHOP DISCUSSION

- As a result of competition and for efficiency reasons, evaluations tend to be planned and conducted with minimum resources, both human and financial.
- Inadequate resourcing often impacts the quality of evaluations, too.
- When resources are scarce, it is even more important to prioritise evaluation questions.

SUMMARY OF META-EVALUATION RATINGS

Fulfillment of the criteria (N= 9)



4. Meta-analysis of innovation programmes

- methodology, findings and conclusions, lessons

META-ANALYSIS OF INNOVATION PROGRAMMES

- Evaluation reports with no major shortcomings have been included in the meta-analysis.
- Exception to this is ALICT evaluation, which has been left out due to weak methodology
- The qualitative analysis of what has worked, what not and why in the projects is conducted through an assessment of document review and evaluation criteria, which pays attention to:
 1. Relevance
 2. Impact
 3. Effectiveness
 4. Efficiency
 5. Sustainability
 6. Aid effectiveness
 7. Coherence

WORKSHOP DISCUSSION

- When looking into the results of development innovation programmes, it is important to understand that direct beneficiaries of this type of programmes are not typically the poorest of the poor.
- This may be challenging /controversial, because the aid effectiveness framework for development programmes requires such impacts.
- Lack of understanding on how innovation processes support social development can lead to biased analysis of relevance and impact.

Q1. RELEVANCE

Generic evaluation questions:

- Are the objectives and achievements of the programme consistent with the problems and priorities of the stakeholders, including all final beneficiaries?
- Whose poverty and inequality is the programme focused to reduce?
- Whose sustainable development is the programme promoting?
- Are the objectives and achievements of the programme consistent with the policies of the partner country?
- Are the objectives and achievements of the programme consistent with Finland's development policy?
- Are the objectives consistent with poverty reduction objectives?
- Are the commitments of the partner country's national policies and strategies, and of the international and regional conventions on promotion and enjoyment of human rights and gender equality, reduction of inequalities and promotion of climate sustainability integrated into programme design and implementation?

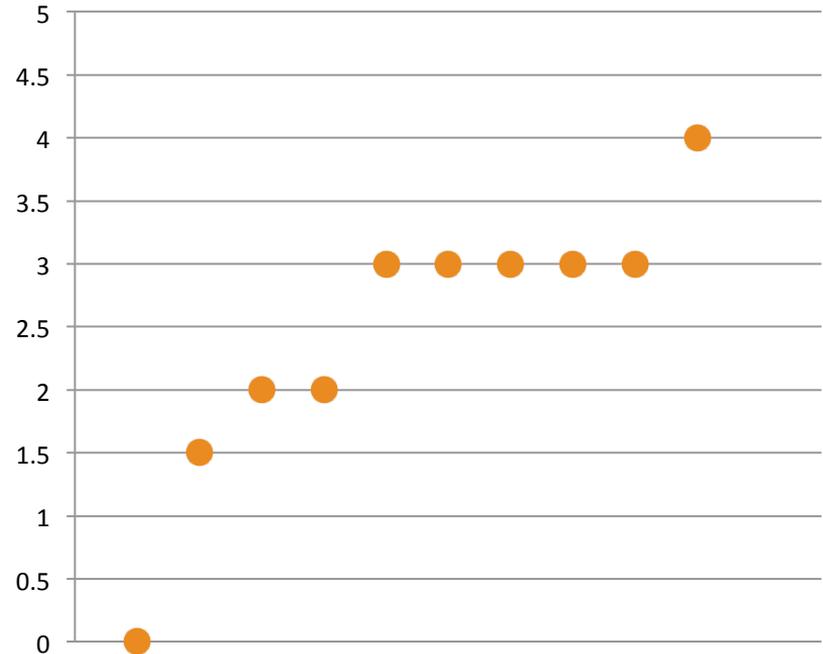
RELEVANCE

Relevance focuses on problems and policy priorities

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of relevance



FINDINGS & CONCLUSIONS

- All programmes are seen to be in line with policies and priorities of partner countries – at least at the high level.
- In practice this is not always the case. Examples:
 - Better and earlier needs assessment/baseline needed
 - “Not necessarily addressing the right issues and challenges”
 - “Operationally perhaps not fully in line”
 - “Gap between expressed priorities and the capacity to fulfill them”
- Many evaluations did not directly discuss relevance
- HRBA-issues and poverty reduction mostly only indirectly addressed

Q2. IMPACT

Generic evaluation questions:

- Has progress been made towards achieving the overall objective(s) of the programme?
- Did the programme reduce the poverty of all intended final beneficiaries?
- Did the programme impact on the lives of the poor women and men through prices, employment, transfers, access, authority, assets or empowerment?
- What are the overall impacts of the programme, intended and unintended, long term and short term, positive and negative?
- Are there real and long lasting positive changes in the lives of all intended beneficiaries in terms of human rights and gender equality, reduction of inequalities and promotion of climate sustainability?

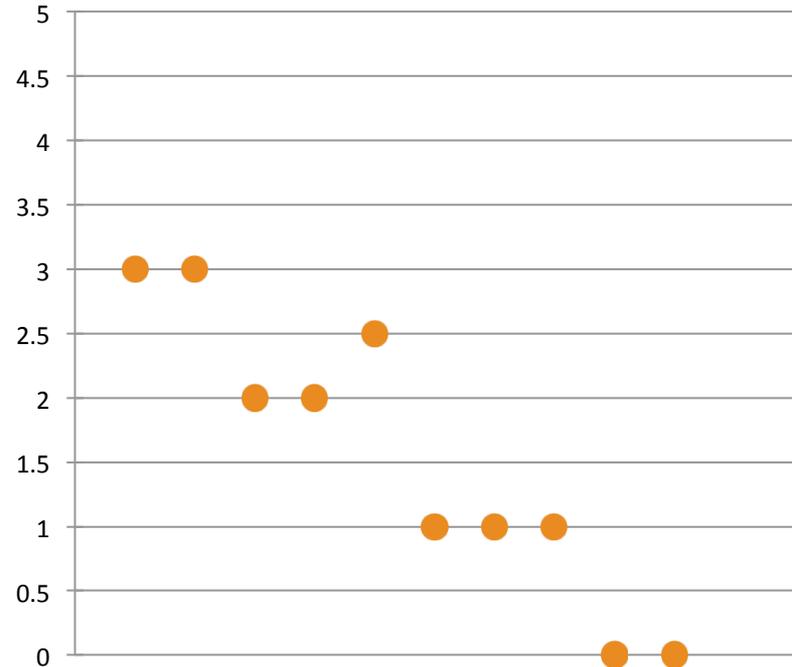
Q2. IMPACT

Impact focuses on achievement of wider objectives

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of impact



FINDINGS & CONCLUSIONS

- Large variance, problematic topic for implementation and for evaluation.
- Most of the mid-term reviews said it was too early to evaluate impact
- Some general challenges for impact:
 - Lack of human resources
 - Partner government priorities and commitment insufficient
 - Not enough local partnerships created
- Some evaluation reports concentrated on the outputs, analysis of impact was not there.

WORKSHOP DISCUSSION

- One should understand the limits of evaluating impact. Only after some years it is possible to assess if a programme has generated impact. Until that, evaluation can validate the anticipated impact logic and estimate whether the implemented actions have guided the development onto the right tracks that likely are to result in impacts.

Q3. EFFECTIVENESS

Generic evaluation questions:

- Has the intervention achieved its purpose or will it do so in the future?
- Are the results and the programme purpose making a contribution towards reducing poverty?
- To what extent have the objectives related to promotion, enjoyment and accountability for human rights been achieved during implementation of the programme?
- To what extent have gender equality, reduction of inequalities and promotion of climate sustainability been achieved during implementation of the programme?

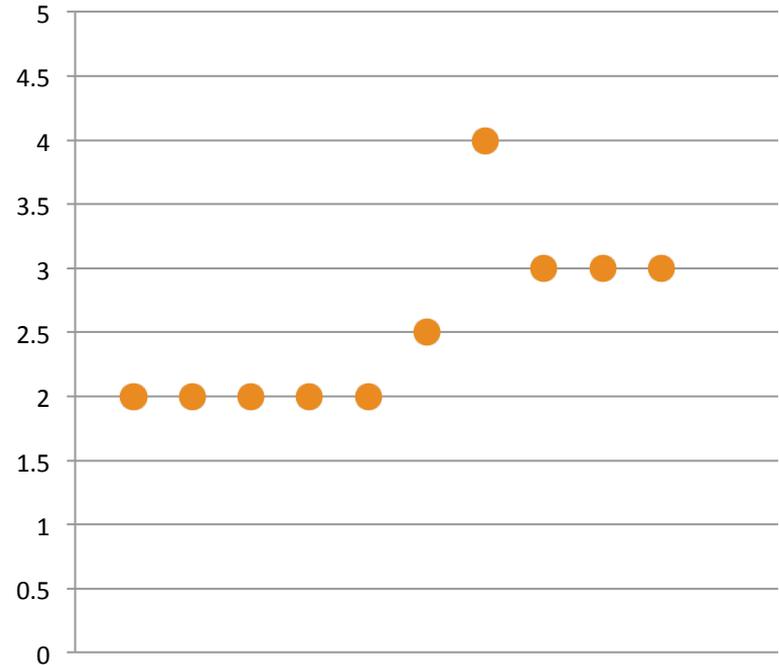
Q3. EFFECTIVENESS

Effectiveness focuses on evaluating the achievement of the programmes immediate objectives

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of effectiveness



FINDINGS & CONCLUSIONS

- Difficulties in evaluating effectiveness were common. As with impact, many MTR's stated it was too early.
- Lack of baselines, indicators, objectives and targets especially related to reducing poverty, reducing inequality, promoting gender equality, promoting human rights etc.
- Some accomplishments are difficult to attribute to programmes.
- Most evaluations do not discuss cross-cutting issues related to effectiveness

WORKSHOP DISCUSSION

- MFA's recommendation is that MTRs should not be commissioned too early. Instead, management reviews can and should be made in early phases to provide accurate feedback for the project management after the initial phase.
- More emphasis should be given to prior analysis of the context and monitoring of the project activities since their beginning.

Q4. EFFICIENCY

Generic evaluation questions:

- How well have the activities transformed the available resources into the intended outputs or results, in terms of quantity, quality and time?
- Can the costs of the intervention be justified by the achievements?
- What is the quality of the management of the programme, including e.g. work planning, monitoring and reporting, resource and personnel management, cooperation and communication?
- Have important assumptions been identified? Have risks been appropriately managed?
- Have resources been provided and efficiently used for participation of all stakeholders (rights holders), inclusiveness and other short-term process achievements?
- Have resources been provided and efficiently used for long-term investments in enabling environment, capacity development etc. for promotion and enjoyment of human rights by duty bearers and rights holders, for promotion of gender equality, reduction of inequalities and promotion of climate sustainability?

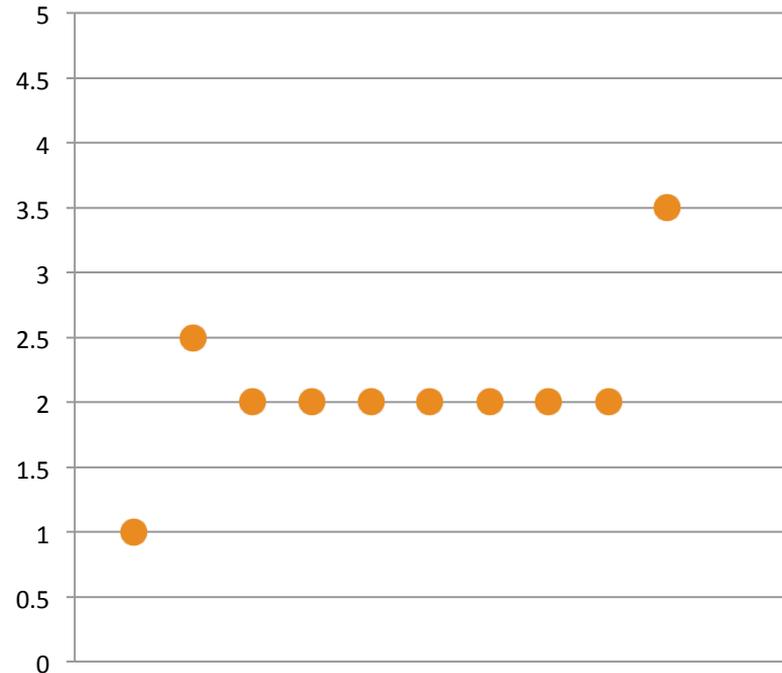
EFFICIENCY

Efficiency focuses on value for money, other available resources and sound management

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of efficiency



FINDINGS & CONCLUSIONS

- In MTR phase several programmes had had a slow start, which made management costs to seem unproportionally high. Expectation was for this to even out.
- In many cases inefficiencies were beyond the influence of the program (mostly partner government process –related)
- Financial management procedures, funding disbursement, guidelines and general programme management set up were weak in some cases, and took some time to set up.
- Some variation but in general efficiency is not the biggest challenge

Q5. SUSTAINABILITY

Generic evaluation questions:

- Will the benefits produced by the programme be maintained after the termination of external support?
- What are the possible factors that enhance or inhibit sustainability, including ownership/commitment, economic/financial, institutional, technical, socio-cultural and environmental sustainability aspects?
- Has the phasing out of external support been planned, and will the plan ensure sustainability?
- What is the likelihood that the achievements in human rights and gender equality, reduction of inequalities and promotion of climate sustainability are sustained after the programme is completed?

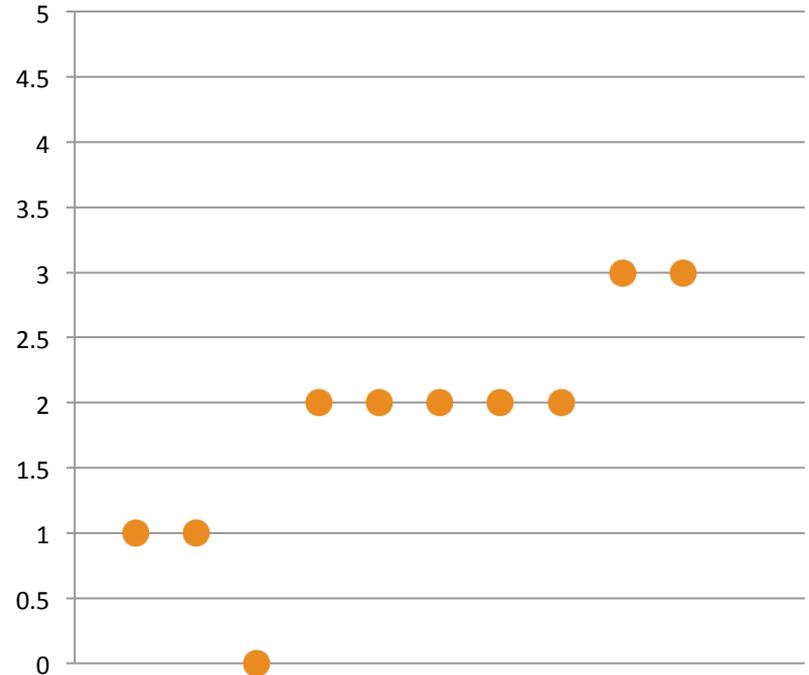
SUSTAINABILITY

Sustainability focuses on evaluating the likely continuations of achievements

Rating scale:

- 1 = criteria mostly not fulfilled or totally absent
- 2 = criteria partially fulfilled
- 3 = criteria mostly fulfilled
- 4 = criteria entirely fulfilled
- 5 = criteria entirely fulfilled in a clear and original way
- n/a = not addressed (shown as 0)

Ratings of sustainability



FINDINGS & CONCLUSIONS

- Sustainability seems to be challenging to assess.
- Again, MTR's found it too early to assess.
- The issue of sustainability is discussed in the reports, but in most cases not analysed with any depth.
- Pilots and projects, as well as human capital development, are the areas where long term improvements are seen
- In some cases, one programme cycle was seen as a too short time for real sustainability to be built.

WORKSHOP DISCUSSION

- Similar to the impact, it is difficult to evaluate actual sustainability of the programme during its early phase.
- The paradigm of having to have measurable outcomes and results for projects' reporting purposes leads to outcomes that have a great impact in a short-term but don't have good long-term sustainability. Working with local governments, on the other hand, typically result poorly measurable impacts, but can generate sustainable developments in a longer-term.
- Well carried background analysis, gained contextual understanding and a change theory linked to national and local development agendas supports the sustainability of innovation projects as well.

Q6. AID EFFECTIVENESS

Aid effectiveness focuses on evaluating the implementation of Paris Declaration principles.

Generic evaluation questions:

- Has the programme promoted ownership, alignment, harmonisation, management for development results and mutual accountability
- Has the programme promoted coordination and complementarity?
- Has the implementation of Paris Declaration principles contributed to the achievement of the cross-cutting objectives?

Rating scale:

1 = criteria mostly not fulfilled or totally absent

2 = criteria partially fulfilled

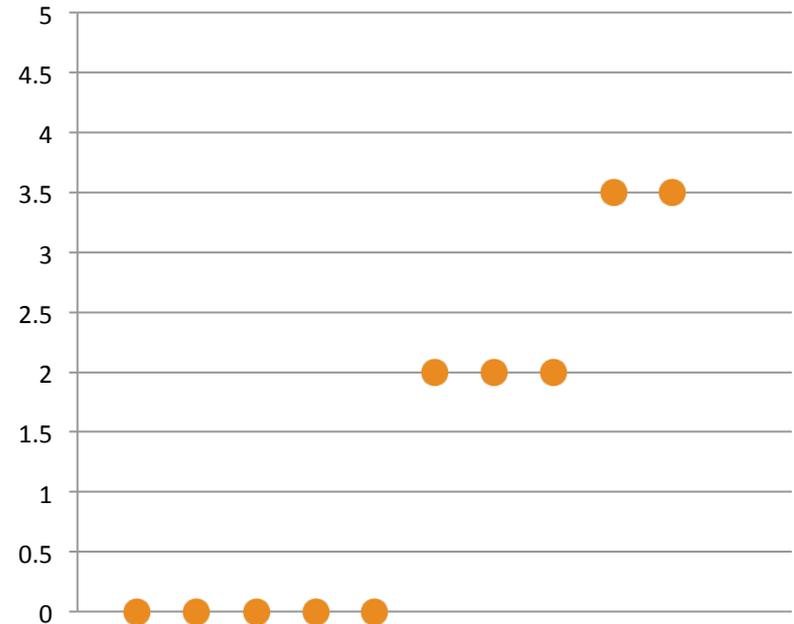
3 = criteria mostly fulfilled

4 = criteria entirely fulfilled

5 = criteria entirely fulfilled in a clear and original way

n/a = not addressed (shown as 0)

Ratings of aid effectiveness



FINDINGS & CONCLUSIONS

- Most evaluation reports did not address aid effectiveness.
- In some cases there were some mentions of coordination, ownership, alignment, etc.

Q7. COHERENCE

Coherence focuses on evaluating issues beyond development cooperation.

Generic evaluation questions:

- Have contradictions with other policies prevented the implementation and achievement of the development objectives, or are they mutually reinforcing?
- Are other policies consistent with the human rights based approach and cross-cutting objectives and their integration into the programme?

Rating scale:

1 = criteria mostly not fulfilled or totally absent

2 = criteria partially fulfilled

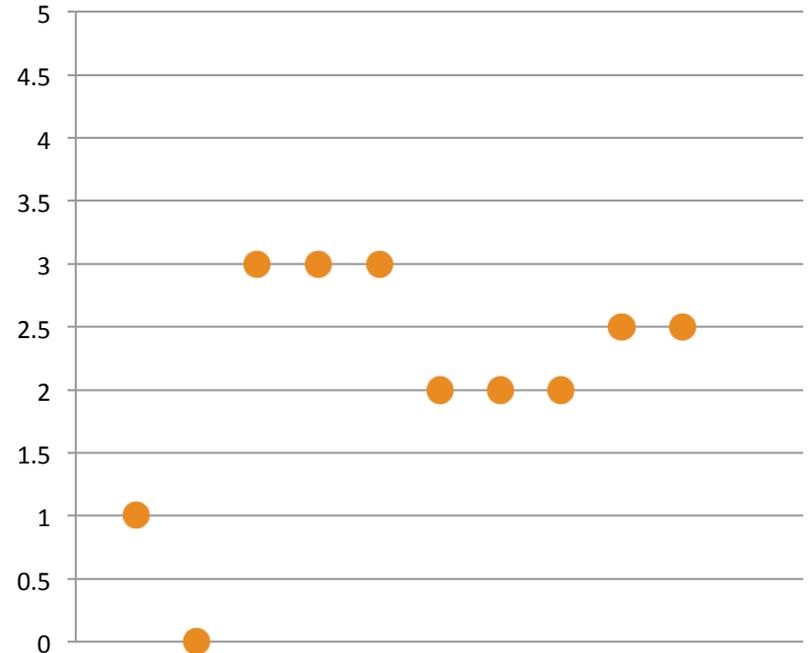
3 = criteria mostly fulfilled

4 = criteria entirely fulfilled

5 = criteria entirely fulfilled in a clear and original way

n/a = not addressed (shown as 0)

Ratings of coherence

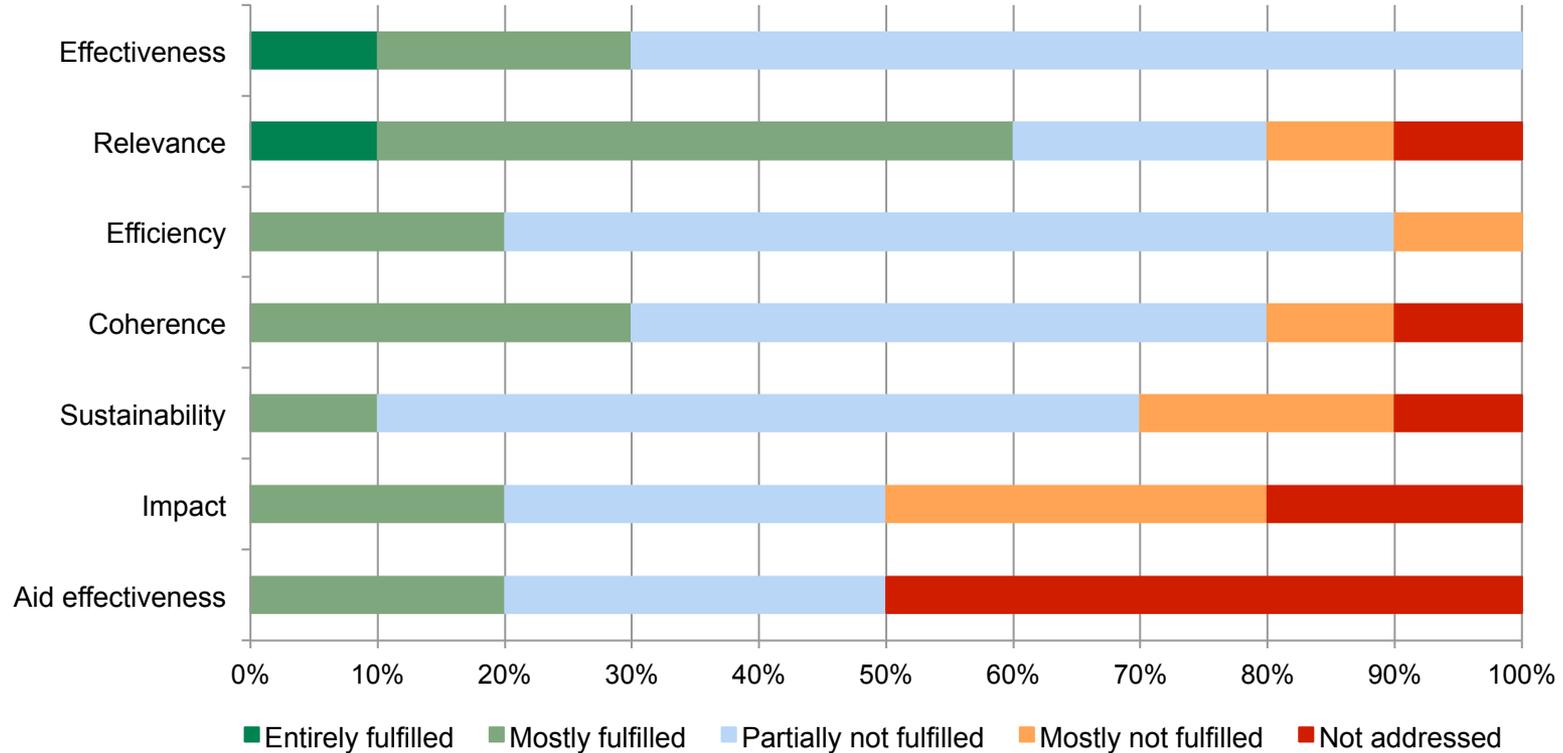


FINDINGS & CONCLUSIONS

- As with relevance, many projects were seen to be coherent on the high level and on paper but this is not necessarily apparent in the activities.
- Most reports did not directly address coherence, but policy complementarity or contradiction was mentioned.
- Discussion related to coherence was mainly general and non-specific.

SUMMARY OF META-ANALYSIS RATINGS

Fulfillment of the criteria (N=10)



5. Summa summarum

OVERALL CONCLUSIONS

- Meta-evaluation and meta-analysis provide efficient ways to compare and assess the overall situation and trends in programming
- They are particularly useful for 'checking the overall status' of evaluation practices
- The MFA meta-evaluation tool could be refined further. It currently ensures that the evaluation includes enough information for the procurer to be able to check the quality, but it does not by itself investigate the quality of the evaluation
 - For example, it ensures the methodology is explained and justified, but it does not ask whether the chosen methodology is relevant for the evaluation

META-EVALUATION

1. The evaluation quality of MFA innovation projects, according to OECD / DAC standards, varies.
2. Innovation programmes are broad, systemic, experimental and anticipate impact over a long term. They are often unique in their design and context as well. Straight-forward evaluation approaches may have limited capability to address the full nature of such programmes.
3. Regardless of the above, there are (too) many technical shortcomings in the programme evaluations. Particularly:
 - a. Description of programme context and its changes
 - b. Description of overlying assumptions
 - c. Explaining the choice of methods
 - d. Documentation and analysis of qualitative data (e.g. stakeholder views)
 - e. Responding to ToR questions

META-ANALYSIS

1. **Relevance:** all programmes seem to be in line with partner country policies, but in practice there are gaps. Many evaluations do not discuss relevance.
2. **Impact:** Evaluation of impact challenging in the current format. Most cases report it is too early to assess impact. The criteria could be changed to evaluate how well the programme has been planning for impact.
3. **Effectiveness:** difficulties in evaluating this. Lack of baselines, targets, etc. Accomplishments difficult to attribute to programmes.
4. **Efficiency:** Many programmes had slow start. Inefficiencies often beyond the control of the programme.
5. **Sustainability:** As with impact, generally too early. Short-term indicators and long-term sustainability do not always correlate.
6. **Aid effectiveness:** Most evaluations do not report directly on aid effectiveness
7. **Coherence:** High in programme documents, not always visible in practical activities