

META-EVALUATION

Presentation of the Meta-Evaluation of MFA's Project and Programme Evaluations in 2017-2020

Dr. Susanne Johanna Väh
Dr. Stefan Silvestrini
Dr. Hansjörg Gaus
Petra Mikkolainen
Dr. Maja Flaig
Janis Wicke

March 31



The Assignment

Purpose

- Learn from decentralised evaluations
- Exploit findings from a comparison with previous meta-evaluation results
- Generate aggregated results for Development Policy Results Report 2022

Scope

- Decentralised evaluations by regional or thematic MFA units, embassies, or by multilateral partners between 2017 and 2020
- Perspectives of MFA commissioners, Finnish embassies, & project implementers on these evaluations

Goals

- Enhancing quality of evaluations, evaluation management practices, evaluation capacity development
- Increasing the usefulness of decentralised evaluations
- Supporting improvements of development cooperation

Methodological Approach

Replication of 2017 meta-evaluation



Component 1:

80 Methodological quality assessments

- Standardised assessment
- Aggregation to aspects and overall report quality
- Statistical analysis

~~~~~ **MINIMUM REPORT QUALITY THRESHOLD** ~~~~~



### Component 2:

Content assessment of 72 evaluation reports

- Transferring evaluators' assessments on OECD DAC criteria and CCOs
- Standardised aggregation and statistical analysis
- Qualitative synthesis of evaluators' lessons and recommendations

## New features of recent meta-evaluation



### Component 3:

Use and usefulness assessment

- 119 survey responses and 26 interviews (MFA staff and intervention implementers)
- Descriptive analysis, qualitative content analysis, triangulation



### Component 1:

Dynamic methodological quality assessment

- Acknowledging increased evaluation standards in an additional report quality score



### Comparison:

(recent vs. previous meta-evaluation)

- Comparison of replicated quality scores.
- Comparison of dynamic quality score with replicated quality score

# R1

*On quality, use and usefulness of evaluation reports*

# R2

*To improve quality of Finnish development cooperation*

# R3

*For future meta-evaluations*

**Some limitations, but coping strategies**

# Limitations and Coping Strategies

Unclear representativeness of Finnish bi-, multi- and multi-bilateral interventions

→ Partially mitigated by simplifications and equal weighting of small and large interventions, but at a same time limitation

Risk of subjective assessment

→ Ratings as yes/no or on a four-step scale, standardised aggregations, strictly following tools and protocols

→ Extensive pre-test with calibration workshop, cross-analysis showed robustness, joint analysis workshops

→ Strictly limited results interpretation to an aggregated level

No consultation of project documents, limited possibility to triangulate evaluators' perspectives (components 1 & 2)

→ Indirectly possible, if results validation mentioned in reports

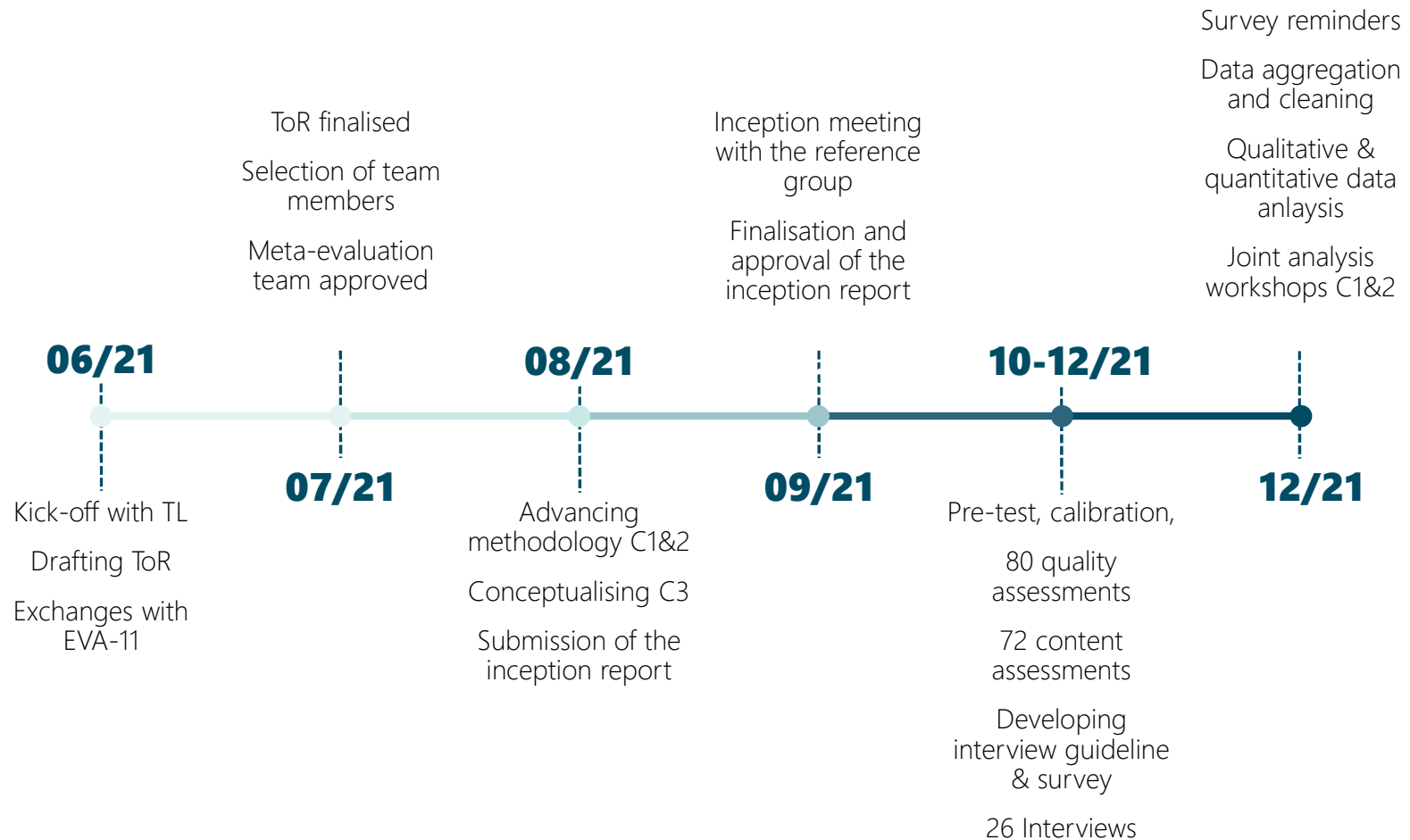
Self-selection into survey fosters upwards biased results of usefulness assessment (component 3)

→ Within case triangulation if multiple perspectives

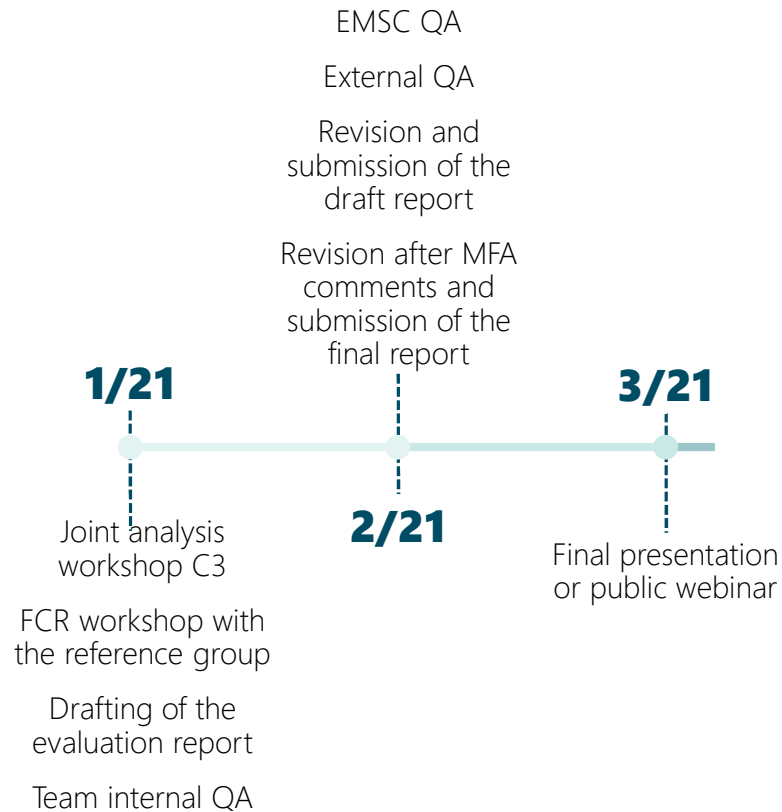
→ Estimation of potential biases and provision of lower- and upper-boundary estimates

→ Contextualisation with less biased qualitative interview data

# Timeline



# Timeline and Challenges



## Extreme Resource Constraints (time & man power)

- MFA Evaluation plan not up to date, identification of sample cumbersome, serious delays from the beginning
- Tracing of HQ and embassies staff for C3 caused several circles and dead ends, more than 300 contact attempts
- Summer break, different holiday periods, illness and pandemic-related effects not factored into the schedule

- High flexibility of the evaluation team
- New team member to comply with timeline

## Lessons learnt

- Enhance calendar time and man power to ensure healthy working conditions
- Alternatively: closing earlier the door for identification and data collection to save sufficient resources for analytical tasks and reporting



# Component 1 - Findings

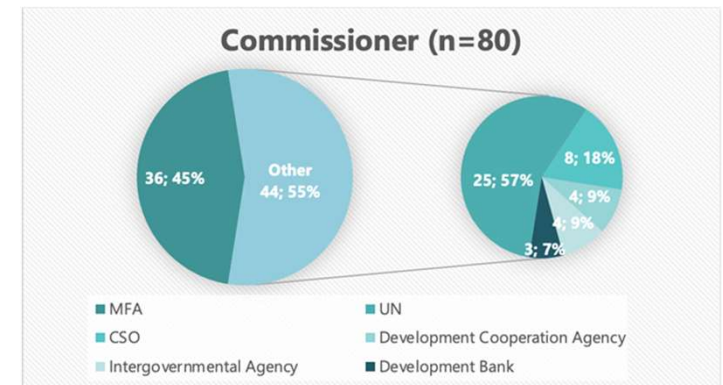
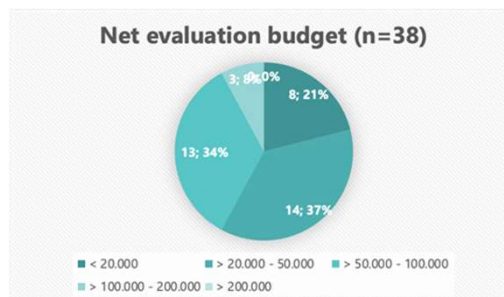
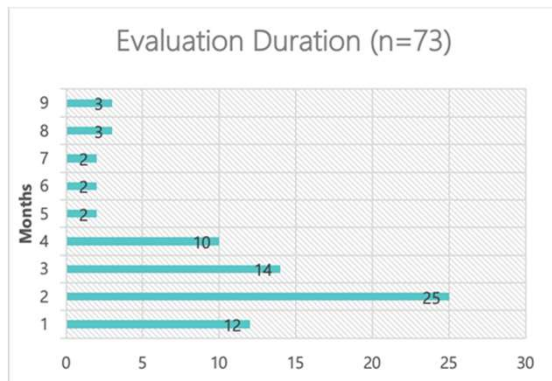
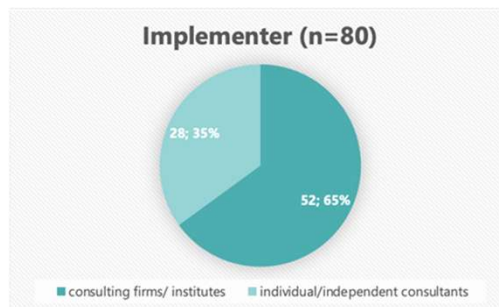
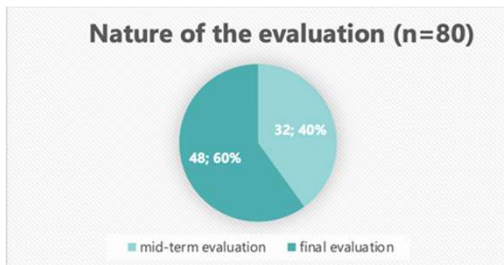
Quality assessment of 80 evaluation reports and 70 ToRs



# MFA's Decentralised Evaluation Portfolio (EQ1)

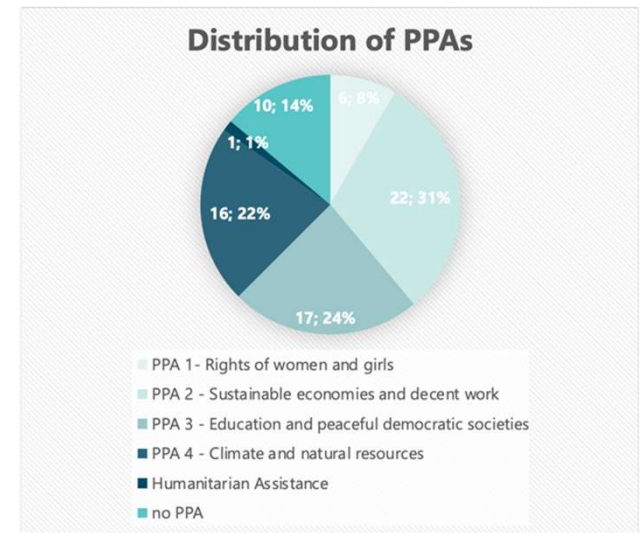
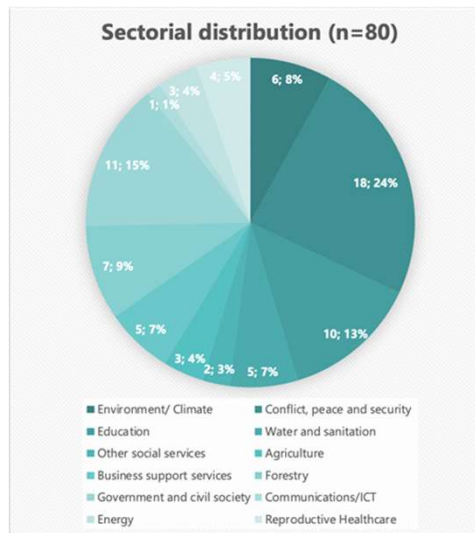
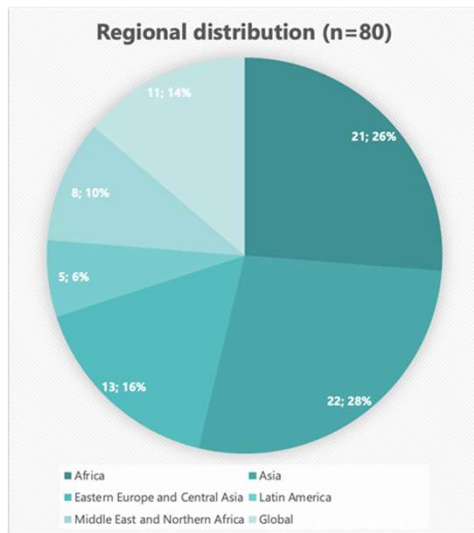
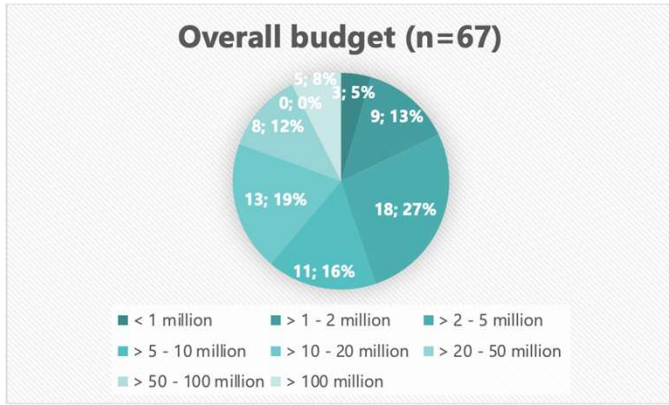
1. How can MFA's decentralised evaluation portfolio be described?

- 80 reports, comparable to earlier meta-evaluation but new fractions of atypical evaluations (funds, institutions)





- Variance regarding intervention characteristics poses limitations to comparability of intervention quality

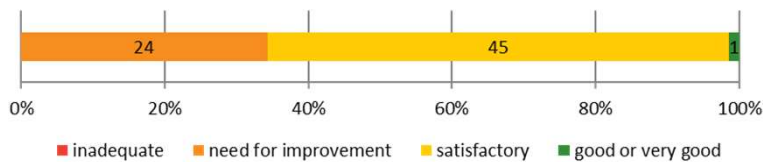


# Quality of ToRs (EQ3)



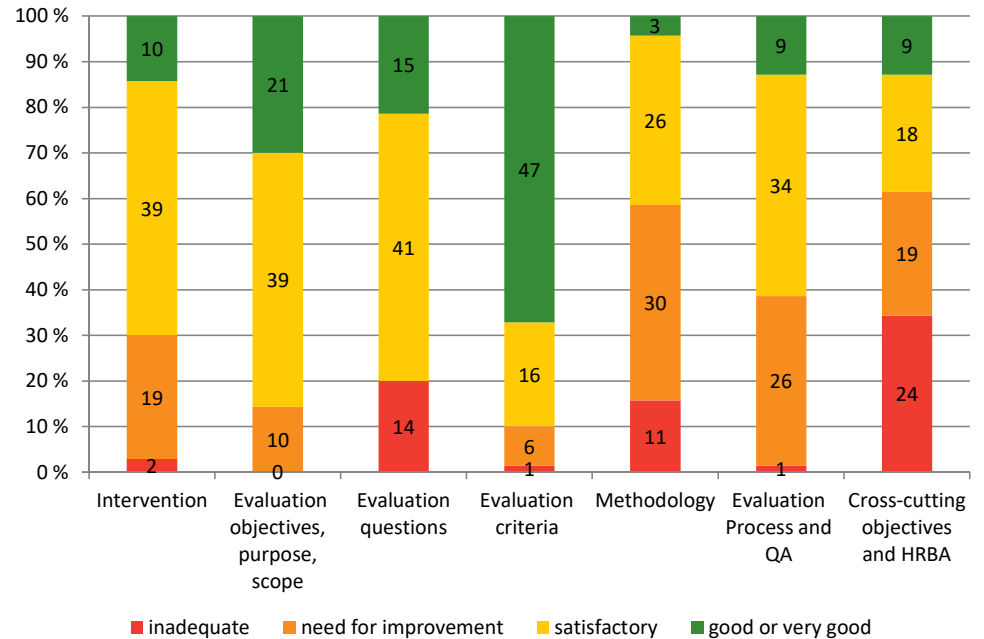
3. How is the quality of the ToRs of decentralised evaluations? Are there systematic linkages?

**Overall Quality of ToR (n = 70)**



- Largest room for improvement leave
  - Methodological guidance
  - Consideration of cross-cutting objectives
- Evaluation criteria more frequently addressed than in the previous meta, no differences on other aspects

**Quality of ToR (n=70)**



# Linkages: Quality of ToRs and Reports (EQ3)



- Quality of ToR is a driver of quality of evidence.
- A one-unit increase of ToR quality yields nearly to a one-unit increase of quality of evidence.
- Statistically significant effect holds when controlling for other variables in ordinary least squares regression (OLS).
- Further, no statistically significant differences among sub-groups (i.e. MFA vs. non-MFA commissioner, individual vs associated evaluators)

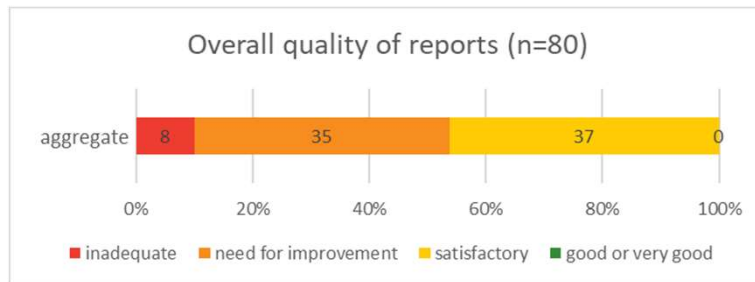
|                             | OLS                 |
|-----------------------------|---------------------|
| (Intercept)                 | -0.277<br>(1.310)   |
| ToR Quality                 | 0.942 **<br>(0.290) |
| Evaluation Budget (log)     | 0.086<br>(0.099)    |
| Evaluation Duration (monts) | 0.043<br>(0.031)    |
| Independent Consultant      | -0.063<br>(0.195)   |
| non-MFA commissioned        | -0.306<br>(0.180)   |
| Geographical scope          | -0.146<br>(0.084)   |
| N                           | 33                  |
| R2                          | 0.444               |

Standard errors are heteroskedasticity robust. \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05; model fit: F(6,26)=3.450, p=0.012\*\*, adjusted R2=0.316.

# Quality of Decentralised Evaluation Reports (EQ2)



2. How is the quality and reliability of decentralised evaluation reports? Are there differences between sub-groups?



- Majority of inadequate reports are atypical evaluations (funds, partnerships, institutions)
- Quality of findings, methodologies, consideration of CCOs and summaries are mostly in need for improvement.

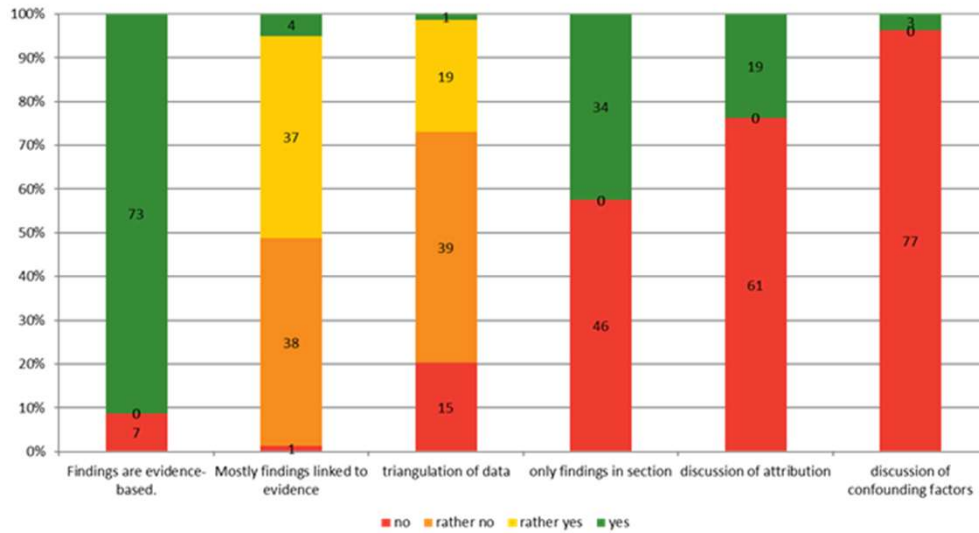




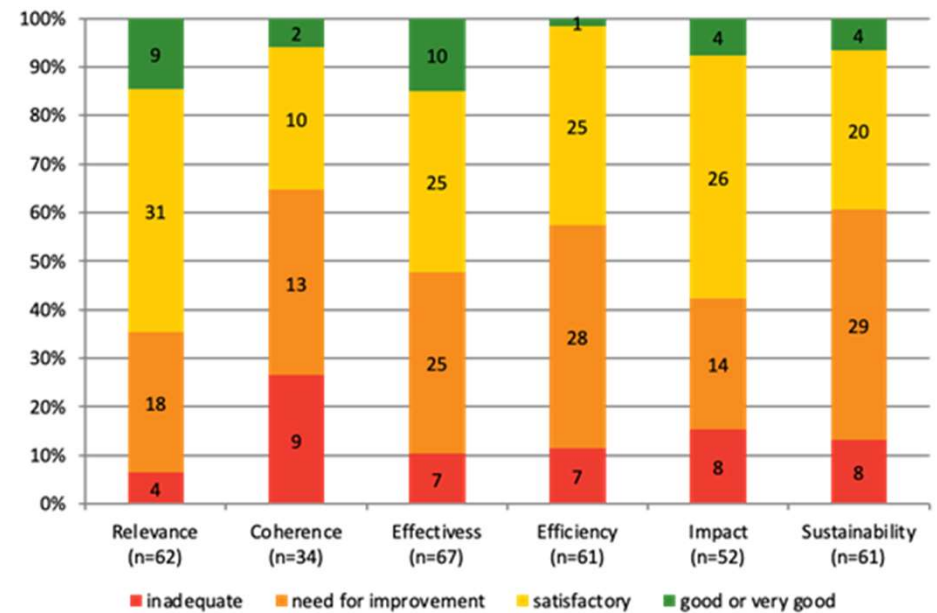
# Quality of Reports (EQ2) – Zooming in



Quality of findings (n=80)



Appropriate capturing of DAC criteria



- Nearly all reports are to some extent evidence-based, about half rather link findings to evidence.
- Coherence, efficiency, and sustainability are less appropriately captured than other criteria.

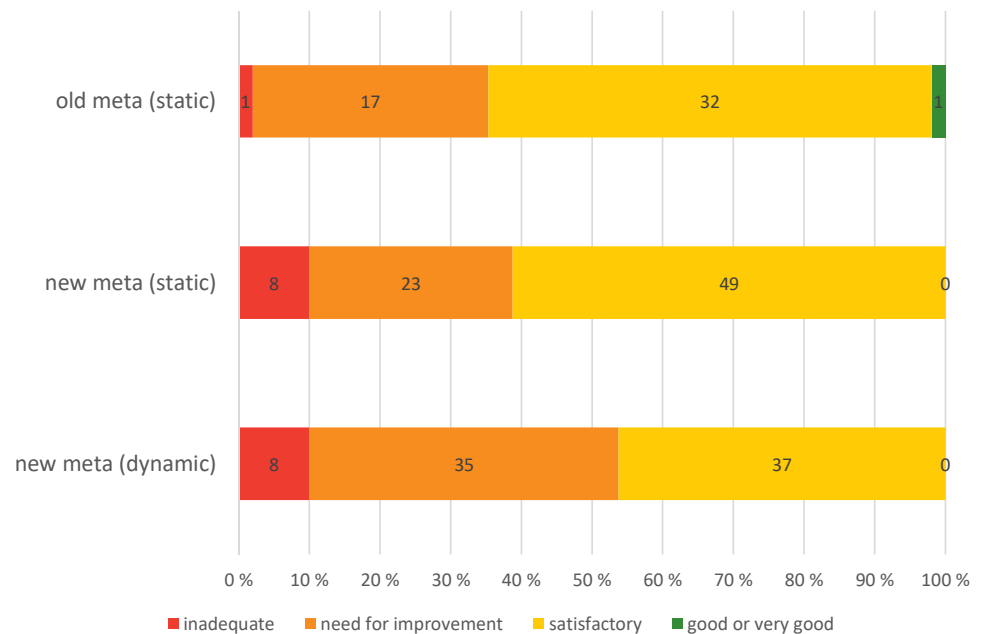
# Differences & Trends to Previous Meta (EQ4)



4. What are key differences and possible trends compared to the previous meta-evaluation? Are there gaps regarding MFA's evaluation capacity?

- The static comparison (i.e., replication) reveals no differences.
- The dynamic comparison (i.e., anticipating increased evaluation standards) suggests deterioration.
- Likely evaluators have not yet fully adapted.
- Deterioration may be also caused by inclusion of conceptually different evaluations.

Quality for different meta and aggregates



# Gaps Regarding MFA's Evaluation Capacity (EQ4)



## Summary of findings

- About one third of the ToR is in need of improvement.
- About half of the reports' quality is in need of improvement, few even inadequate.
- Half of the reports do not consider gender adequately, 80% lack non-discrimination, climate sustainability and HRBA.
- Deterioration of report quality when anticipating increased quality standards.

## Gaps derived

- Evaluation capacity inside and outside MFA does not yet tap its full potential.
- Evaluation capacity at evaluators' level is not yet sufficient and MFA guidance is not yet generally applied.
- Evaluation capacity inside MFA is partly limited and internal guidance is not yet fully followed.
- Acceptance of weak reports discloses that enforcement of external guidelines are not yet assured.

# Component 2 - Findings

Content assessment of 72 evaluation reports

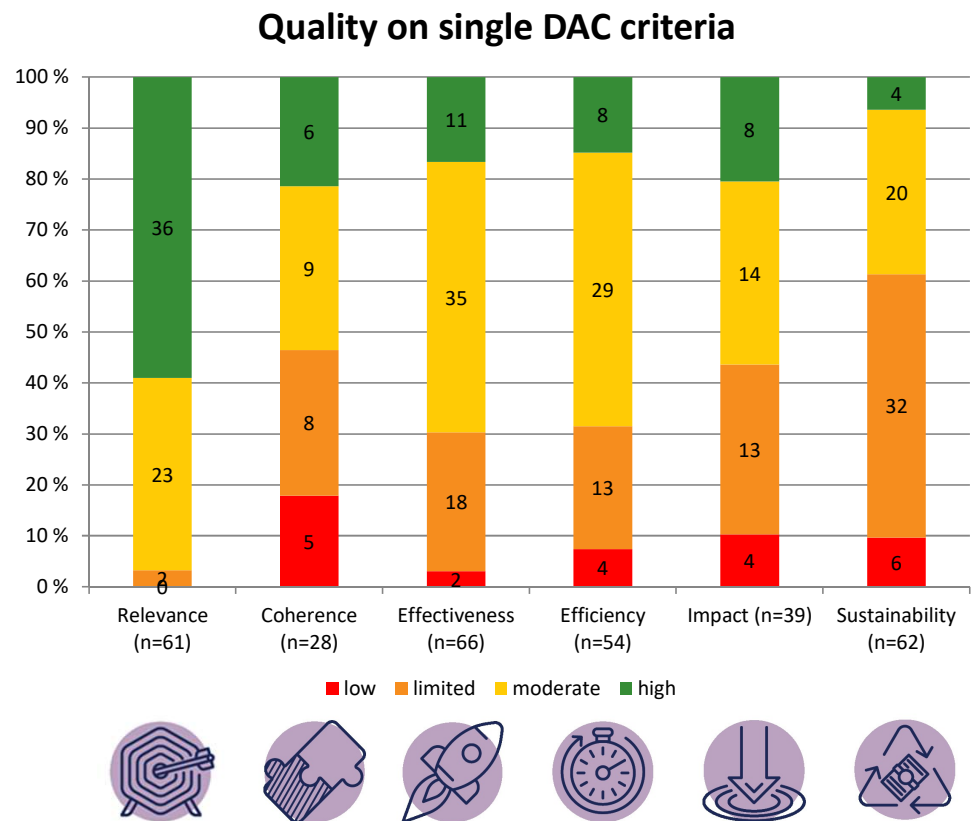
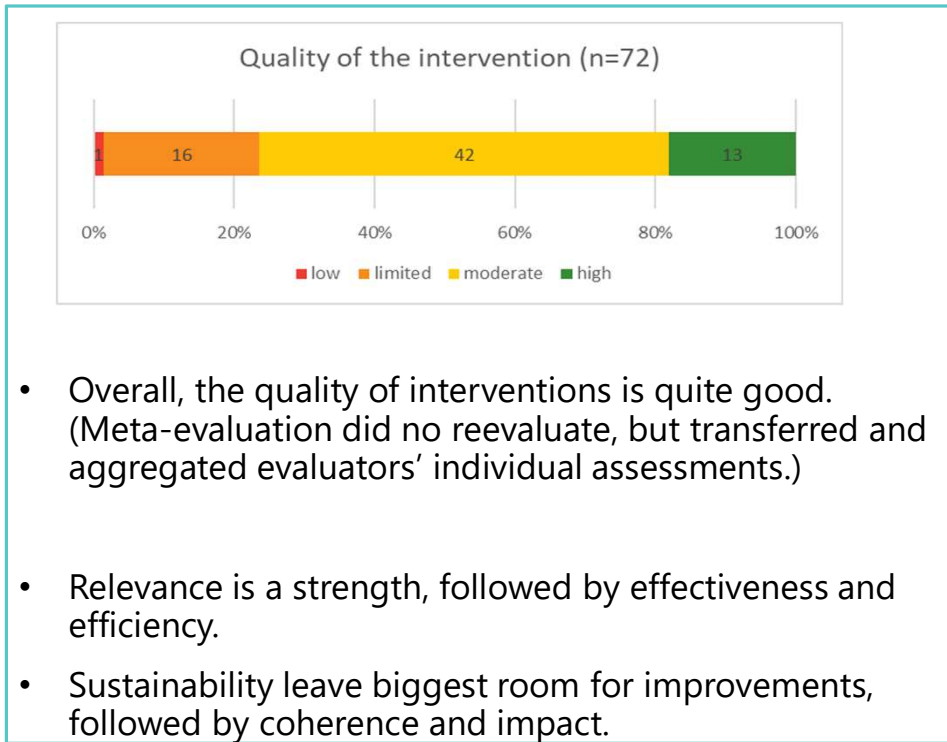




# Quality on DAC criteria (EQ5)



- 5. What about the quality of Finnish development cooperation according to the OECD DAC criteria based on reliable reports?
- 8. What about overall quality, major strengths and challenges of Finnish development cooperation based on reliable reports?





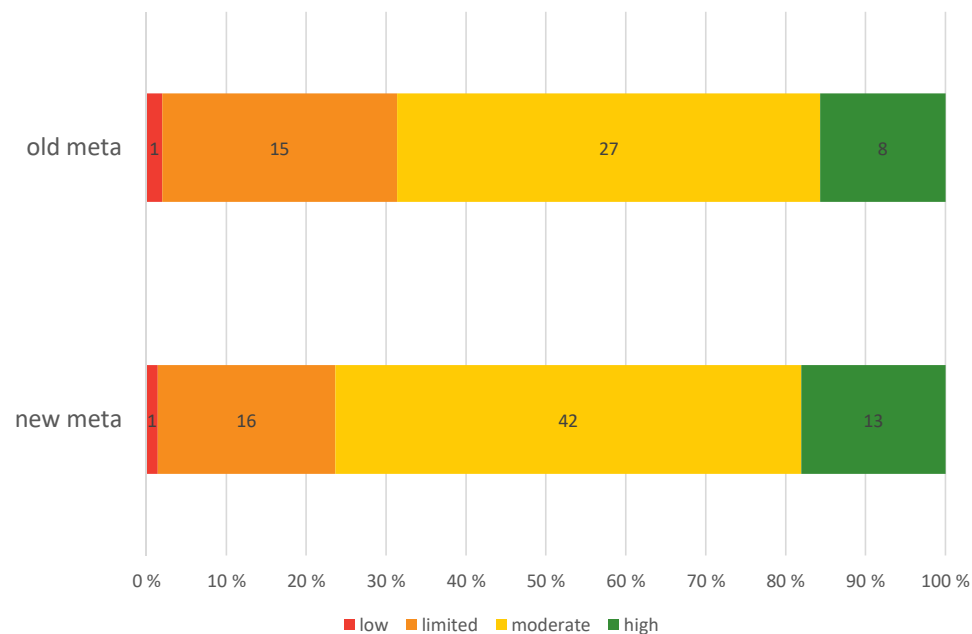
# Differences & Trends to Previous Meta (EQ9)



9. What are key differences and possible trends compared to the previous meta-evaluation?

- The overall quality of the interventions did not change at large from the previous meta-evaluation.
- However, direct comparisons at the intervention level are not possible as
  - both meta-evaluations looked at different interventions
  - the sample composition varied with respect to intervention budgets and sectoral distribution.

**Possible trends on overall intervention quality**



# Evaluators' Recommendations and Lessons (EQ7)



7. What are major recommendations and lessons learnt by evaluators to improve Finnish development cooperation?

## Recommendations

- Three evaluation reports lack recommendations; the remainder provides 1,123 recommendations.
- About two-thirds of the reports contain recommendations on M&E, coherence and planning.
- About half of the reports provide recommendations on sustainability, management, financial aspects, and capacity development.

## Lessons learnt

- Only 53% of all evaluation reports contain lessons learnt. In total, 171 lessons learnt were identified.
- Lessons learnt are spread over a wide range of themes. Hence, no typical lessons could be identified.
- Capacity, planning, and coherence are the most frequent ones, appearing in a bit less than 20% of the reports.



# Component 3 - Findings

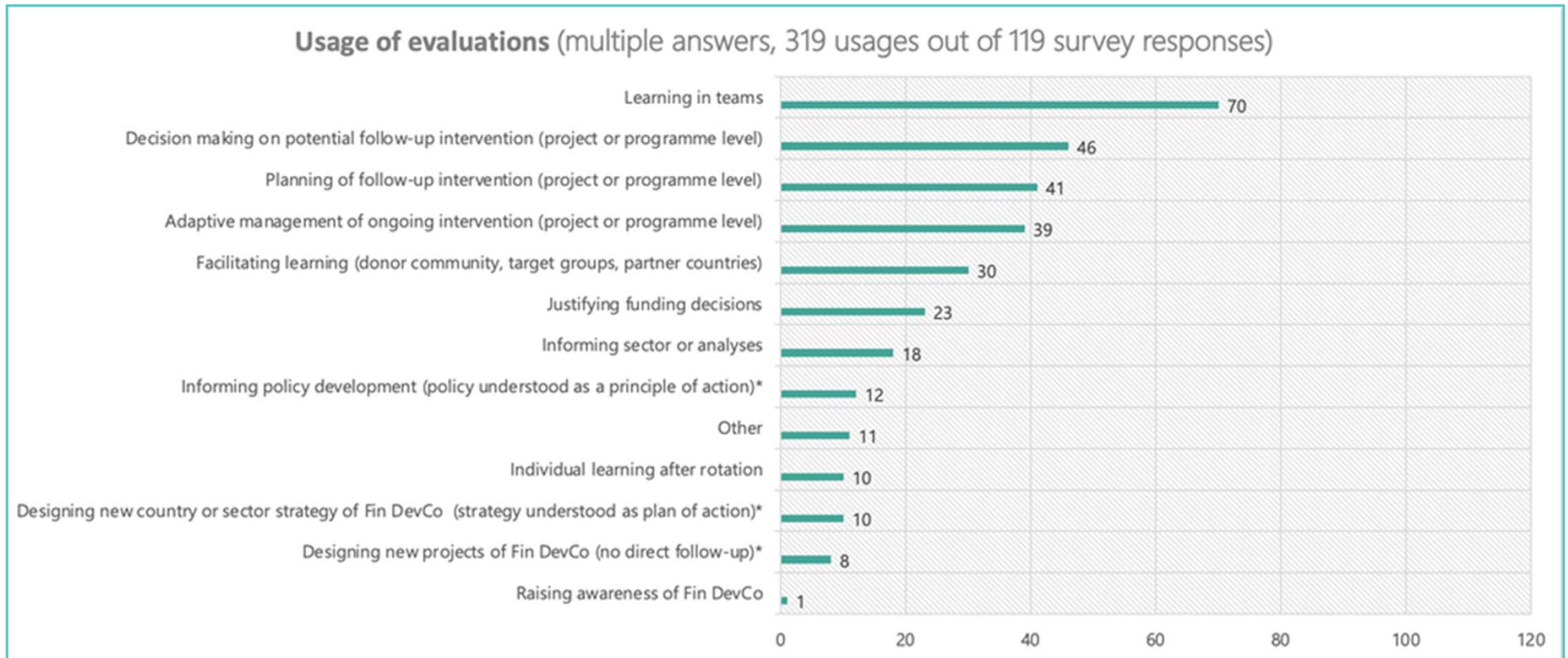
Use and usefulness assessment of evaluations



# Use of Decentralised Evaluations (EQ10)



10. How are decentralised evaluation reports used by key stakeholders?



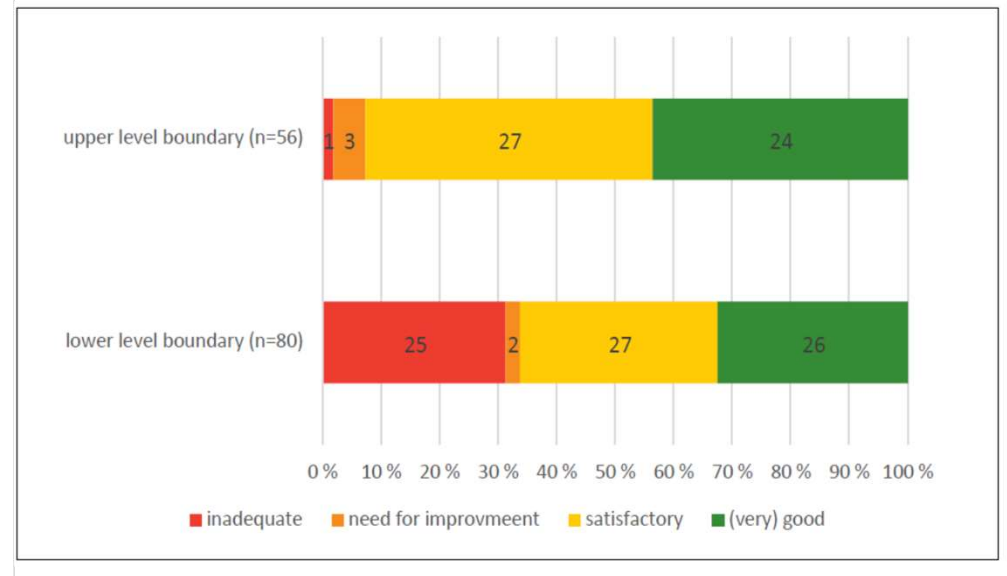
# Overall Usefulness of Evaluations (EQ11)



11. How useful are decentralised evaluations (i.e., reports and processes)?

- Overall, usefulness is quite positively assessed.
- No major upward bias of potentially overoptimistic implementers' assessments of the evaluations could be observed.
- A lower-level boundary assuming inadequateness for all cases for which survey data is missing (i.e., about a third) comes still to a quite positive overall usefulness.
- Sub-aspects like timing, timeliness, relevance of recommendations are similarly positive assessed.

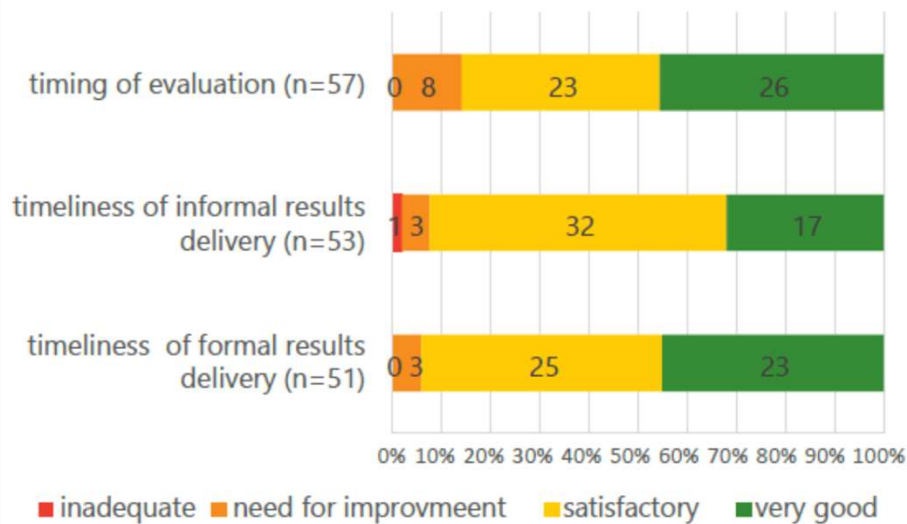
## Overall usefulness of evaluations (case level, ranges)



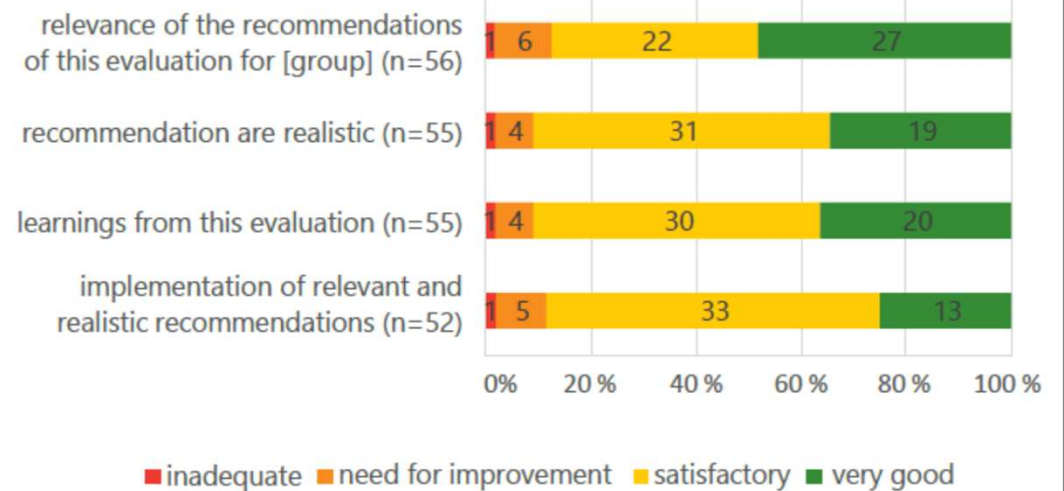
# Usefulness of Timing and Recommendations



## Timing and timeliness



## Recommendations





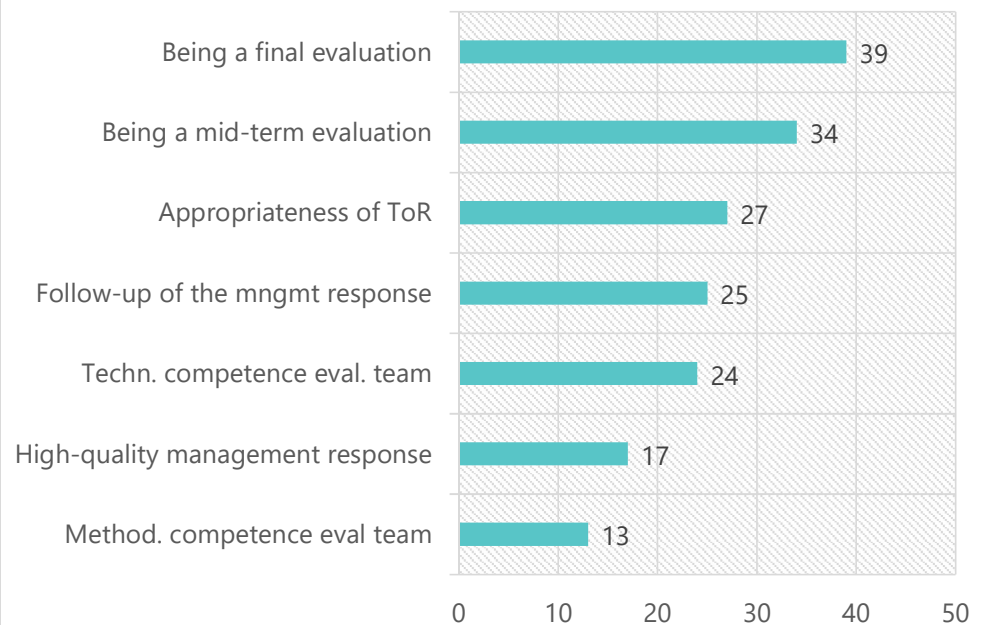
# Facilitating Factors for Usefulness (EQ12)



10. Are there typical factors facilitating or hampering the usefulness of decentralised evaluations? What are major recommendations and lessons learnt from stakeholders to improve the use and usefulness?

- Most of the main **facilitating factors** are under MFA's sphere of influence. Typical factors hampering the usefulness could not be identified.
- Facilitating factors were twice as likely to be identified than hampering factors (252 in 85 responses vs 116 in 64).
- **Interviewees' recommendations** comprise mainly organisational aspects, timing and timeliness of the evaluations, and aspects of report quality.
- Support to enhance clarity on scope and focus of evaluations, efficiency of learning and results dissemination at institutional level is desired by MFA staff

## Main facilitating factors





# Conclusions and Recommendations

Methodological quality, intervention quality, and usefulness of evaluations





## CONCLUSIONS

**COMPONENT 2**  
Summative analysis of the quality of interventions (EQ5, EQ6, EQ7, EQ8 & EQ9)

C4: OECD DAC criteria sustainability and coherence leave room for improvement.  
C5: Finland's Development Policy on CCOs is not fully reflected in its interventions.  
C6: Evaluators' recommendations points to continuing areas for improvement for MFA.

**COMPONENT 1**  
Methodological QA of evaluation reports and ToRs (EQ2, EQ3 & EQ4)

C1: Enforcement gap inside MFA result in weak evaluation reports.  
C2: Evaluation capacity inside MFA is still constrained.  
C3: MFA is at risk to fall behind attained quality level

**COMPONENT 3**  
Use and usefulness assessment of the evaluations (EQ10, EQ11 & EQ12)

C7: Evaluation reports are often used, but room to enhance usage remains.  
C8: Usefulness of evaluation reports is often limited to the individual level.  
C9: Exploiting MFA's sphere of influence would enhance the usefulness of its evaluations.

**LIMITATIONS**

## RECOMMENDATIONS

**To improve quality of Finnish development cooperation**

R2.1 Pay stronger attention to sustainability, coherence, and impact, as well as to CCOs and HRBA.  
R2.2 Translate lessons learnt and recommendations provided by evaluators into action.

**On quality, use and usefulness of evaluation reports**

R1.1 Enforce compliance with guidelines inside MFA for better ToRs.  
R1.2 Establish a quality assurance process inside MFA to enforce evaluators' compliance with manual, guidelines, and ToR outside MFA.  
R1.3 Improve knowledge management inside MFA to foster organisational learning and enhance usefulness of evaluations.  
R1.4 Continue further evaluation capacity development for all stakeholders inside and outside MFA on focused topics.  
R1.5 Use facilitating factors identified in this meta-evaluation like appropriate ToRs, follow-up of management response as an entry point to enhance use and usefulness of evaluations.

**For future meta-evaluations**

R3.1 Set-up an inventory of interventions and evaluations to assess sample representativeness and save resources for future meta-evaluations.  
R3.2 Introduce digitised feedback sheets on evaluations for MFA staff, intervention implementers, and evaluators to gain continuous insights on use and usefulness and enhance data quality.  
R3.3 Ensure sufficient resourcing of future meta-evaluations & replicating the methodology to gain the best evidence for future learning and systematic review.

**R2**  
*To improve quality of Finnish development cooperation*

**R1**  
*On quality, use and usefulness of evaluation reports*

**R3**  
*For future meta-evaluations*



# Thank you

---



# Recommendations

## **R1.1 Enforce compliance with guidelines inside MFA for better ToRs.**



### HOW

- Adherence to guidelines can be strengthened in an institutionalised evaluation quality assurance process. (→ R1.2)
- As immediate reaction, high-level MFA management shall launch a call to underline importance of adherence, e.g.,
  - delivered in a speech during a physical meeting,
  - distributing official action letter,
  - during meta-evaluation results dissemination,
  - linked to updated evaluation development norm communication,
  - in the context of launching the upcoming Framework Agreement for Decentralized Evaluations and Reviews (FADER)
- Brown-bag meetings on good practices of ToR-drafting for follow-up.
  
- Main implementation responsibility: high-level MFA staff, EVA-11, regional units
- Urgency: high, priority: high

# Recommendations

**R1.2 Establish a quality assurance process inside MFA** to enforce evaluators' compliance with manual guidelines and ToR outside MFA.



## WHAT

- An evaluation quality assurance process (EQAP) is an institutionalised workflow commencing with selection of right point in time for evaluation of a pre-selected intervention and lasting to feedback provision on the evaluation process, or further to follow-up on implementation of management response.

## HOW

- Bring all steps into a logical order, display responsibilities of different actors, calendar time and working time required, point to supporting materials, indicate focal points for advice.
- Possible EQAP features:
  - more detailed annotation for inception and evaluations reports inclusive of elaborations on aspects and their level of detail to be covered (e.g., on discussion of intervention logics, gender-disaggregated assessments).
  - involve qualified reviewers to assure compliance of inception and evaluation reports with annotations, to relieve commissioners and to act as their sparing partners
  - establishing an EQAP help desk to support commissioners or outsource QA elements from commissioners.
- Link EQAP to FADER
- Main implementation responsibility: EVA-11 for process development, regional units for process application
- Urgency: high, should be systematically integrated in 2022 and reviewed in 2023, priority: high



# Recommendations

**R1.3 Improve knowledge management inside MFA** to foster organisational learning and enhance the usefulness of evaluations.



## WHAT

- A functional institutional knowledge management system ensures that relevant information on interventions and evaluations is regularly fed into a system and that such information is tagged to be accessible to interested users.

## HOW

- Seek professional support to inquire on the technical functionality and user-friendliness of existing system, including in the context of the development policy system reform (KeTTU).
- Seek advise on improvements to foster complete and continuous uploading of data by the majority of stakeholders.
- To avoid users' delay in uploading data, this should be anchored into EQAP.
  
- Main implementation responsibility: EVA-11 for setting up the system, regional units for archiving evaluation insights
- Urgency: high, should be systematically integrated in 2022, priority: high

# Recommendations

**R1.4 Continue further evaluation capacity development** for stakeholders inside and outside MFA on focused topics to adapt to increased evaluation quality standards, i.e., methodological rigour, provision of high-quality, evidence-based findings, and CCOs.



## HOW

- Prioritise and tailor training towards high-quality, evidence-based findings and appropriate capturing of CCOs and HRBA.
- Capacity development formats for commissioners of evaluations, evaluators and intervention implementers (M&E!).
- Consider compulsory web-based training and introduce certificates as precondition to award assignments to evaluators or comply with on-the-job training requirements inside MFA.
- Plan refresher courses from the start.
- Seek support from an evaluation capacity specialist and consult with key stakeholders.
- Inquire how efforts can be linked or streamlined with the upcoming FADER.
  
- Main implementation responsibility: EVA-11 to secure provision of services, regional units to support advertising for external evaluators and to foster the participation of own commissioners
- Urgency: high, immediately and with refreshers, priority: high

# Recommendations

**R1.5 Use facilitating factors** identified in this meta-evaluation like appropriate ToRs, follow-up of management response **as an entry point to enhance use and usefulness of evaluations.**



## HOW

- For better ToRs → R1.1., R.1.4
- For strengthening the follow-up on management responses link it to EQAP → R1.2
- Make the provision of a management response compulsory (or at least highly recommended) for decentralised evaluations
- Institutionalise follow-up on their implementation.
- Set clear and mandatory timelines for both and ensure enforcement.
- For selecting the right evaluation type introduce and maintain an easy-to-use inventory → R3.1
  
- Main implementation responsibility: EVA-11
- Urgency: medium, priority: medium

# Recommendations

**R2.1 Pay stronger attention to sustainability, coherence, and impact, as well as to CCOs and HRBA to work towards more comprehensive, high-quality interventions.**

**R2**  
To improve quality of  
Finnish development  
cooperation

## HOW

- Follow and participate in thematic, country-level, regional, and international exchanges to foster peer-to-peer learning
- Inside MFA lay stronger foundations ensuring better consideration of CCOs and HRBA already at planning stage of interventions.
  - Deliver mandatory focussed capacity development to obtain a “driving licence for appropriate CCOs and HRBA consideration”
  - Balance length, mode, and contents with standards set by Finnish Development Policy and corresponding guidelines and stakeholders’ needs.
  - Exploit synergies with implementation of R1.4.
- Main implementation responsibility: regional units, project planners, chief technical assistants, M&E officers
- Urgency: medium, priority: high

# Recommendations

**R2.2 Translate lessons learnt and recommendations provided by evaluators into action**, mainly in the fields of M&E, planning, sustainability, management, capacity development, financial aspects, and coherence to uplift the quality of Finnish development cooperation.



## HOW

- Appoint a technical expert group to work on translation of synthesised lessons learnt and recommendations into specific options for actions (e.g., topic-wise fact sheets, facilitation of communities of practice).
- Disseminate knowledge among MFA key stakeholders to support planning and implementation of Finnish aid.  
Potential options:
  - Well-established exchange fora
  - Tailored web-based inputs on key topics
  - General workshop formats with training inputs and discussion sections.
- Use the Framework Agreement for Planning Bilateral Development Cooperation to facilitate planning for better quality interventions.
- Involve an M&E specialist to improve functioning M&E system development at intervention level which is key to generate intervention-specific insights to improve Finnish development cooperation.
- Main implementation responsibility: regional units, project planners, chief technical assistants, M&E officers
- Urgency: medium, priority: high



# Recommendations

**R3.1 Set up an inventory of interventions and evaluations** to assess sample representativeness and save resources for future meta-evaluations.



## WHAT

- An inventory allows the application of efficient sampling strategies and increases the rigour of MFA's evaluation decisions.

## HOW

- Set up a simple inventory with key characteristics on interventions and corresponding evaluations.
- Keep the inventory as simple as possible and seek professional support for its development.
- To ensure usefulness and maintenance consult key stakeholders during all stages of inventory development and testing.
- Consider advantages of an interface to AHA-systems.
- Explore linkages to the development policy system reform (KeTTU).
- Start the system in the context of launching FADER.
  
- Main implementation responsibility: EVA-11 for set-up, regional units for archiving
- Urgency: low, priority: high

# Recommendations

**R3.2 Introduce digitised feedback sheets on evaluations** for MFA staff at headquarters and embassy level, intervention implementers, and evaluators to gain continuous insights on use and usefulness and enhance data quality and coverage for future meta-evaluations.



## WHAT

- Feedback sheets allow:
  - Immediate quality assessments of evaluations and thus support decision making on future evaluators.
  - Just-in-time peer learning from useful evaluation processes.
  - Self-assessment by commissioners

## HOW

- Keep feedback sheets short and simple
- Anchor distribution, inclusive of friendly reminders, in EQAP
- Go for a digital solution and support the introduction by a professional M&E expert
  
- Main implementation responsibility: EVA-11: provision of the tool, regional units: distribution of the tool and archiving feedbacks, stakeholders: provision of feedback
- Urgency: low, priority: medium

# Recommendations

**R3.3 Ensure sufficient resourcing of future meta-evaluations** and replicating the methodology to gain the best evidence for future learning and systematic review.

A green square logo with the text 'R3' in large, bold, dark green letters, and 'For future meta-evaluations' in smaller, dark green letters below it.

**R3**  
For future  
meta-evaluations

## WHAT

- Enlarge the budget or exclude analyses of lower interest from future assignments to save resources
- Possible options:
  - Skip some qualitative content assessment of underlying reasons for evaluators' assessment on OECD DAC in one meta-evaluation and then look at it again in a subsequent meta-evaluation.
  - Provide surplus funding to fully exploit gained evidence for future learning and systematic reviews.
- Main implementation responsibility: EVA-11
- Urgency: low, priority: high