

# EVALUATION

META-EVALUATION OF MFA'S PROJECT AND PROGRAMME  
EVALUATIONS IN 2017-2020



Ministry for Foreign  
Affairs of Finland



Evaluation on Finland's Development Policy and Cooperation

2022/1



© Ministry for Foreign Affairs of Finland 2022

This report can be downloaded through the home page of the Ministry for Foreign Affairs  
<https://um.fi/development-cooperation-evaluation-reports-comprehensive-evaluations>

Contact: [EVA-11@formin.fi](mailto:EVA-11@formin.fi)

ISBN 978-952-281-712-9 (PDF)

ISSN 2342-8341

Layout: Grano

Cover design: Dustin Go, NIRAS International Consulting



## EVALUATION

### METAEVALUATION OF MFA'S PROJECT AND PROGRAMME EVALUATIONS IN 2017-2020

Dr. Susanne Johanna Väth (Team Leader)

Dr. Stefan Silvestrini

Dr. Hansjörg Gaus

Petra Mikkolainen

Dr. Maja Flaig

Janis Wicke

**Consortium composed of:**



*Lead Company*

In collaboration with:



**2022/1**

This evaluation was commissioned by the Ministry for Foreign Affairs of Finland to Particip GmbH. This report is the product of the authors, and responsibility for the accuracy of the data included in this report rests with the authors. The findings, interpretations, and conclusions presented in this report do not necessarily reflect the views of the Ministry for Foreign Affairs of Finland.



# Contents

<b>ACRONYMS AND ABBREVIATIONS</b> .....	<b>IX</b>
<b>YHTEENVETO</b> .....	<b>XI</b>
<b>TÄRKEIMMÄT LÖYDÖKSET, JOHTOPÄÄTÖKSET JA SUOSITUKSET</b> .....	<b>XV</b>
<b>SAMMANFATTNING</b> .....	<b>XX</b>
<b>NYCKELRESULTAT, SLUTSATSER OCH REKOMMENDATIONER</b> .....	<b>XXIV</b>
<b>SUMMARY</b> .....	<b>XXIX</b>
<b>KEY FINDINGS, CONCLUSIONS AND RECOMMENDATIONS</b> .....	<b>XXXIII</b>
<b>1. Introduction</b> .....	<b>2</b>
1.1. Scope, purpose and objectives of the metaevaluation .....	2
1.2. Evaluation questions .....	3
1.3. Previous metaevaluation and MFA's management response .....	5
1.4. Structure of this report .....	6
<b>2. Approach, Methodology and Limitations</b> .....	<b>7</b>
2.1. General approach and methodology .....	8
2.2. Data sources and identification strategies .....	10
2.3. Limitations and coping strategies .....	11
<b>3. Context Analysis</b> .....	<b>14</b>
3.1. Finland's Development Policies .....	14
3.2. Delivery of Finnish aid .....	15
3.3. Evaluation reports in light of the Finnish development context (EQ1) .....	16
<b>4. Findings of the Metaevaluation</b> .....	<b>24</b>
4.1. Quality of underlying ToRs and their influence on the quality of evaluations (EQ3) .....	25
4.1.1. Quality of ToRs .....	25
4.1.2. Linkages between the quality of ToRs and quality of reports .....	27
4.2. Overall quality and reliability of decentralised evaluation reports (EQ2) .....	29
4.3. Differences, trends and evaluation capacity gaps (EQ4) .....	33
4.3.1. Differences and trends to the previous metaevaluation .....	33
4.3.2. Gaps regarding MFA's evaluation capacity .....	35
<b>5. Findings of the Summative Analysis</b> .....	<b>36</b>
5.1. Overall quality, strengths and weaknesses of the interventions (EQ8 & EQ5) .....	37
5.2. Consideration of cross-cutting objectives (CCOs), human rights-based approach (HRBA) and policy priority areas (PPAs) (EQ6) .....	39
5.3. Differences and trends to the previous metaevaluation (EQ9) .....	43
5.4. Lessons learnt and recommendations by evaluators (EQ7) .....	44



5.4.1. Lessons learnt presented in the evaluation reports .....	44
5.4.2. Recommendations identified in the evaluation reports .....	52
<b>6. Findings of the Use and Usefulness Assessment .....</b>	<b>66</b>
6.1. Underlying data for the assessment .....	66
6.2. Use of decentralised evaluations (EQ10) .....	68
6.3. Usefulness of evaluations (EQ11) .....	69
6.3.1. Overall usefulness of the evaluations .....	69
6.3.2. Usefulness of timing and recommendations .....	71
6.4. Drivers of usefulness (EQ12) .....	72
6.4.1. Facilitating and hampering factors for usefulness .....	72
6.4.2. Interviewees' lessons & recommendations .....	74
<b>7. Conclusions .....</b>	<b>76</b>
7.1. Component 1: Methodological quality assessment of evaluation reports and ToRs (EQ2, EQ3 & EQ4) .....	76
7.2. Component 2: Summative analysis of the quality of interventions (EQ5, EQ6, EQ7, EQ8 & EQ9) .....	78
7.3. Component 3: Use and usefulness assessment of the evaluations (EQ10, EQ11 & EQ12) ...	79
<b>8. Recommendations .....</b>	<b>81</b>
8.1. Recommendations on quality, use and usefulness of evaluation reports (EQ13 & 15) .....	82
8.2. Recommendations to improve the quality of Finnish DevCo (EQ14) .....	86
8.3. Recommendations for future metaevaluation (EQ16) .....	87
<b>REFERENCES .....</b>	<b>90</b>
<b>THE ASSIGNMENT, THE EVALUATION TEAM, AND THE QUALITY ASSURANCE PROCESS .....</b>	<b>92</b>
<b>ANNEX 1: TERMS OF REFERENCE .....</b>	<b>97</b>
<b>ANNEX 2: PEOPLE INTERVIEWED .....</b>	<b>108</b>
<b>ANNEX 3: DOCUMENTS CONSULTED .....</b>	<b>110</b>
<b>ANNEX 4: ANALYSIS GRID .....</b>	<b>111</b>
<b>ANNEX 5: METHODOLOGICAL DETAILS .....</b>	<b>113</b>
<b>ANNEX 6: QUALITY ASSESSMENT TOOL FOR REPORTS AND TOR .....</b>	<b>117</b>
<b>ANNEX 7: CONTENT ASSESSMENT TOOL .....</b>	<b>139</b>
<b>ANNEX 8: INTERVIEW GUIDELINE .....</b>	<b>156</b>
<b>ANNEX 9: SURVEY QUESTIONS AND RESPONSES .....</b>	<b>160</b>
<b>ANNEX 10: LIST OF EVALUATION REPORTS RECEIVED AND USED .....</b>	<b>163</b>
<b>ANNEX 11: QUALITY ASSESSMENT ON SINGLE SECTIONS OF THE REPORTS .....</b>	<b>172</b>
<b>ANNEX 12: INTERVENTIONS' QUALITY ON SINGLE OECD DAC CRITERIA .....</b>	<b>186</b>



<b>ANNEX 13: FURTHER DETAILS ON CCOS AND HRBA</b> .....	<b>228</b>
---	------------

<b>ANNEX 14: FURTHER QUALITY ASSESSMENTS BY SURVEY RESPONDENTS</b> .....	<b>232</b>
--	------------

<b>ANNEX 15: STATISTICAL TESTS</b> .....	<b>234</b>
--	------------

## **FIGURES**

Figure 1: Conclusions and recommendations at a glance .....	XXXVII
Figure 2: Summary of the section .....	7
Figure 3: Channels of Finnish aid .....	15
Figure 4: Year of publication of the evaluation report (n=80) .....	17
Figure 5: Nature of the evaluation (n=80) .....	17
Figure 6: Commissioner of the evaluation (n=80) .....	18
Figure 7: The implementer of the evaluation (n=80) .....	18
Figure 8: Finland's budget of the intervention (n=59) .....	19
Figure 9: Overall budget of the intervention (n=67) .....	19
Figure 10: Net evaluation budget of the intervention (n=38) .....	20
Figure 11: Duration of the evaluation (n=73) .....	20
Figure 12: Regional distribution of interventions (n=80) .....	21
Figure 13: Sectorial distribution of interventions (n=80) .....	21
Figure 14: Distributions of PPAs (n=72) .....	22
Figure 15: Overall quality of ToRs (n=70) .....	25
Figure 16: Quality of ToR (n=70) .....	26
Figure 17: Overall quality of reports (n=80) .....	30
Figure 18: Quality Dimensions (n=80) .....	30
Figure 19: Quality of findings (n=80) .....	31
Figure 20: Appropriate capturing of DAC criteria .....	32
Figure 21: Quality aggregates of the previous and current metaevaluation .....	34
Figure 22: Quality of the intervention (n=72) .....	37
Figure 23: Quality on single DAC criteria .....	38
Figure 24: Integration of CCOs and HRBA (n=80) .....	40
Figure 25: Word cloud of all keywords .....	40
Figure 26: Possible trends on overall intervention quality .....	43
Figure 27: Survey responses (n=119) and share of perspectives per case (n=80) .....	67
Figure 28: Usage of evaluations (multiple answers, 319 usages out of 199 survey responses) ...	69
Figure 29: Overall usefulness of evaluations (case level, ranges) .....	70
Figure 30: Timing and timeliness (case level) and Recommendations (case level) .....	72
Figure 31: Main facilitating factors .....	73
Figure 32 Summary of the section .....	81
Figure 33: Overall rating of introductions (n=80) .....	173
Figure 34: Contents of introduction (n=80) .....	173
Figure 35: Overall rating of context analysis (n=60) .....	174
Figure 36: Contents of context analysis (n=60) .....	174
Figure 37: Overall rating on methodology (n=80) .....	175
Figure 38: Description and appropriation of methods (n=80) .....	176
Figure 39: Quality of findings (n=80) .....	178
Figure 40: Appropriate capturing of DAC criteria .....	180
Figure 41: Conclusions are derived from findings (n=80) .....	181
Figure 42: Recommendations are derived from findings and conclusions (n=80) .....	182



Figure 43: Quality of recommendations (n=80) .....	182
Figure 44: Overall quality of executive summaries (n=80) .....	183
Figure 45: Projects' relevance according to the evaluation reports (n=72) .....	186
Figure 46: Number of reports assessing different aspects of relevance (n=72) .....	187
Figure 47: Assessment of different aspects of relevance .....	188
Figure 48: Projects' coherence according to the evaluation reports (n=72) .....	195
Figure 49: Number of reports assessing different aspects of coherence (n=72) .....	196
Figure 50: Assessment of different aspects of coherence .....	196
Figure 51: Projects' effectiveness according to the evaluation reports (n=72) .....	197
Figure 52: Number of reports assessing different aspects of effectiveness (n=72) .....	198
Figure 53: Assessment of different aspects of effectiveness .....	198
Figure 54: Projects' efficiency according to the evaluation reports (n=72) .....	206
Figure 55: Number of reports assessing different aspects of efficiency (n=72) .....	207
Figure 56: Assessment of different aspects of efficiency .....	207
Figure 57: Interventions' impact according to the evaluation reports (n=72) .....	217
Figure 58: Number of reports assessing different aspects of impact (n=72) .....	218
Figure 59: Assessment of intervention quality on different aspects of impact .....	219
Figure 60: Interventions' sustainability according to the evaluation reports (n=72) .....	223
Figure 61: Number of reports assessing different aspects of sustainability (n=72) .....	224
Figure 62: Assessment of intervention quality on different aspects of sustainability .....	224
Figure 63: Comparison of recent (n=80) and previous (n=51) metaevaluation findings on CCOs and HRBA .....	228
Figure 64: Quality assessment of evaluations and ToRs by survey respondents (case level) ..	233

## TABLES

Table 1: OLS Regression Analysis .....	28
Table 2: Results on the frequencies of keyword groups in evaluation reports (n=72) .....	41
Table 3: Number of reports including lessons learnt categorised under different themes .....	45
Table 4: Number of reports including recommendations categorised under different themes ..	52
Table 5: Work plan with a division of assigned tasks and timeline .....	95
Table 6: Communication and dissemination plan .....	96
Table 7: Example: four-step scale question .....	114
Table 8: Frequencies of gender equality related keywords in evaluation reports .....	229
Table 9: Percentage of reports that mention selected keywords related to non-discrimination .....	230
Table 10: Percentage of reports that mention selected keywords related to climate sustainability .....	230
Table 11: Percentage of reports that mention selected keywords related to HRBA .....	231
Table 12: Percentage of reports that mention additional keywords from MFA's Theories of Change .....	231
Table 13: Mann-Whitney tests PPA2 vs other PPAs .....	234
Table 14: Mann-Whitney tests PPA3 vs other PPAs .....	235
Table 15: Mann-Whitney tests PPA4 vs other PPAs .....	235
Table 16: OLS Regression Analysis .....	236
Table 17: Mann-Whitney test on survey responses of MFA headquarters staff vs embassy staff .....	237
Table 18: Mann-Whitney test on survey responses of implementers vs MFA embassy staff ....	238
Table 19: Mann-Whitney test on survey responses of implementers vs all MFA staff .....	239



Table 20: Spearman Correlation between Report Quality (QA) and Overall Usefulness (Survey) .....	239
Table 21: Spearman Correlation between Intervention Quality (CA) and Intervention Quality (Survey) .....	239

## BOXES

Box 1: Examples for lessons referring to the theme capacity .....	46
Box 2: Examples for lessons learnt in the theme planning .....	47
Box 3: Examples for lessons in the theme coherence .....	48
Box 4: Examples for lessons in the theme participation .....	49
Box 5: Examples for lessons in the theme financial aspects .....	50
Box 6: Examples for lessons in the theme of sustainability .....	51
Box 7: Examples for recommendations on M&E .....	54
Box 8: Examples for recommendations on coherence .....	56
Box 9: Examples for recommendations on planning .....	57
Box 10: Examples for recommendations on sustainability .....	59
Box 11: Examples for recommendations on management .....	61
Box 12: Examples for recommendations on financial aspects .....	63
Box 13: Examples for recommendations on capacity development .....	64
Box 14: Examples of reasons for the assessment of relevance related to the needs of the target groups .....	191
Box 15: Examples of reasons for the assessment of relevance related to the needs of the final beneficiaries .....	194
Box 16: Examples of reasons for the assessment related to outcome achievement .....	200
Box 17: Examples of intervention resulting in benefits for the target group .....	202
Box 18: Examples of reasons for the assessment related to intervention resulting in benefits for the final beneficiaries .....	205
Box 19: Examples of reasons for the assessment related to timeliness of the intervention ...	209
Box 20: Examples of reasons for the assessment related to cost efficiency of the intervention .....	211
Box 21: Examples of reasons for the assessment related efficiency regarding personnel .....	213
Box 22: Examples of reasons for the assessment related to management efficiency .....	215
Box 23: Examples of unintended positive impacts .....	220
Box 24: Examples of unintended negative impacts .....	221
Box 25: Examples of reasons related to the assessment of sustainability of interventions ....	226





# Acronyms and Abbreviations

<b>AHA</b>	MFA's Database Management System
<b>CCO</b>	Cross-cutting Objectives
<b>CSO</b>	Civil Society Organisations
<b>DAC</b>	Development Assistance Committee
<b>DRR</b>	Disaster Risk Reduction
<b>DW</b>	Decent Work
<b>EMS</b>	Evaluation Management System
<b>EQ</b>	Evaluation Question
<b>EU</b>	European Union
<b>EUR</b>	Euro
<b>EVA-11</b>	Development Evaluation Unit
<b>FADER</b>	Framework Agreement for Decentralized Evaluations and Reviews
<b>FCR</b>	Findings, conclusions and recommendations
<b>FGM</b>	Female Genital Mutilation
<b>Fin DevCo</b>	Finnish Development Cooperation
<b>GBV</b>	Gender-Based Violence
<b>GDP</b>	Gross Domestic Product
<b>GNI</b>	Gross National Income
<b>HQ</b>	Headquarters
<b>HRBA</b>	Human Rights Based Approach
<b>ICT</b>	Information and Communications Technology
<b>ILO</b>	International Labour Organization
<b>IOM</b>	International Organization for Migration
<b>IT</b>	Information Technology
<b>KeTTU</b>	MFA's comprehensive reform of its development cooperation practices
<b>M&amp;E</b>	Monitoring and Evaluation
<b>MENA</b>	Middle East and North Africa
<b>MFA</b>	Ministry for Foreign Affairs (Finland)
<b>MIDA FINNSOM</b>	Migration for Development in Africa Finnish Somali
<b>MTR</b>	Mid-Term Review
<b>NDC</b>	Nationally Determined Contributions
<b>NGO</b>	Non-Governmental Organisation
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>OLS</b>	ordinary least squares regression
<b>PPA</b>	Policy Priority Area
<b>QA</b>	Quality Assurance
<b>RBM</b>	Result Based Management
<b>SDG</b>	Sustainable Development Goals
<b>SRHR</b>	Sexual and reproductive health and rights
<b>SUFORD</b>	Sustainable Forestry and Rural Development Project
<b>SUFORD-AF</b>	Sustainable Forestry and Rural Development Project Additional Financing Project



<b>SUFORD-SU</b>	Scaling-up Participatory Sustainable Forest Management Project
<b>TA</b>	Technical Assistance
<b>ToR</b>	Terms of Reference
<b>UN</b>	United Nations
<b>UNDP</b>	United Nations Development Programme
<b>UNFCCC</b>	United Nations Framework Convention on Climate Change
<b>UNFPA</b>	United Nations Population Fund
<b>UNICEF</b>	United Nations Children's Fund
<b>WASH</b>	Water, Sanitation and Hygiene
<b>WB</b>	World Bank



# Yhteenveto

## Evaluoinnin tarkoitus, laajuus ja tavoitteet

Tämän seitsemännen hanke- ja ohjelmaevaluointien metaevaluoinnin tarkoituksena oli tukea ulkoministeriön evaluointitoimintaa ja oppimista. Toimeksiannon tarkoituksena oli myös tuottaa tietoa eduskunnalle laadittavaa kehityspolitiikan tulosraporttia varten.

Metaevaluointi kattoi vuosina 2017–2020 valmistuneet hanke- ja ohjelmaevaluointiraportit ja niiden tehtäväkuvaukset. Ne käsittelivät Suomen tukemia kahdenvälisiä, monenkeskisiä tai ns. multi-bi hankkeita, joissa yhteistyötä tehdään maatasolla monenkeskisen toimijan kanssa.<sup>1</sup>

Metaevaluointi koostui kolmesta osasta, joilla kullakin oli omat tavoitteensa:

1. Metaevaluointi tarkasteli evaluointien tehtäväkuvausten ja raporttien menetelmällistä laatua
2. Summatiivinen eli kokoava analyysi arvioi kehitysyhteistyön sisällöllistä laatua ja kokosi yhteen evaluointien opit sekä suositukset
3. Analyysi evaluointien käytettävyydestä ja hyödyllisyydestä tuotti tietoa kehitysyhteistyön toteuttajien ja evaluointien tilaajien näkökulmasta.

Metaevaluoinnin tavoitteena oli muodostaa näyttöön perustuvia johtopäätöksiä ja esittää ulkoministeriölle suosituksia, joita toteuttamalla ulkoministeriö (i) kehittää evaluointien laatua, hallintokäytänteitä ja osapuolten evaluointiosaamista, (ii) parantaa Suomen tukemien kahdenvälisen ja monenkeskisten hankkeiden sisällöllistä laatua, ja (iii) vahvistaa hanke- ja ohjelmaevaluointien hyödyllisyyttä niiden käyttäjille.

## Metodologia

Metaevaluoinnin metodologia käytti suurelta osin edellisen metaevaluoinnin metodologiaa. Standardoituja arviointityökaluja käytettiin sekä arvioimaan (i) 80:n evaluointiraportin ja niihin liittyvien 70:n tehtäväkuvauksen laatua että (ii) kehitysyhteistyön sisällöllistä laatua perustuen 72:n evaluointiraporttiin, jotka olivat läpäisseet laaduntarkastuksen. Laadullisessa sisältöanalyysissä tarkasteltiin evaluointiraporttien esiintuomia oppeja, suosituksia, taustalla vaikuttavia syitä evaluoijien arviointeihin ja sukupuolten väliseen tasa-arvoon liittyviä tuloksia. Avainsanahauulla selvitettiin läpileikkaavien tavoitteiden ja HRBA:n huomioimista evaluointiraporteissa. Uutena ominaisuutena haastateltiin 26 ulkoministeriön edustajaa Helsingissä ja edustustoissa. Lisäksi saatiin 199 vastausta kyselyyn, joka lähetettiin evaluointien yhteyshenkilöille ulkoministeriössä sekä kehitysyhteistyöhankkeiden ja -ohjelmien toteuttajille. Kyselyn avulla saatiin tietoa evaluointien käytöstä ja hyödyllisyydestä.

---

<sup>1</sup> Tästä eteenpäin yhteenvedossa käytetään ilmaisua ”kahdenväliset ja monenkeskiset hankkeet” kattamaan kaikki nämä kolme hankemuotoa.



## Vastaukset evaluointikysymyksiin

### 1. Ulkoministeriön hanke- ja ohjelmaevaluointien kokonaisuus

Ulkoministeriö ei ylläpidä kattavaa listaa kahdenvälisistä ja monenkeskisistä hankkeista. Siten ei voitu määrittellä, miten kattavasti metaevaluoinnissa arvioidut hanke- ja ohjelmaevaluointiraportit edustavat Suomen kehitysyhteistyötä. Vertailu edelliseen metaevaluointiin paljasti hanke- ja ohjelmaevaluoinneissa yhtäläisyyksiä (evaluoinnin luonne, evaluoinnin tilaajat, maantieteellinen kattavuus, Suomen rahoitusosuudet) ja eroja (hankkeiden ja evaluoinnin budjetit, sekä teema-kohtainen jakautuminen).

### 2. Evaluointiraporttien laatu ja luotettavuus sekä raporttityyppien väliset erot.

Noin puolet raporteista oli laadultaan tyydyttäviä. Lopuissa oli parantamisen varaa ja muutamat olivat olennaisilta osin puutteellisia. Heikkouksia esiintyi liittyen metodologiaan, korkealaatuiseen ja näyttöön perustuviin löydöksiin sekä OECD:n kehitysapukomitean kriteerien (erityisesti johdonmukaisuuden, kestävyyden ja tehokkuuden) asianmukaiseen huomiointiin. Evaluointiraportit olivat kuitenkin kaiken kaikkiaan melko luotettavia eikä niiden laadussa ollut eroja eri laisten raporttityyppien välillä esim. ulkoministeriön ja muiden kuin ulkoministeriön teettämien evaluointien välillä).

### 3. Tehtävänkuvauksen laatu sekä niiden ja raporttien laadun välinen yhteys

Kaksi kolmesta tehtävänkuvauksesta oli laadultaan tyydyttäviä, ja lopuissa oli parantamisen tarvetta. Regressioanalyysi osoitti, että paremmat tehtävänkuvaukset johtavat laadukkaampiin evaluointituloksiin ja siten parempiin raportteihin.

### 4. Erot verrattuna edelliseen metaevaluointiin ja ulkoministeriön evaluointiosaamisen puutteet

Staattinen vertailu, eli menetelmä, jossa toistettiin täsmälleen sama analyysi kuin edellisessä metaevaluoinnissa, ei paljastanut eroja tämän ja edellisen metaevaluoinnin tulosten välillä. Sen sijaan dynaaminen vertailu, jossa otettiin huomioon tiukentuneet evaluointistandardit, viittasi jonkinasteiseen evaluointiraporttien laadun heikkenemiseen. Molemmat havainnot osoittavat, että ulkoministeriön kehitysyhteistyön evaluointiosaamisessa on joitakin puutteita ja että haasteet tulevat todennäköisesti lisääntymään tulevaisuudessa. Toimeksiantojen tehtävänkuvauksen perusteella suurimmat puutteet liittyivät riittävän syvällisen evaluointimetodologian ohjeistuksen tarjoamiseen ja läpileikkaavien tavoitteiden ja ihmisoikeuksiin perustuvan lähestymistavan järjestelmälliseen huomioimiseen.

### 5. Hankkeiden laatu OECD:n kehitysapukomitean kriteerien perusteella

Lähes kaikki evaluoidut hankkeet olivat erittäin tai kohtalaisen tarkoituksenmukaisia. Noin kaksi kolmasosaa evaluoiduista hankkeista oli vähintään kohtalaisen tuloksellisia ja tehokkaita. Lähes kaksi kolmasosaa evaluoiduista hankkeista oli heikkohkoja tai hyvin heikkoja kestävyyden osalta. Lähes puolet evaluoiduista hankkeista olivat heikkohkoja tai hyvin heikkoja myös johdonmukaisuuden ja vaikuttavuuden näkökulmasta, mutta näiden evaluointikriteerien kohdalla hanke- ja ohjelmaevaluointien kattama tieto oli rajallista, mikä saattaa vaikuttaa tämän metaevaluoinnin löydöksiin.

### 6. Läpileikkaavat tavoitteet ja ihmisoikeusperustainen lähestymistapa

Sukupuolten välistä tasa-arvoa oli käsitelty osassa raportteja ja niiden perusteella se on osittain valtavirtaistettu. Sen sijaan syrjimättömyyden, ilmastokestävyyden ja ihmisoikeusperustaisen lähestymistavan huomioonottamista ei voitu analysoida, koska näitä teemoja ei oltu yleensä käsitelty analysoiduissa hanke- ja ohjelmaevaluointiraporteissa.



## **7. Tärkeimmät suositukset ja opit Suomen kehitysyhteistyön parantamiseksi**

Noin kaksi kolmasosaa raporteista sisälsi suosituksia seurantaan ja evaluointiin, johdonmukaisuuteen ja suunnitteluun, ja noin puolet kestävytyteen, hallintoon, rahoitukseen ja kapasiteetin kehittämiseen liittyen. Pääasialliset opit liittyivät kapasiteetin vahvistamiseen, hankkeiden suunnitteluun ja johdonmukaisuuteen. Niitä kuitenkin esiintyi vain hieman alle 20 %:ssa raporteista, joten niitä ei voi pitää kehitysyhteistyölle luonteenomaisina oppeina.

## **8. Suomen kehitysyhteistyön sisällöllinen laatu; keskeisimmät vahvuudet ja haasteet**

Kaiken kaikkiaan kahden- ja monenvälisen hankkeiden laatu on melko hyvä. Noin 20 % hankkeista on korkealaatuisia, noin 60 % kohtalaisen laadukkaita ja noin 20 % heikkolaatuisia. Toiminnan tarkoituksenmukaisuutta voidaan pitää Suomen kehitysyhteistyön vahvuutena. Kestävytydessä on eniten parantamisen varaa kuten myös johdonmukaisuudessa ja vaikuttavuudessa.

## **9. Keskeisimmät erot hankkeiden laadussa verrattuna edelliseen metaevaluointiin**

Yleisesti ottaen hankkeiden laadussa ei havaittu muutoksia edelliseen metaevaluointiin verrattuna. Suorat vertailut eivät olleet mahdollisia, koska metaevaluoinneissa tarkasteltiin eri hankkeita ja otos oli eri.

## **10. Evaluointiraporttien käyttö**

Hanke- ja ohjelmaevaluointien raportteja on käytetty tyypillisesti ulkoministeriön tiimien oppimiseen, päätöksentekoon, yhteistyön jatkon suunnitteluun ja käynnissä olevien hankkeiden hallintoihin. Suomen kehitysyhteistyön kannalta strategisempia käyttötapoja mainittiin harvemmin, mutta niillä vaikuttaisi olevan merkitystä.

## **11. Evaluointien hyödyllisyys, ajoitus, oikea-aikaisuus, suositukset ja suositusten täytäntöönpano**

Valtaosa evaluoinneista arvioitiin hyödyllisyyden näkökulmasta tyydyttäväksi tai erittäin hyödylliseksi. Evaluointien ajoitus sekä virallisten ja epävirallisten tulosten toimittamisen oikea-aikaisuus arvioitiin yleensä hyväksi. Sama koskee suositusten asianmukaisuutta ja realistisuutta sekä evaluoinneista oppimista ja suositusten täytäntöönpanoa.

## **12. Tyypilliset evaluointien käytettävyyttä lisäävät ja vähentävät tekijät**

Yleisiä evaluointien hyödyllisyyttä lisääviä tekijöitä olivat sopivan evaluointityypin valinta (esim. välievaluointi vai loppuevaluointi), asianmukaisten tehtävänkuvausten laatiminen, johdon vastineiden laatiminen, johdon vastineissa määriteltyjen toimien seuranta, sekä evaluointitiimin temaattinen ja metodologinen pätevyys. Yleisiä evaluointien hyödyllisyyttä vähentäviä tekijöitä ei sen sijaan noussut selkeästi esiin. Haastateltujen ulkoministeriön edustajien esittämät suositukset ja kokemukset evaluointien hyödyllisyyden lisäämiseksi koskivat lähinnä organisaation toimivuuteen liittyviä seikkoja, evaluointien ajoitusta ja oikea-aikaisuutta sekä raporttien laatuun liittyviä näkökohtia. Kehitysyhteistyön parissa toimivat henkilöt toivoivat tukea i) evaluointien laajuuden ja painopisteiden selkiyttämiseksi, ii) prosessin tarjoaman oppimisen tehostamiseksi ja iii) evaluoinneista saatujen tulosten viestimiseksi laajemmin ulkoministeriössä.

## **13. Metaevaluoinnin suositukset evaluointien laadun parantamiseksi.**

Joka kolmannessa tehtävänkuvauksessa ilmeni parannustarpeita, joten ulkoministeriön evaluointiosaaminen on edelleen puutteellista. Evaluointiosaamisen vahvistaminen on tärkeää ulkoministeriölle, sillä laadukkaat tehtävänkuvaukset parantavat evaluointiraporttien laatua (2. johtopäätös). Metaevaluointi suosittelee, että ulkoministeriö varmistaa, että evaluointiohjeita noudatetaan. Jo tämä auttaisi tuottamaan parempia tehtävänkuvauksia. Koska useimpien hanke- ja ohjelmaevaluointien metodologiat ovat olleet heikkoja, raporttien löydökset ovat vain jokseenkin luotettavia



(1. johtopäätös). Ulkoministeriö on myös vaarassa jäädä jälkeen aiemmin saavuttamastaan laatu-  
tutasosta, koska evaluointistandardit ovat tiukentuneet viime vuosina (3. johtopäätös). Tämän  
perusteella suosittelemme, että ulkoministeriö vakiinnuttaa laadunvarmistusprosessin, jolla  
taataan, että evaluoijat noudattavat evaluointikäsi kirjaa, muita olemassa olevia linjauksia ja oh-  
jeistuksia, sekä toimeksiannolle asetettua tehtävänkuvausta (suositus 1.2). Lisäksi suosittelemme,  
että jatketaan evaluointiosaamisen ja omistajuuden vahvistamista, jotta henkilökunta valvoo, että  
evaluointien laatuvaatimukset toteutuvat (esim. metodologinen osaaminen, laadukkaiden, näyt-  
töön perustuvien löydösten tuottaminen ja läpileikkaavat tavoitteet) (suositus 1.4).

#### **14. Metaevaluoinnin suositukset hankkeiden laadun parantamiseksi**

Kahdenvälisen ja monenkeskisten hankkeiden osalta laatu todettiin kohtuulliseksi, vaikka joka  
viidennessä hankkeessa havaittiinkin puutteita. Erityisesti OECD:n kehitysapukomitean kestä-  
vyyttä ja johdonmukaisuutta koskevien kriteerien osalta kaikilla kehitysyhteistyökumppaneilla on  
vielä huomattavasti parantamisen varaa (4. johtopäätös). Vaikka sukupuolten välinen tasa-arvo  
on otettu huomioon useammin kuin muut läpileikkaavat tavoitteet ja ihmisoikeusperustainen lä-  
hestymistapa, Suomen kehityspolitiikan linjauksia ei ollut vielä sisällytetty hankkeisiin riittävän  
järjestelmällisesti (5. johtopäätös). Tämän vuoksi suosittelemme, että ulkoministeriössä kiinni-  
tetään enemmän huomiota toiminnan kestävyteen, johdonmukaisuuteen ja vaikuttavuuteen,  
läpileikkaaviin tavoitteisiin ja ihmisoikeusperustaiseen lähestymistapaan kokonaisvaltaisemman  
ja laadukkaamman kehitysyhteistyön toteuttamiseksi (suositus 2.1). Evaluointien suositukset  
viittasivat siihen, että Suomen rahoittamissa hankkeissa on vielä paljon parantamisen varaa.  
Seuranta- ja evaluointijärjestelmät sekä suunnittelu, kestävyys, johdonmukaisuus, hallinnointi,  
kapasiteetin vahvistaminen ja rahoitusnäkökohdat eivät ole vielä toivotulla tasolla (6. johtopää-  
tös). Siksi suosittelemme, että evaluointiraportteihin kirjatut näihin liittyvät opit ja suositukset  
käytännössä toteutetaan (suositus 2.2).

#### **15. Metaevaluoinnin suositukset evaluointien käytön ja hyödyllisyyden parantamiseksi**

Evaluointiraporttien hyöty jää usein yksittäisen virkamiehen tasolle. Tiedolla johtamiseen liittyvät  
puutteet haittaavat ulkoministeriön sisäistä oppimista (8. johtopäätös). Näin ollen suosittelemme,  
että ulkoministeriön tiedolla johtamista parannetaan sisäisten oppimisprosessien edistämiseksi  
ja evaluointien hyödyllisyyden lisäämiseksi (suositus 1.3). Hajautetut evaluoinnit tukevat hyvin  
oppimista, päätöksentekoa, suunnittelua ja mukautuvaa johtamista ulkoministeriössä. On silti  
mahdollista lisätä evaluointien käyttöä yksittäisiä hankkeita laajemmalle ja välttää käyttämästä  
resursseja evaluointeihin, joista on vain vähän hyötyä (7. johtopäätös). Monet näistä kehitettävissä  
olevista asioista kuuluvat ulkoministeriön vastuun ja vaikutusmahdollisuuksien piiriin (9. johto-  
päätös). Ulkoministeriön tulisi siis ottaa tilanteen parantamisessa lähtökohdaksi tekijät, joiden on  
todettu edesauttavan evaluointien käyttöä ja hyödyllisyyttä (kuten laadukkaat tehtävänkuvaukset,  
johdon vastineissa määriteltujen toimien seuranta ja oikean evaluointityypin valinta) (suositus 1.5).

#### **16. Metaevaluoinnin suositukset tulevien metaevaluointien parantamiseksi**

Tässä toimeksiannossa kohdattujen haasteiden perusteella ehdotamme, että ulkoministeriö alkaa  
ylläpitää kattavaa luetteloa kehitysyhteistyöhankkeista ja niiden evaluoinneista. Tämä helpottaa  
metaevaluoinnin otoksen edustavuuden analysointia. Järjestelmällinen luettelointi mahdollistaa  
myös evaluointitiimin resurssien säästämisen muihin tehtäviin tulevaisuudessa metaevaluoinneissa  
(suositus 3.1). Lisäksi evaluointien digitaaliset palautelomakkeet ulkoministeriön henkilöstölle ja  
hankkeiden toteuttajille tulisi ottaa käyttöön. Näin saadaan jatkuvaa tietoa evaluointien käytöstä  
ja hyödyllisyydestä sekä parannetaan analysoitavan tiedon laatua ja kattavuutta tulevia metaevalu-  
ointeja varten (Suositus 3.2). Suosittelemme myös, että tulevaisuudessa metaevaluoinneissa varmistetaan  
riittävät resurssit ja toistetaan käytettyä menetelmää, mikä edesauttaa systemaattista tarkastelua  
ja oppimista. (Suositus 3.3).



# Tärkeimmät löydökset, johtopäätökset ja suositukset

LÖYDÖKSET	JOHTOPÄÄTÖKSET	SUOSITUKSET
<b>OSA 1. RAPORTTIEN JA TEHTÄVÄNKUVAUSTEN LAATU</b>		
<b>4.2 Raporttien laatu</b> <ul style="list-style-type: none"><li>Lähes puolet raporteista oli laadultaan tyydyttäviä ja lähes puolessa oli parantamisen varaa.</li><li>Löydösten, niiden perustana olevien menetelmien, läpileikkaavien tavoitteiden ja ihmisoikeusperustaisen lähestymistavan ja yhteenvetojen laadussa oli useimmiten parantamisen varaa.</li></ul>	<b>1.</b> Ulkoministeriö ei useinkaan valvo laatimiensa ohjeistusten noudattamista. Tämän vuoksi useimmat evaluointiraportit perustuivat suhteellisen heikkoihin menetelmiin. Evaluointiraporteissa esitettyjä tuloksia voidaan silti pitää jokseenkin luotettavina.	<b>1.2.</b> Ulkoministeriö perustaa laadunvarmistusprosessin, jolla taataan, että evaluoijat noudattavat evaluointikäsi kirjaa, muita olemassa olevia linjauksia ja ohjeistuksia, sekä toimeksiannon tehtäväkuvausta.  <b>1.4.</b> Ulkoministeriö jatkaa henkilökunnan evaluointiosaamisen ja omistajuuden kehittämistä puutteellisilta osin, jotta he pystyvät noudattamaan tiukentuvia evaluointien laatuvaatimuksia (esim. metodologinen osaaminen, laadukkaiden, näyttöön perustuvien löydösten tuottaminen ja läpileikkaavat tavoitteet).
<b>4.1.1 Tehtäväkuvausten laatu</b> <ul style="list-style-type: none"><li>Tehtäväkuvausten yleinen laatu oli tyydyttävä, mutta huomattavassa osassa oli parantamisen tarvetta.</li></ul> <b>4.1.2 Tehtäväkuvausten ja raporttien laadun välinen yhteys</b> <ul style="list-style-type: none"><li>Tehtäväkuvausten laatu vaikuttaa evaluoinneissa esitetyn näytön laatuun.</li></ul>	<b>2.</b> Joka kolmannessa tehtäväkuvauksessa ilmeni parannustarpeita, joten ulkoministeriön evaluointiosaaminen on edelleen puutteellista. Evaluointiosaamisen vahvistaminen on tärkeää ulkoministeriölle, sillä laadukkaat tehtäväkuvaukset parantavat evaluointiraporttien laatua.	<b>1.1.</b> Evaluoinnista vastaavat tahot UM:ssä varmistavat, että tehtäväkuvausten laadintaan tehtyjä ohjeita noudatetaan, jotta voidaan varmistaa tehtäväkuvausten parempi laatu.  <b>1.4.</b> Ulkoministeriö jatkaa henkilökunnan evaluointiosaamisen kehittämistä puutteellisilta osin.



LÖYDÖKSET	JOHTOPÄÄTÖKSET	SUOSITUKSET
<p><b>4.3.1 Erot ja trendit edelliseen metaevaluointiin verrattuna</b></p> <ul style="list-style-type: none"><li>• Staattinen vertailu, eli menetelmä, jossa toistettiin täsmälleen sama analyysi kuin edellisessä meta-evaluoinnissa, ei paljastanut eroja tämän ja edellisen metaevaluoinnin välillä. Sen sijaan dynaaminen vertailu, jossa otettiin huomioon tiukentuneet evaluointistandardit, viittasi jonkinasteiseen evaluointiraporttien laadun heikkenemiseen.</li><li>• Tässä meta-evaluoinnissa tarkasteltiin kuitenkin eri raporteja kuin edellisessä. Lisäksi tällä kertaa aineistossa oli mukana jokseenkin erityyppisiä evaluointeja (mm. rahastoja käsitteleviä toimeksiantoja), mikä saattaa osaltaan selittää laadun heikentymistä.</li></ul> <p><b>4.3.2 Puutteet ulkoministeriön evaluointiosaamisessa</b></p> <ul style="list-style-type: none"><li>• Laadunarviointi paljasti heikkouksia osaamisessa ulkoministeriön sisällä ja ulkopuolella. Parannettavaa on noin kolmasosassa tehtäväkuvauksissa ja noin puolessa evaluointiraporteissa. Molemmista puuttuu useimmiten järjestelmällinen läpileikkaavien tavoitteiden ja ihmisoikeusperustaisen lähestymistavan huomioon ottaminen.</li><li>• Raporttien laadun heikentyminen saattaa johtua ainakin osin viime vuosina tiukentuneista evaluointistandardeista.</li></ul>	<p><b>3.</b> Ulkoministeriö on vaarassa jäädä jälkeen aiemmin saavuttamastaan evaluointien laatutasosta, koska evaluointistandardit ovat tiukentuneet viime vuosina.</p>	<p><b>1.2.</b> Ulkoministeriöperustaa laadunvarmistusprosessin, jolla taataan, että evaluointinoudattavat evaluointikäsi kirjaa, muita olemassa olevia linjauksia ja ohjeistuksia, sekä toimeksiannon tehtäväkuvauksista.</p> <p><b>1.4.</b> Ulkoministeriö jatkaa henkilökunnan valuointiosaamisen kehittämistä puutteellisilta osin.</p>
<b>OSA 2. HANKKEIDEN SISÄLLÖLLINEN LAATU</b>		
<p><b>5.1 Hankkeiden yleinen laatu, vahvuudet ja heikkoudet.</b> Kaiken kaikkiaan kahdenvälisen ja monenkeskisten hankkeiden laatu on melko hyvä. Noin 20 % hankkeista on korkealaatuisia, noin 60 % kohtalaisen laadukkaita ja noin 20 % heikkolaatuisia.</p> <ul style="list-style-type: none"><li>• Tarkoituksenmukaisuutta voidaan pitää Suomen kehitysyhteistyön vahvuutena. Kestävydessä on eniten parantamisen varaa, ja sen lisäksi johdonmukaisuudessa ja vaikuttavuudessa.</li></ul>	<p><b>4.</b> Kahdenvälisen ja monenkeskisten hankkeiden osalta laatu todettiin kohtuulliseksi, vaikka joka viidennessä hankkeessa havaittiin puutteita. Erityisesti OECD:n kehitysapukomitean kestävyttä ja johdonmukaisuutta koskevien kriteerien osalta kaikilla kehitysyhteistyökumppaneilla on vielä huomattavasti parantamisen varaa.</p>	<p><b>2.1.</b> Ulkoministeriö kiinnittää enemmän huomiota kestävyyteen, johdonmukaisuuteen ja vaikuttavuuteen, läpileikkaaviin tavoitteisiin ja ihmisoikeusperustaiseen lähestymistapaan kokonaisvaltaisemman ja laadukkaamman kehitysyhteistyön toteuttamiseksi.</p>





LÖYDÖKSET	JOHTOPÄÄTÖKSET	SUOSITUKSET
<p><b>5.2 Lämpileikkaavat tavoitteet ja ihmisoikeusperustainen lähestymistapa.</b></p> <ul style="list-style-type: none"><li>Sukupuolten välisen tasa-arvon edistäminen nousee esiin yhtenä Suomen kehitysyhteistyön vahvuutena.</li><li>Evaluoititimet ottivat syrjimättömyyden, ilmastokestävyyden ja ihmisoikeusperustaisen lähestymistavan harvoin huomioon. Näin ollen hankkeiden laatua tältä osin ei voitu arvioida tässä meta-evaluoinnissa.</li></ul>	<p><b>5.</b> Vaikka sukupuolten välinen tasa-arvo oli otettu huomioon useammin kuin muut lämpileikkaavat tavoitteet ja ihmisoikeusperustainen lähestymistapa, Suomen kehityspolitiikan linjauksia ole ollut vielä sisällytetty hankkeisiin riittävän järjestelmällisesti</p>	<p><b>2.1.</b> Ulkoministeriö kiinnittää enemmän huomiota kestävyteen, johdonmukaisuuteen ja vaikuttavuuteen, lämpileikkaaviin tavoitteisiin ja ihmisoikeusperustaiseen lähestymistapaan kokonaisvaltaisemman ja laadukkaamman kehitysyhteistyön toteuttamiseksi.</p>
<p><b>5.4 Evaluointiraportteihin kirjatut opit ja suositukset</b></p> <ul style="list-style-type: none"><li>Noin kaksi kolmasosaa raporteista sisälsi suosituksia liittyen seurantaan ja arviointiin, johdonmukaisuuteen ja suunnitteluun.</li><li>Noin puolet raporteista antoivat suosituksia liittyen kestävyteen, hallintoon, rahoitukseen ja kapasiteetin kehittämiseen.</li></ul> <p><b>5.3 Erot ja trendit verrattuna edelliseen metaevaluointiin</b></p> <ul style="list-style-type: none"><li>Yleisesti ottaen hankkeiden laadussa ei ollut havaittavissa muutoksia edelliseen metaevaluointiin verrattuna.</li><li>Suorat vertailut eivät kuitenkaan olleet mahdollisia, koska meta-evaluoinneissa tarkasteltiin eri hankkeita ja otoksen rakenne oli erilainen sekä budjettien että sektoreiden suhteen.</li></ul>	<p><b>6.</b> Evaluojien laatimat suositukset viittasivat siihen, että Suomen rahoittamissa hankkeissa on vielä paljon parantamisen varaa. Seuranta- ja evaluointijärjestelmät sekä johdonmukaisuus, hallinnointi, kapasiteetin vahvistaminen ja rahoitusnäkökohdat eivät ole vielä toivotulla tasolla.</p>	<p><b>2.1.</b> KUIkoministeriö kiinnittää enemmän huomiota kestävyteen, johdonmukaisuuteen ja vaikuttavuuteen, lämpileikkaaviin tavoitteisiin ja ihmisoikeusperustaiseen lähestymistapaan kokonaisvaltaisemman ja laadukkaamman kehitysyhteistyön toteuttamiseksi.</p> <p><b>2.2.</b> Hankkeista ja ohjelmista vastaavat yksiköt huolehtivat, että evaluointiraportteihin kirjatut opit ja suositukset toteutetaan, erityisesti seurannan ja evaluoinnin, suunnittelun, kestävyden, kapasiteetin kehittämisen, rahoituksen ja johdonmukaisuuden osalta.</p>
<b>OSA 3. EVALUOINTIEN KÄYTTÖ JA HYÖDYLLISYYS</b>		
<p><b>6.3.1 Evaluointien yleinen hyödyllisyys</b></p> <ul style="list-style-type: none"><li>Keskeiset sidosryhmät pitivät evaluointeja hyödyllisinä tai erittäin hyödyllisinä. Varovaisempikin arvio hyödyllisyydestä, jossa otetaan huomioon kaikki mukana olleet evaluoinnit (eli myös ne, joista on puutteelliset tiedot), johtaa varsin myönteiseen tulokseen hyödyllisyydestä.</li></ul> <p><b>6.2 Hajautettujen evaluointien käyttö</b></p> <ul style="list-style-type: none"><li>Ulkoministeriön tiimien oppiminen ilmoitettiin yleisimmäksi evaluointien käyttötarkoitukseksi. Lisäksi evaluointeja käytettiin päätöksentekoon, suunnitteluun ja hallinnointiin liittyen meneillään oleviin hankkeisiin ja niiden jatkovaiheisiin.</li></ul>	<p><b>7.</b> Hajautetut evaluoinnit tukevat hyvin oppimista, päätöksentekoa, suunnittelua ja mukautuvaa johtamista ulkoministeriossa. Voidaan kuitenkin lisätä evaluointien käyttöä yksittäisiä hankkeita laajemmalle ja välttää resurssien käyttämistä evaluointeihin, joista on vain vähän hyötyä.</p>	<p><b>1.5.</b> Ulkoministeriön tulisi ottaa lähtökohdaksi tekijät, joiden on todettu parantavan evaluointien käyttöä ja hyödyllisyyttä (kuten laadukkaat tehtävänkuvaukset ja johdon vastineissa määriteltyjen toimien seuranta).</p>



LÖYDÖKSET	JOHTOPÄÄTÖKSET	SUOSITUKSET
<p><b>6.2 Hajautettujen evaluointien käyttö</b></p> <ul style="list-style-type: none"><li>Hanke- ja ohjelmaevaluointien raportteja oli käytetty tyypillisesti ulkoministeriön tiimien oppimiseen, päätöksentekoon, jatkotoimien suunnitteluun ja käynnissä olevien hankkeiden hallinnointiin.</li><li>Suomen kehitysyhteistyön kannalta strategisempia käyttötapoja mainittiin harvemmin, mutta kuitenkin niin usein, että niillä vaikuttaisi olevan merkitystä.</li></ul>	<p><b>8.</b> Evaluointiraporttien hyöty jää usein yksittäisen virkahenkilön tasolle. Tiedolla johtamisen puutteet haittaavat ulkoministeriön sisäistä oppimista.</p>	<p><b>1.3.</b> Ulkoministeriö parantaa tiedolla johtamista organisaation sisäisten oppimisprosessien edistämiseksi ja evaluointien hyödyllisyyden lisäämiseksi.</p>
<p><b>6.4.1 Evaluointien käytettävyyttä lisäävät ja haittaavat tekijät</b></p> <ul style="list-style-type: none"><li>Evaluointien hyödyllisyyttä lisääviä tekijöitä olivat sopivan evaluointityypin valinta (esim. väli- vai loppuevaluointi), asianmukaisten tehtäväkuvausten laatiminen, johdon vastineiden laatiminen, johdon vastineissa määriteltyjen toimenpiteiden seuranta, sekä evaluointitiimin temaattinen ja metodologinen osaaminen.</li><li>Haastatellut henkilöt listasivat evaluointien hyödyllisyyttä kuvaavia tekijöitä kaksi kertaa niin paljon kuin haittaavia tekijöitä (252 vs 116). Evaluointien hyödyllisyyttä haittaavien tekijöiden osalta ei noussut selkeästi esiin tyypillisiä, tilastollisesti merkittäviä seikkoja.</li></ul> <p><b>6.4.2 Haastateltujen henkilöiden antamat suositukset</b></p> <ul style="list-style-type: none"><li>Haastateltujen ulkoministeriön edustajien esittämät suositukset evaluointien hyödyllisyyden parantamiseksi koskivat lähinnä organisaation toimivuutta, evaluointien ajoitusta ja oikea-aikaisuutta sekä raporttien laatuun liittyviä näkökohtia.</li><li>Kehitysyhteistyön parissa toimivat henkilöt toivoivat tukea i) evaluointien laajuuden ja painopisteiden selkiyttämiseksi, ii) evaluointiprosessista oppimisen tehostamiseksi ja iii) evaluoinneista saatujen tulosten viestimiseksi laajemmin ulkoministeriössä.</li></ul>	<p><b>9.</b> Monet parannettavat asiat kuuluvat ulkoministeriön vastuun ja vaikutusmahdollisuuksien piiriin.</p>	<p><b>1.5.</b> Ulkoministeriön tulisi ottaa lähtökohdaksi tekijät, joiden on todettu parantavan evaluointien käyttöä ja hyödyllisyyttä (kuten laadukkaat tehtäväkuvaukset ja johdon vastineissa määriteltyjen toimien seuranta).</p>



LÖYDÖKSET	JOHTOPÄÄTÖKSET	SUOSITUKSET
<b>METAVALUOINNIN RAJOITTEET JA SUOSITUKSET TULEVIEN METAVALUOINTIEN PARANTAMISEKSI</b>		
<ul style="list-style-type: none"><li>• Metaevaluoinnin edustavuutta suhteessa Suomen koko kehitysyhteistyön kahdenvälisten ja monenkeskisten hankkeiden kokonaisuuteen ei pystytty määrittämään.</li><li>• Käyttöä ja hyödyllisyyttä koskevien tutkimustietojen laatuun vaikuttavat todennäköisesti muistivajeet ja valikoiva muistaminen. Avainhenkilöiden haastattelujen avulla näitä rajoitteita voitiin ainakin osittain lieventää. Toisaalta yksittäisten henkilöiden haastattelujen heikkous on niiden suppea edustavuus.</li></ul>		<p><b>3.1</b> Ulkoministeriö alkaa ylläpitää kattavaa luetteloa kehitysyhteistyöhankkeista ja evaluoinneista. Tämä helpottaa metaevaluoinnin otoksen edustavuuden analysointia. Järjestelmällinen luettelointi mahdollistaa myös evaluointitiimin resurssien säästämisen muihin tehtäviin tulevissa metaevaluoinneissa.</p>
<ul style="list-style-type: none"><li>• Trendien tarkastelu pidemmällä aikavälillä on mahdollista vain, jos samaa menetelmää sovelletaan tulevissa meta-evaluoinneissa.</li></ul>		<p><b>3.2.</b> Ulkoministeriö ottaakäyttöön evaluointeja koskevat digitaaliset palautelomakkeet henkilöstölle ja hankkeiden toteuttajille. Näin saadaan jatkuvaa tietoa evaluointien käytöstä ja hyödyllisyydestä sekä parannetaan analysoitavan datan laatua ja kattavuutta tulevia meta-evaluointeja varten.</p>
<ul style="list-style-type: none"><li>• Aiempaan metaevaluointiin verrattuna huomattavasti suurempi määrä aineistoa ja kyselyyn vastanneiden ja haastateltavien kartoittamisen ja tavoittamisen työläys rasittivat käytettävissä olevia resursseja valtavasti eivätkä mahdollistaneet käytettävissä olevan datan täysimääräistä hyödyntämistä.</li></ul>		<p><b>3.3.</b> EVA-11 varmistaa riittävät resurssit tuleviin meta-evaluointeihin ja aikaisemmin käytettyjen menetelmien toistamiseen. Siten saadaan paras mahdollinen aineisto tulevaa oppimista ja systemaattista tarkastelua varten.</p>



# Sammanfattning

## Syfte, omfattning och mål

Syftet **med** den sjunde metautvärderingen av decentraliserade utvärderingar är att underlätta lärande och ansvarsutkrävande med avseende på UM:s utvärderingsfunktion och att bidra till UM:s utvecklingspolitiska resultatrapport till riksdagen.

Metautvärderingen omfattar samtliga decentraliserade utvärderingar av finländska bi-, multi- och multibilaterala insatser och motsvarande uppdragsbeskrivningar som påbörjats sedan den tidigare metautvärderingen 2017 och som avslutats under 2020.

Meta-utvärderingen består av tre delar med följande **mål**:

4. Meta-utvärderingen för att bedöma metodologiska kvaliteten på utvärderingsrapporterna och motsvarande uppdragsbeskrivningarna.
5. Den sammanfattande analysen för att bedöma insatsernas övergripande kvalitet och sammanställa lärdomar och rekommendationer från utvärderingarna.
6. Bedömningen av utvärderingarnas användning och användbarhet från UM:s och genomförandeaktörers perspektiv.

Det **övergripande målet** är att förse UM med evidensbaserade slutsatser och rekommendationer för att (i) förbättra utvärderingarnas kvalitet, hanteringen av utvärderingar och utvärderingskapaciteten, (ii) förbättra det finländska bi-, multi- och multi-bilaterala utvecklingssamarbetet, och (iii) göra decentraliserade utvärderingar mer användbara. De **primära användarna** av meta-utvärderingen är UM:s personal på olika nivåer.

## Metodologi

I stort sett har utvärderingen använt samma metodologi som den tidigare metautvärderingen. Standardiserade verktyg användes för att bedöma (i) kvaliteten på 80 utvärderingsrapporter och motsvarande 70 uppdragsbeskrivningar och (ii) syntetisera kvaliteten på de insatser som berördes av de 72 utvärderingsrapporter som uppfyllde kvalitetskriterierna. En kvalitativ innehållsanalys gjordes för att ta hänsyn till lärdomar, rekommendationer, bakomliggande orsaker till utvärderarnas bedömning och resultat med avseende på jämställdhet. Tvärgående målsättningar och rättighetsperspektivet granskades genom kvantitativa sökordsanalyser. Utöver detta genomfördes 26 intervjuer med UM:s personal i Finland och på ambassader, samt en enkätundersökning. Från den senare erhöles 119 svar från UM:s personal och genomförandeaktörer (motsvarande 71% av urvalet och 38% av totalen), som bidrog till bedömningen av utvärderingarnas användning och användbarhet.



## Svar på utvärderingsfrågor

### Utvärderingsfråga 1: UM:s decentraliserade utvärderingsportfölj

Eftersom det saknas en sammanställning av bi-, multi- och multi-bilaterala insatser har det inte varit möjligt att bedöma utvärderingsrapporternas representativitet. En jämförelse med den tidigare meta-utvärderingen visar på både likheter (utvärderingens typ, uppdragsgivare, geografisk omfattning, Finlands ekonomiska bidrag) och olikheter (insats- och utvärderingsbudgetar, sektoriell fördelning).

### Utvärderingsfråga 2: Utvärderingsrapporternas kvalitet och tillförlitlighet, samt skillnader mellan undergrupper

Ungefär hälften av rapporterna är av tillfredsställande kvalitet. De övriga har brister och bland dessa är några otillräckliga. Brister förekommer i fråga om metodologi, evidensbaserade resultat, och hur OECD/DAC:s utvärderingskriterier (särskilt koherens, bärkraft och effektivitet) har hanterats. Generellt sätt är dock utvärderingsrapporterna någorlunda tillförlitliga. Kvaliteten på rapporterna skiljer sig inte åt mellan undergrupper (t.ex. utvärderingar som har beställts av UM och av andra aktörer).

### Utvärderingsfråga 3: Uppdragsbeskrivningarnas kvalitet och kopplingar mellan kvaliteten på uppdragsbeskrivningar och rapporter

Två av tre uppdragsbeskrivningar är av tillfredsställande kvalitet. Resten har brister. Regressionsanalysen visar att ju högre kvalitet på uppdragsbeskrivningar desto bättre utvärderingsresultat och rapporter.

### Utvärderingsfråga 4: Skillnader mot tidigare metautvärdering och brister i UM:s utvärderingskapacitet

En statisk jämförelse som baseras på exakt samma metod avslöjar inga skillnader. En dynamisk analys som tar hänsyn till ökade utvärderingskrav tyder dock på en viss försämring. Dessa resultat pekar på befintliga och sannolikt ökande brister i UM:s utvärderingskapacitet. De största bristerna i uppdragsbeskrivningarna är relaterad till metodologisk vägledning och systematisk beaktande av tvärgående målsättningar och rättighetsperspektivet.

### Utvärderingsfråga 5: Insatsernas kvalitet enligt OECD DAC kriterierna

Nästan alla bedömda insatser är av hög eller måttlig relevans. Ungefär två tredjedelar av de bedömda insatserna visar på i sämsta fall måttlig måluppfyllelse och kostnadseffektivitet. Nästan två tredjedelar av de bedömda insatserna har begränsad eller låg bärkraft, och nästan hälften av de bedömda insatserna har begränsad eller låg koherens och effekt. På grund av den begränsade tillgången på information om effekt och koherens är det dock inte säkert hur representativt utvärderingens resultat är på dessa områden.

### Utvärderingsfråga 6: Beaktandet av tvärgående målsättningar och rättighetsperspektivet

Jämställdhetsperspektivet, i den mån det har berörts av utvärderare, är delvis integrerat. Icke-diskriminering, klimathållbarhet och rättighetsperspektivet kunde inte bedömas eftersom de i de flesta fall inte berördes i utvärderarnas analyser.

### Utvärderingsfråga 7: Rekommendationer och lärdomar för att förbättra det finländska utvecklingssamarbetet

Ungefär två tredjedelar av rapporterna innehåller rekommendationer om resultatstyrning, koherens och planering, och ungefär hälften om hållbarhet, administration, ekonomiska aspekter och



kapacitetsutveckling. De viktigaste lärdomarna berör kapacitetsutveckling, planering, och koherens, men eftersom de bara förekommer i lite mindre än 20% av rapporterna betraktas de inte som representativa lärdomar.

#### **Utvärderingsfråga 8: Den övergripande kvaliteten på det finländska utvecklingssamarbetet; väsentliga styrkor och utmaningar**

Sammantaget är kvaliteten på bi-, multi- och multi-bilaterala insatser ganska bra: Ungefär 20% av insatserna är av hög kvalitet, ungefär 60 % av måttlig kvalitet, och cirka 20 % av begränsad kvalitet. Insatsernas relevans är relativt hög, vilket kan betraktas som en styrka. Störst utrymme för förbättringar återfinns i fråga om bärkraft, följt av koherens och effekt.

#### **Utvärderingsfråga 9: Skillnader i insatsers kvalitet jämfört med den tidigare meta-utvärderingen**

I stort sett kan ingen förändring från den tidigare meta-utvärderingen observeras. Direkta jämförelser är inte möjliga då de två utvärderingarna har tittat på olika insatser och urvalskriterier också har skilt sig åt.

#### **Utvärderingsfråga 10: Intressenters användning av utvärderingsrapporter**

Decentraliserade utvärderingsrapporter används ofta för lärande i team, beslutsfattande, planering av kommande samarbete och styrning av pågående insatser. De används mer sällan på strategisk nivå inom det finländska utvecklingssamarbetet.

#### **Utvärderingsfråga 11: Övergripande användbarhet, timing, rekommendationer och genomförande**

En stor majoritet av utvärderingarna bedöms som tillfredsställande eller mycket användbara. Tidpunkten för utvärderingarna, såväl som timingen av de insatser som de berör, bedöms överlag på ett positivt sätt. Detsamma gäller rekommendationernas relevans och realism, lärdomarna från utvärderingen och genomförandet av rekommendationer.

#### **Utvärderingsfråga 12: Främjande och hämmande faktorer och rekommendationer för att förbättra användbarheten**

Utvärderingars användbarhet underlättas av faktorer såsom val av tidpunkt för utvärderingen (halvtidsutvärderingar och slutgiltiga utvärderingar), uppdragsbeskrivningens kvalitet, i vilken mån utvärderingen genererar ett management respons och hur detta följs upp, samt utvärderingsteamets ämnesmässiga och metodmässiga kompetens. Inga utmärkande faktorer som hämmar användbarheten påträffades. De rekommendationer och lärdomar som fångades upp genom intervjuer berör framförallt organisatoriska aspekter, utvärderingarnas timing, samt rapportkvalitet. UM:s personal efterfrågar stöd för att (i) tydliggöra utvärderingarnas omfattning och inriktning, (ii) effektivisera lärandet och (iii) öka spridningen av utvärderingars resultat inom organisation, vilket sammantaget förväntas göra utvärderingar mer användbara.

#### **Utvärderingsfråga 13: Rekommendationer från meta-utvärderingen i syfte att förbättra utvärderingskvalitet**

Var tredje uppdragsbeskrivning har brister, vilket antyder att UM:s utvärderingskapacitet är begränsad. Åtgärder på detta område bör prioriteras av UM då det finns en nära koppling mellan kvaliteten på uppdragsbeskrivningar och utvärderingsrapporter (C2). På basis av denna slutsats görs följande rekommendation (R1.1): tillförsäkra att UM:s riktlinjer för utvärderingar följs för bättre uppdragsbeskrivningar. Dessutom ser UM ofta inte till att utvärderare följer riktlinjerna, vilket resulterar i att de flesta utvärderingsrapporter uppvisar metodologiska brister. Trots detta är



utvärderingsresultatet någorlunda tillförlitliga (C1). Därutöver, även om standarden för utvärderingar har höjts, riskerar kvalitetsnivån på UM:s utvärderingar att sjunka (C3). Vi rekommenderar därför UM (R1.2) att inrätta en kvalitetssäkringsprocess för att tillförsäkra att utvärderarna följer manualer, riktlinjer och uppdragsbeskrivningar. Dessutom rekommenderar vi (R1.4): fortsatt att utveckla utvärderingskapaciteten på särskilda områden för att uppfylla de ökade krav som ställs på utvärderingar (dvs. metodisk stringens, rapportering av högkvalitativa och evidensbaserade resultat, och tvärgående målsättningar).

#### **Utvärderingsfråga 14: Rekommendationer från metautvärderingen i syfte att förbättra insatsernas kvalitet**

Även om resultatuppfyllelsen generellt är god när det gäller bi-, multi- och multi-bilaterala insatser är var femte insats av begränsad kvalitet. Det finns en stor förbättringspotential särskilt vad gäller OECD/DAC-kriterier för hållbarhet och koherens (C4). Jämställdhet integreras oftare än andra tvärgående målsättningar och rättighetsperspektivet, och Finlands utvecklingspolitik återspeglas ännu inte fullt ut på insatsnivå (C5). Detta föranleder oss att rekommendera (R2.1): att större uppmärksamhet ges åt bärkraft, koherens och effekt, tvärgående målsättningar samt rättighetsperspektivet, för att tillförsäkra mer omfattande, högkvalitativa insatser. Den stora mängden rekommendationer och lärdomar som återfinns i utvärderingsrapporter pekar också på behovet av en bättre organisatorisk struktur för utvärderingar inom UM. Befintliga system för resultatuppföljning och utvärderingar fungerar dåligt och brister finns inom planering, bärkraft, koherens, ledning, kapacitetsutveckling och finansiella aspekter (C6). Vi rekommenderar därför dessutom (K2.2): omsatt lärdomar och viktiga rekommendationer från utvärderare i praktiken.

#### **Utvärderingsfråga 15: Rekommendationer från metautvärderingen i syfte att förbättra användning och användbarhet**

Utvärderingsrapporter används framförallt av enskilda individer. Avsaknaden på ett fungerande, organisations-omfattande kunskapshanteringsystem hämmar användandet av utvärderingar inom UM (C8). Därför rekommenderar vi (R1.3): förbättra kunskapshanteringen inom UM för att främja organisatoriskt lärande och förbättra användandet av utvärderingar. Även om decentraliserade utvärderingsrapporter ofta främjar lärande, beslutsfattande, planering och omställning inom UM kvarstår behovet av att öka användandet på organisationsnivå, och undvika kostnader för utvärderingar som inte används (C7). Dessutom kan UM, genom sin intressesfär, öka användbarheten av decentraliserade utvärderingar (C9). Detta föranleder att vi rekommenderar (R1.5): ta de främjande faktorer, såsom uttömmande uppdragsbeskrivningar, uppföljning av management respons, och val av rätt utvärderingstyp, som identifieras genom denna utvärdering som utgångspunkt för att öka användningen och användbarheten av utvärderingar.

#### **Utvärderingsfråga 16: Rekommendationer från metautvärderingen i syfte att förbättra meta-utvärderingar**

Utifrån de begränsningarna som vi stött på inom ramen för detta uppdrag har vi identifierat en sista uppsättning rekommendationer och föreslår (K3.1) gör en kartläggning av insatser och utvärderingar för att kunna bedöma hur representativt ett urval är och därmed spara på resurser i framtida meta-utvärderingar, (K3.2); Inför digitala feedback-enkäter om utvärderingar för UM:s personal och genomförandeaktörer för att kontinuerligt samla in uppgifter om användning och användbarhet, och för att förbättra datakvalitet och omfattning av framtida meta-utvärderingar; och (R3.3) säkerställ tillräckliga resurser för framtida metautvärderingar och återanvänd samma metod för bästa evidens-baserade resultat för framtida lärande och systematisk granskning.



# Nyckelresultat, slutsatser och rekommendationer

RESULTAT	SLUTSATSER	REKOMMENDATIONER
<b>DEL 1: KVALITETEN PÅ RAPPORTER OCH UPPDRAGSBESKRIVNINGAR</b>		
<b>4.2 Rapporternas kvalitet</b> <ul style="list-style-type: none"><li>Nästan hälften av rapporterna är av tillfredsställande kvalitet, och nästan hälften har brister</li><li>Kvaliteten på resultaten, underliggande metod, granskning av tvärgående målsättningar och rättighetsperspektivet, samt sammanfattningar är oftast i behov av förbättringar.</li></ul>	<b>C1</b> <p>UM tillförsäkrar inte att riktlinjer efterlevs. Av denna anledning är de flesta utvärderingsrapporterna baserade på bristfälliga metoder. Trots detta är resultaten någorlunda tillförlitliga.</p>	<b>R1.2</b> <p>Inrätta en kvalitetssäkringsprocess inom UM för att tillförsäkra att utvärderare följer manualer, riktlinjer och uppdragsbeskrivningar.</p> <b>R1.4</b> <p>Fortsätt att utveckla utvärderingskapaciteten hos intressenter inom och utanför UM på specifika områden för att uppfylla de ökade krav som ställs på utvärderingar, dvs i fråga om metod, rapportering av högkvalitativa, evidensbaserade resultat och tvärgående mål.</p>
<b>4.1.1 Kvaliteten på uppdragsbeskrivningarna</b> <ul style="list-style-type: none"><li>På ett generellt plan är kvaliteten på uppdragsbeskrivningar tillfredsställande, men en betydande del av de uppdragsbeskrivningar som granskats uppvisar brister.</li></ul> <b>4.1.2 Kopplingen mellan kvaliteten på uppdragsbeskrivningar och utvärderingsrapporter</b> <ul style="list-style-type: none"><li>Kvaliteten på uppdragsbeskrivningar är en kritisk faktor för kvaliteten på utvärderingars resultat.</li></ul>	<b>C2</b> <p>Var tredje uppdragsbeskrivning uppvisar brister, vilket antyder att utvärderingskapaciteten inom UM är fortsatt begränsad. Åtgärder på detta område bör prioriteras av UM då det finns en nära koppling mellan kvaliteten på uppdragsbeskrivningar och utvärderingsrapporter.</p>	<b>R1.1</b> <p>Tillförsäkra att UM:s riktlinjer för uppdragsbeskrivningar följs</p> <b>R1.4</b> <p>Fortsätt att utveckla utvärderingskapaciteten hos olika intressenter inom och utanför UM, på särskilda områden (enligt ovan)</p>





RESULTAT	SLUTSATSER	REKOMMENDATIONER
<p><b>4.3.1 Skillnader och trender i förhållande till den tidigare meta-utvärderingen</b></p> <ul style="list-style-type: none"><li>• En statisk jämförelse påvisar inga skillnader men den dynamiska analysen (som tar hänsyn till ökande krav på utvärderingar) tyder dock på en försämring</li><li>• Utvärderingsportföljen har dock förändrats, och försämringen kan delvis förklaras av att olika typer av utvärderingar av fonder och organisationer har inkluderats i den senaste meta-utvärderingen.</li></ul> <p><b>4.3.2 Brister avseende UM:s utvärderingskapacitet</b></p> <ul style="list-style-type: none"><li>• Kvalitetsbedömningen visade på brister inom och utanför UM. Ungefär en tredjedel av uppdragsbeskrivningarna och ungefär hälften av utvärderingsrapporterna har brister. En systematisk hantering av tvärgående målsättningar och rättighetsperspektivet saknas ofta i båda.</li><li>• En försämrade rapportkvalitet jämte ökade krav på utvärderingar pekar på ökande kapacitetsbrister.</li></ul>	<p><b>C3</b></p> <p>Då kraven på utvärderingar har höjts riskerar kvalitetsnivån på UM:s utvärderingar att sjunka</p>	<p><b>K1.2</b></p> <p>Inrätta en kvalitetssäkringsprocess inom UM (enligt ovan)</p> <p><b>R1.4</b></p> <p>Fortsätt att utveckla utvärderingskapaciteten på särskilda områden (enligt ovan)</p>
<b>DEL 2: INSATSKVALITET</b>		
<p><b>5.1 Övergripande kvalitet, styrkor och brister hos insatserna</b></p> <ul style="list-style-type: none"><li>• Sammantaget är kvaliteten på de bi-, multi- och multibilaterala insatser som granskats ganska hög: Ungefär 20 % av insatserna är av hög kvalitet, ungefär 60 % av måttlig kvalitet, och ungefär 20 % begränsad kvalitet.</li><li>• Relevans, följt av måluppfyllelse och kostnadseffektivitet, hör till the finländska utvecklingsarbetets starka sidor. Den största potentialen för förbättringar finns vad gäller bärkraft, följt av koherens och effekt.</li></ul>	<p><b>C4</b></p> <p>Även om resultatuppfyllelsen generellt är god när det gäller bi-, multi- och multi-bilaterala insatser är var femte insats av begränsad kvalitet. Det finns en stor förbättringspotential särskilt vad gäller OECD/DAC-kriterier för hållbarhet och koherens, och detta gäller samtliga intressenter</p>	<p><b>R2.1</b></p> <p>Fäst större uppmärksamhet vid bärkraft, koherens och effekt, tvärgående målsättningar samt rättighetsperspektivet, för att tillförsäkra mer omfattande, högkvalitativa insatser.</p>
<p><b>5.2 Beaktande av tvärgående målsättningar och rättighetsperspektivet</b></p> <ul style="list-style-type: none"><li>• Jämställdhetsperspektivet är en relativ styrka i det finländska utvecklingsarbetet.</li><li>• Icke-diskriminering, klimathållbarhet och rättighetsperspektivet beaktas sällan av utvärderare. Insatsernas kvalitet i detta avseende kan därför inte bedömas.</li></ul>	<p><b>C5</b></p> <p>Jämställdhet integreras i större utsträckning än andra tvärgående målsättningar och rättighetsperspektivet, och Finlands utvecklingspolitik återspeglas ännu inte full ut i bi-, multi- och multibilaterala insatser.</p>	<p><b>R2.1</b></p> <p>Fäst större uppmärksamhet vid bärkraft, koherens och effekt, tvärgående mål samt rättighetsperspektivet (enligt ovan)</p>



RESULTAT	SLUTSATSER	REKOMMENDATIONER
<p><b>5.4 Lärdomar och rekommendationer från utvärderare</b></p> <ul style="list-style-type: none"> <li>• Ungefär två tredjedelar av rapporterna innehåller rekommendationer om resultatuppföljning och utvärdering, koherens och planering.</li> <li>• Ungefär hälften av rapporterna innehåller rekommendationer om bärkraft, administration, ekonomiska aspekter och kapacitetsutveckling.</li> </ul> <p><b>5.3 Skillnader och trender i förhållande till den tidigare meta-utvärderingen</b></p> <ul style="list-style-type: none"> <li>• Ingen övergripande förändring in insatsernas kvalitet noteras sedan den tidigare meta-utvärderingen.</li> <li>• Direkta jämförelser på insatsnivå är dock inte möjliga eftersom det två meta-utvärderingarna granskade olika insatser och urvalet varierade med avseende på insatsbudget och sektoriell fördelning.</li> </ul>	<p><b>C6</b></p> <p>Den stora mängden rekommendationer och lärdomar pekar behovet att fortsätta förbättringar inom UM. Dåligt fungerande resultatuppföljnings- och utvärderingssystem och bristfällig planering, bärkraft, koherens, ledning, kapacitetsutveckling och ekonomiska aspekter är ännu inte på en acceptabel nivå.</p>	<p><b>R2.1</b></p> <p>Fäst större uppmärksamhet vid bärkraft, koherens och effekt, tvärgående målsättningar samt rättighetsperspektivet (enligt ovan).</p> <p><b>R2.2</b></p> <p>Applicera lärdomar och rekommendationer från utvärderare i praktiken, främst inom områdena för resultatuppföljning och utvärdering, planering, bärkraft, administration, kapacitetsutveckling, ekonomiska aspekter och koherens, för att höja kvaliteten på det finländska utvecklingssamarbetet.</p>
<p><b>DEL 3: UTVÄRDERINGSRAPPORTERS ANVÄNDNING OCH ANVÄNDBARHET</b></p>		
<p><b>6.3.1 Utvärderingarnas användbarhet i stort</b></p> <ul style="list-style-type: none"> <li>• De granskade utvärderingarna bedömdes som (mycket) användbara av intressenter. En mer konservativ uppskattning av användbarheten, inklusive i de fall där information saknas, ger fortfarande ett ganska positivt bild av användbarheten (lägre gränsvärde).</li> </ul> <p><b>6.2 Användning av decentraliserade utvärderingar</b></p> <ul style="list-style-type: none"> <li>• Team-baserat lärande var den vanligaste användningen, följt av beslutsfattande, planering, hantering och uppföljning av pågående insatser.</li> </ul>	<p><b>C7:</b> Decentraliserade utvärderingsrapporter främjar ofta lärande, beslutsfattande, planering och omställningar inom UM, men det finns utrymme för att öka användningen på organisationsnivå och undvika kostnader för utvärderingar med liten användbarhet.</p>	<p><b>R1.5</b> Ta de främjande faktorer, såsom adekvata uppdragsbeskrivningar och uppföljning av management respons, som identifieras genom denna utvärdering som utgångspunkt för att öka användningen och användbarheten av utvärderingar.</p>
<p><b>6.2 Användning av decentraliserade utvärderingar</b></p> <ul style="list-style-type: none"> <li>• Team-baserat lärande var den vanligaste användningen, följt av beslutsfattande, planering, hantering och uppföljning av pågående insatser.</li> <li>• Utvärderingar strategiska användning inom det finska utvecklingssamarbetet var mer ovanlig, men inte försumbar eftersom det bara gäller för UM i Finland och ambassadpersonal och inte för genomförandeaktörer.</li> </ul>	<p><b>C8:</b> Utvärderingsrapporters användbarhet är fortsatt individberoende. Avsaknaden på ett fungerande kunskapshanteringssystem hämmar det organisatoriska lärande inom UM.</p>	<p><b>R1.3</b> Förbättra kunskapshanteringen inom UM för att främja organisatoriskt lärande och förbättra användbarheten av utvärderingar.</p>



RESULTAT	SLUTSATSER	REKOMMENDATIONER
<p><b>6.4.1 Främjande och hämmande faktorer för användbarhet</b></p> <ul style="list-style-type: none"><li>• De flesta av de viktigaste faktorerna som främjar utvärderingars användbarhet är inom UM:s kontroll, t.ex. val av utvärderingstyp (halvtidsutvärdering kontra slutlig utvärdering), uppdragsbeskrivningars kvalitet, i vilken mån utvärderingar genererar ett management respons och hur detta följs upp. Utvärderingsteamets ämnesmässiga och metodologiska kompetens nämns också ofta.</li><li>• Främjande faktorer identifierades i dubbelt så hög utsträckning som hämmande faktorer (252 gentemot 116), och inga utmärkande faktorer som hämmande användbarheten kunde identifieras.</li></ul> <p><b>6.4.2 Intervjupersoners rekommendationer</b></p> <ul style="list-style-type: none"><li>• Intervjupersoners rekommendationer och lärdomar för att förbättra utvärderingarnas användbarhet berör framförallt organisatoriska aspekter, utvärderingarnas tidpunkt och timing, samt olika aspekter i fråga om rapportkvalitet.</li><li>• UM:s personal efterfrågar mer stöd för att tydliggöra utvärderingarnas omfattning och inriktning, effektivisera lärandet och öka spridningen av utvärderingars resultat inom organisationen, vilket sammantaget förväntas göra utvärderingar mer användbara.</li></ul>	<p><b>C9:</b> UM kan öka användbarheten av decentraliserade utvärderingar genom att utöva aktivt inflytande.</p>	<p><b>K1.5</b> Använd främjande faktorer som identifierats i denna meta-utvärdering (enligt ovan)</p>



RESULTAT	SLUTSATSER	REKOMMENDATIONER
<b>BEGRÄNSNINGAR OCH REKOMMENDATIONER FÖR FRAMTIDA META-UTVÄRDERINGAR</b>		
<ul style="list-style-type: none"><li>• Det går inte att bedöma hur representativt urvalet är för hela Finlands portfölj av bi-, multi- och multilaterala insatser.</li><li>• Enkätdata om användning och användbarhet kommer sannolikt att påverkas av minnesluckor och selektivt hågkomst. Sådana begränsningar kunde åtminstone delvis hanteras genom intervjuer, men detta påverkade förmågan att tillförsäkra ett representativt urval.</li><li>• Att granska förändringar över en längre tidsperiod är endast möjligt om samma metod tillämpas av framtida meta-utvärderingar.</li><li>• Den ökande mängden relevanta resultat från den tidigare meta-utvärderingen och det ineffektiva urvalet av enkätresponder och intervjupersoner var mycket betungande för utvärderingen och tillät inte ett fullt utnyttjande av tillgänglig data.</li></ul>		<p><b>R3.1</b> Gör en kartläggning av insatser och utvärderingar för att kunna bedöma hur representativt ett urval är och därmed spara resurser för framtida meta-utvärderingar.</p> <p><b>R3.2</b> Inför digitala feedback-enkäter om utvärderingar för UM:s personal i Finland och på ambassadnivå samt för genomförandeaktörer och utvärderare, för att kontinuerligt samla in uppgifter om användning och användbarhet, och för att förbättra datakvalitet och omfattning av framtida meta-utvärderingar.</p> <p><b>R3.3</b> Säkerställ tillräckliga resurser för framtida meta-utvärderingar och återanvänd samma metod för bästa evidens-baserade resultat för framtida lärande och systematisk granskning.</p>



# Summary

## Purpose, scope and objectives

The **purpose** of the seventh metaevaluation of decentralised evaluations is to facilitate learning and accountability with respect to MFA's evaluation function and to provide insights for MFA's results reporting to the parliament.

All decentralised evaluation reports on Finnish bi-, multi- and multi-bilateral interventions and corresponding ToRs completed since the previous metaevaluation in 2017 and lasting until the end of 2020 fall under the **scope** of the metaevaluation.

The metaevaluation consists of three components with the following **objectives**:

7. The metaevaluation to assess the methodological quality of the evaluation reports and corresponding ToRs.
8. The summative analysis to assess the overall quality of the interventions and synthesise lessons learnt and recommendations provided by the evaluators.
9. And the assessment of actual use and usefulness to obtain MFA commissioners' and intervention implementers' perspectives.

The **overarching goal** is delivering evidence-based conclusions and recommendations to MFA for (i) enhancing the quality of evaluations, evaluation management practices and evaluation capacity development, (ii) improving Finnish bi-, multi- and multi-bilateral development cooperation, and (iii) increasing the usefulness of decentralised evaluations. **Primary users** of the metaevaluation are MFA staff at different levels.

## Methodology and data sources

Overall, the methodology of the previous metaevaluation has been replicated. Standardised assessments tools were applied to assess (i) the quality of 80 evaluation reports and 70 corresponding ToRs and (ii) to synthesise intervention quality for 72 reliable evaluation reports which passed the quality assessment. Lessons learnt, recommendations, underlying reasons for evaluators' assessment and gender results were considered in qualitative content analysis. To gain insights on consideration of CCOs and HRBA, quantitative keyword analyses were performed. As a new feature, 26 key informant interviews were conducted with MFA staff at headquarters and embassy level, and 119 survey responses were obtained by MFA staff and intervention implementers (case coverage: 71%, sample coverage rate 38%) to assess the use and usefulness of the evaluations.



## Answers to evaluation questions

### **EQ1 on MFA's decentralised evaluation portfolio**

A lack of an inventory of bi-, multi- and multi-bilateral interventions does not allow to assess the representativeness of the evaluation reports. A comparison with the previous metaevaluation reveals similarities (nature of the evaluation, commissioner, geographical scope, Finland's budget contributions) and differences (intervention and evaluation budgets, sectorial distribution).

### **EQ2 on the quality and reliability of evaluation reports and differences among sub-groups**

About half of the reports are of satisfactory quality, the rest with a need for improvement and few are inadequate. Limitations regarding methodological rigour, provision of high-quality, evidence-based findings, appropriate capturing of OECD DAC criteria (particularly coherence, sustainability, and efficiency) exist. Yet, overall evaluation reports are somewhat reliable. The quality of reports, however, does not differ among sub-groups (e.g., MFA vs non-MFA commissioned).

### **EQ3 on the quality of the ToRs and linkages between ToRs and report quality**

Two out of three ToRs are of satisfactory quality, and the rest displays the need for improvement. Regression analysis shows: Better ToR yield to a better quality of evaluation findings and thus better reports.

### **EQ4 on differences to previous metaevaluation and gaps regarding MFA's evaluation capacity?**

A static comparison replicating exactly the same methodology reveals no differences; a dynamic analysis anticipating increased evaluation standards suggests some deterioration. Both findings identify partially existing and likely to increase evaluation capacity gaps inside MFA. Major gaps identified from the ToRs are related to providing sound methodological guidance and systematic consideration of cross-cutting objectives (CCOs) and human-rights based approach (HRBA).

### **EQ5 on the interventions' quality according to OECD DAC criteria**

Almost all interventions assessed are of high or moderate relevance. Roughly two-thirds of the interventions assessed are at least of moderate effectiveness and efficiency. Nearly two-thirds of the interventions assessed reveal limited or low sustainability, and nearly half of the interventions assessed revealed limited or low coherence and impact. However, data availability on impact and coherence poses a limitation on sample representativeness of results.

### **EQ6 on the consideration of CCO and HRBA**

Gender equality, if considered by evaluators, is partly mainstreamed. Non-discrimination, climate sustainability, and HRBA could not be assessed as evaluators did not mostly recognise them in their analyses.

### **EQ7 on major recommendations and lessons learnt to improve Finnish development cooperation**

About two-thirds of the reports contain recommendations on M&E, coherence, and planning, about half on sustainability, management, financial aspects, and capacity development. Major lessons learnt were related to capacity development, planning, coherence, but appearing in only a bit less than 20% of the reports, they are not regarded as typical lessons.



### **EQ8 on the overall quality of Finnish development cooperation; major strengths and challenges**

Overall, the quality of bi-, multi- and multi-bilateral interventions is quite good: About 20% of the interventions possess high, roughly 60% moderate and around 20% limited quality. Relevance can be considered a strength. Sustainability leaves the greatest room for improvement, followed by coherence and impact.

### **EQ9 on key differences of intervention quality to the previous metaevaluation**

At large, no change from the previous metaevaluation can be observed. Direct comparisons are not possible as both metaevaluations looked at different interventions and the sample composition varied.

### **EQ10 on the use of evaluation reports by key stakeholders**

Decentralised evaluation reports are often used for learning in teams, decision making, planning of follow-up, and management of ongoing interventions. More strategic usages for Finnish development cooperation were less frequently mentioned but are not negligible.

### **EQ11 on the overall usefulness, timing, timeliness, recommendations, and implementation**

A large majority of the evaluations is assessed as satisfactory or very useful. Timing of the evaluations, as well as timeliness of formal and informal results delivery, are overall positively assessed. The same holds true for the relevance of recommendations, the extent of recommendations being realistic, the learnings from the evaluation, and the implementation of recommendations.

### **EQ12 on typical facilitating and hampering factors and recommendations to improve the usefulness**

Typical factors facilitating the usefulness of evaluations are selecting the adequate nature of the evaluation (mid-term vs final), setting-up appropriate ToRs, following up on the management responses, producing management responses, and technical and methodological competence of the evaluation team. Typical factors hampering the usefulness could not be identified. Interviewees' recommendations and lessons to improve evaluations' usefulness comprise mainly organisational aspects, timing and timeliness of the evaluations, and aspects of report quality. Support to enhance (i) the clarity on the scope and focus of evaluations, (ii) the efficiency of learning and (iii) the dissemination of the results at the institutional level are desired by MFA staff and expected to enhance usefulness.

### **EQ13 on recommendations by the metaevaluation team to improve evaluation quality**

Every third ToR reveals needs for improvement; thus, evaluation capacity inside MFA is still constrained. This marks an important action arena for MFA as sound ToRs leverage the quality of evaluation reports (C2). From this conclusion, we derived a recommendation (R1.1): enforce compliance with guidelines inside MFA for better ToRs. Moreover, MFA does not often enforce that evaluators adhere to its guidelines; most evaluation reports are therefore grounded on weak methodologies. Nevertheless, findings are somewhat reliable (C1). Further, acknowledging increased evaluation standards, MFA is endangered to fall behind its previously attained level of quality (C3). Taking this together, we recommend (R1.2) establish a quality assurance process inside MFA to enforce evaluators' compliance with manual, guidelines, and ToRs. Additionally, we recommend (R1.4): continue further evaluation capacity development on focused topics to adapt to increased evaluation quality standards (i.e., methodological rigour, provision of high-quality, evidence-based findings, and CCOs).



#### **EQ14 on recommendations by the metaevaluation team to improve interventions' quality**

Regarding bi-, multi-, and multi-bilateral interventions, MFA is further on track, yet every fifth intervention is of limited quality. Particularly on the OECD DAC criteria of sustainability and coherence, considerable room for improvement persists for all developmental partners (C4). While gender is more often mainstreamed than other CCOs and HRBA, Finland's Development Policy is also not yet entirely reflected in its interventions (C5). This yields us to recommend (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. On a different note, the vast body of recommendations and lessons learnt by evaluators points to continuing construction sites for MFA. Ill-functioning M&E systems and weak planning, sustainability, coherence, management, capacity development, and financial aspects are not yet at the desired level (C6). We therefore further recommend (R2.2): translate lessons learnt and major recommendations provided by evaluators into action.

#### **EQ15 on recommendations by the metaevaluation team to improve use and usefulness**

The usefulness of evaluation reports often remains at an individual level. A lack of a functioning institutional knowledge management system hampers exploitation for organisational learning inside MFA (C8). Thus, we recommend (R1.3): Improve knowledge management inside MFA to foster organisational learning and enhance the usefulness of evaluations. Furthermore, decentralised evaluation reports often support learning, decision making, planning and adaptive management inside MFA, but room to enhance usage beyond individual levels and to avoid spending on evaluations of little use remains (C7). In addition, MFA can enhance the usefulness of its decentralised evaluations by actively making use of its sphere of influence. (C9). This yields us to recommend (R1.5): use facilitating factors identified in this metaevaluation like appropriate ToRs, follow-up of management response and selection of the right evaluation type as an entry point to enhance use and usefulness of evaluations.

#### **EQ16 on recommendations by the metaevaluation team to improve metaevaluations**

Form the limitations faced in this assignment, we draw a last set of recommendations and suggest (R3.1) Set-up an inventory of interventions and evaluations to assess sample representativeness and save resources for future metaevaluations (R3.2); introduce digitised feedback sheets on evaluations for MFA staff and intervention implementers to gain continuous insights on use and usefulness and enhance data quality and coverage for future metaevaluations; and (R3.3) ensure sufficient resourcing of future metaevaluations and replicating the methodology to gain the best evidence for future learning and systematic review.





# Key Findings, Conclusions and Recommendations

FINDINGS	CONCLUSIONS	RECOMMENDATIONS
<b>COMPONENT 1 ON REPORT AND TOR QUALITY</b>		
<p><b>4.2 Quality of reports</b></p> <ul style="list-style-type: none"> <li>Nearly half of the reports are of satisfactory quality, nearly another half needs improvement.</li> <li>The quality of findings, underlying methodology, consideration of CCOs and HRBA, and executive summaries are mostly in need of improvement.</li> </ul>	<p><b>C1</b></p> <p>MFA does often not enforce compliance with its guidelines. Therefore, most evaluation reports are grounded on weak methodologies; nevertheless, findings are somewhat reliable.</p>	<p><b>R1.2</b></p> <p>Establish a quality assurance process inside MFA to enforce evaluators' compliance with manual, guidelines, and ToRs outside MFA</p> <p><b>R1.4</b></p> <p>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.</p>
<p><b>4.1.1 Quality of ToRs</b></p> <ul style="list-style-type: none"> <li>The overall quality of ToRs is satisfactory, but a considerable part of ToRs discloses a need for improvement.</li> <li>4.1.2 Linkage quality of ToRs and quality of reports</li> <li>The quality of ToRs is a driver of the quality of evidence.</li> </ul>	<p><b>C2</b></p> <p>Every third ToR reveals needs for improvement; thus, evaluation capacity inside MFA is still constrained. This marks an important action area for MFA as sound ToRs leverage the quality of evaluation findings.</p>	<p><b>R1.1</b></p> <p>Enforce compliance with guidelines inside MFA for better ToRs</p> <p><b>R1.4</b></p> <p>Continue further evaluation capacity development for stakeholders inside and outside MFA on focused topics (as specified above)</p>



FINDINGS	CONCLUSIONS	RECOMMENDATIONS
<p><b>4.3.1 Differences and trends to the previous metaevaluation</b></p> <ul style="list-style-type: none"> <li>Regarding report quality, a static comparison reveals no differences to the previous metaevaluation, while a dynamic comparison (taking increased evaluation standards into account) unveils deterioration.</li> <li>However, the evaluation portfolio has changed, and the inclusion of conceptually different evaluations of funds and organisations into the recent metaevaluation may also partially explain deterioration.</li> </ul> <p><b>4.3.2 Gaps regarding MFA's evaluation capacity</b></p> <ul style="list-style-type: none"> <li>The quality assessment unveiled capacity gaps inside and outside MFA. About one-third of the ToRs and roughly half of the evaluation reports need improvement, while systematic consideration of CCOs and HRBA is mostly missing in both.</li> <li>Deterioration of report quality when anticipating increased quality standards provides hints for increasing capacity gaps.</li> </ul>	<p><b>C3</b></p> <p>Acknowledging increased evaluation standards, there is a risk that MFA may fall behind its previously attained level of quality.</p>	<p><b>R1.2</b> Establish a quality assurance process inside MFA (as specified above)</p> <p><b>R1.4</b> Continue further evaluation capacity development on focused topics (as specified above)</p>
<b>COMPONENT 2 ON INTERVENTION QUALITY</b>		
<p><b>5.1 Overall quality, strengths, and weaknesses of the interventions</b></p> <ul style="list-style-type: none"> <li>Overall, the quality of bi-, multi- and multi-bilateral interventions under consideration is quite good: About 20% of the interventions possess high, roughly 60% moderate and around 20% limited quality.</li> <li>While relevance is a strength of Finnish development cooperation, followed by effectiveness and efficiency, sustainability leaves the greatest room for improvement, followed by coherence and impact.</li> </ul>	<p><b>C4</b></p> <p>Regarding bi-, multi-, and multi-bilateral interventions, MFA is further on track, yet every fifth intervention is of limited quality. Particularly on the OECD DAC criteria sustainability and coherence, considerable room for improvement persists for all developmental partners.</p>	<p><b>R2.1</b></p> <p>Pay stronger attention to sustainability, coherence, and impact, as well as to CCOs and HRBA to work towards more comprehensive, high-quality interventions.</p>
<p><b>5.2 Consideration of CCOs and HRBA</b></p> <ul style="list-style-type: none"> <li>Gender equality emerges as a relative strength of Finnish development cooperation.</li> <li>Non-discrimination, climate sustainability and HRBA are rarely considered by evaluators. Thus, interventions' quality in this regard is not assessable.</li> </ul>	<p><b>C5</b></p> <p>Despite gender being more often mainstreamed than other CCOs and HRBA, Finland's Development Policy is not yet entirely reflected in its bi-, multi-, and multi-bilateral interventions.</p>	<p><b>R2.1</b></p> <p>Pay stronger attention to sustainability, coherence, and impact, as well as to CCOs and HRBA (as above)</p>



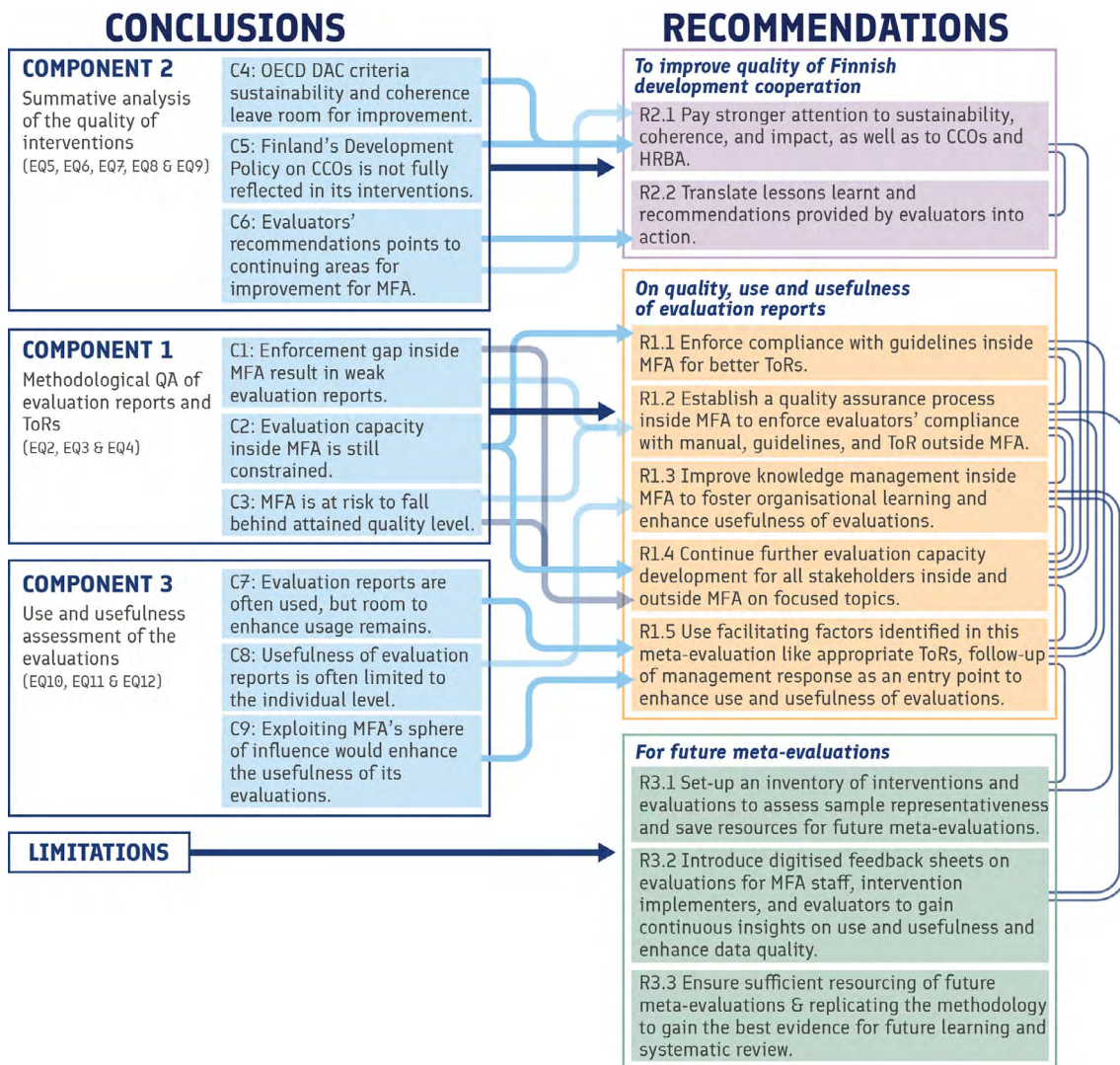
FINDINGS	CONCLUSIONS	RECOMMENDATIONS
<p><b>5.4 Lessons learnt and recommendations by evaluators</b></p> <ul style="list-style-type: none"> <li>About two-thirds of the reports contain recommendations on M&amp;E, coherence, and planning.</li> <li>About half of the reports provide recommendations on sustainability, management, financial aspects, and capacity development.</li> </ul> <p><b>5.3 Differences and trends to the previous metaevaluation</b></p> <ul style="list-style-type: none"> <li>The overall quality of the interventions did not change at large from the previous metaevaluation.</li> <li>However, direct comparisons at the intervention level are not possible as both metaevaluations looked at different interventions and the sample composition varied with respect to intervention budgets and sectorial distribution.</li> </ul>	<p><b>C6</b></p> <p>The vast body of recommendations and lessons learnt points to continuing areas for improvement for MFA. Ill-functioning M&amp;E systems and weak planning, sustainability, coherence, management, capacity development, and financial aspects are not yet at the desired level.</p>	<p><b>R2.1</b></p> <p>Pay stronger attention to sustainability, coherence, and impact, as well as to CCOs and HRBA (as above)</p> <p><b>R2.2</b></p> <p>Translate lessons learnt and recommendations provided by evaluators into action, mainly in the fields of M&amp;E, planning, sustainability, management, capacity development, financial aspects, and coherence to uplift the quality of Finnish development cooperation.</p>
<b>COMPONENT 3 ON USE AND USEFULNESS OF EVALUATION REPORTS</b>		
<p><b>6.3.1 Overall usefulness of the evaluations</b></p> <ul style="list-style-type: none"> <li>The surveyed evaluations were assessed as (very) useful by key stakeholders. A more conservative estimation of the usefulness, including all cases with missing data, still comes to a quite positive result on overall usefulness (lower-level boundary).</li> </ul> <p><b>6.2 Use of decentralised evaluations</b></p> <ul style="list-style-type: none"> <li>Learning in teams was the most frequent use, followed by decision making, planning, and management of ongoing and follow-up interventions.</li> </ul>	<p><b>C7:</b> Decentralised evaluation reports often support learning, decision making, planning and adaptive management inside MFA, but room to enhance usage beyond individual levels and to avoid spending on evaluations of little usefulness remains.</p>	<p><b>R1.5</b> Use facilitating factors identified in this metaevaluation like appropriate ToRs follow-up of management response as an entry point to enhance use and usefulness of evaluations</p>
<p><b>6.2 Use of decentralised evaluations</b></p> <ul style="list-style-type: none"> <li>Learning in teams was the most frequent use, followed by decision making, planning, and management of ongoing and follow-up interventions.</li> <li>More strategic usages for Finnish development cooperation were less frequent, but this is not negligible as only applicable for MFA headquarters and embassy staff and not for implementers.</li> </ul>	<p><b>C8:</b> Usefulness of evaluation reports often remains at an individual level. A lack of a functioning institutional knowledge management system hampers exploitation for organisational learning inside MFA.</p>	<p><b>R1.3</b> Improve knowledge management inside the MFA to foster organisational learning and enhance the usefulness of evaluations</p>



FINDINGS	CONCLUSIONS	RECOMMENDATIONS
<p><b>6.4.1 Facilitating and hampering factors for usefulness</b></p> <ul style="list-style-type: none"> <li>• Most of the main facilitating factors are under MFA's sphere of influence, for example, selecting the adequate nature of the evaluation (mid-term vs final), setting-up appropriate ToRs, producing of and follow-up on the management responses. Also, the technical and methodological competence of the evaluation team were frequently mentioned.</li> <li>• Facilitating factors were twice as much identified as hampering factors (252 vs 116), and typical factors hampering the usefulness could not be identified.</li> </ul> <p><b>6.4.2 Interviewees' recommendations</b></p> <ul style="list-style-type: none"> <li>• Interviewees' recommendations and lessons to improve evaluations' usefulness comprise mainly organisational aspects, timing and timeliness of the evaluations, and aspects of report quality.</li> <li>• Support to enhance the clarity on the scope and focus of the evaluation, efficiency of learning and results dissemination at the institutional level is desired by the MFA staff and expected to enhance usefulness.</li> </ul>	<p><b>C9:</b> MFA can enhance the usefulness of its decentralised evaluations by actively making use of its sphere of influence.</p>	<p><b>R1.5</b> Use facilitating factors identified in this metaevaluation (as above)</p>
LIMITATIONS AND DERIVED RECOMMENDATIONS FOR FUTURE METAEVALUATIONS		
<ul style="list-style-type: none"> <li>• The representativeness of the sample for the whole population of Finland's bi-, multi-, and multi-bilateral interventions cannot be assessed.</li> <li>• The quality of survey data on use and usefulness is likely to be affected by memory gaps and selective recall. Conducting key informant interviews allowed coping with such limitations at least partially but compromised on the representativeness</li> <li>• Inquiring trends over a longer period of time is only possible if the same methodology is applied for future metaevaluations.</li> <li>• The growing body of relevant evidence from the previous metaevaluation and resource inefficient identification of survey respondents and interviewees stressed the available resources tremendously and did not allow full exploitation of available data.</li> </ul>		<p><b>R3.1</b> Set up an inventory of interventions and evaluations to assess sample representativeness and save resources for future metaevaluations.</p> <p><b>R3.2</b> Introduce digitised feedback sheets on evaluations for the MFA staff and headquarters and embassy level, intervention implementers, and evaluators to gain continuous insights on use and usefulness and enhance data quality and coverage for future metaevaluations.</p> <p><b>R3.3</b> Ensure sufficient resourcing of future metaevaluations and replicating the methodology to gain the best evidence for future learning and systematic review.</p>



Figure 1: Conclusions and recommendations at a glance



Source: Evaluation team



# 1. Introduction

## 1.1. Scope, purpose and objectives of the metaevaluation

The Ministry for Foreign Affairs of Finland (MFA) understands metaevaluations as a valuable tool to assess the reliability of evaluation reports, to systematically aggregate evaluation findings and to explore issues and lessons learnt from a wide range of different interventions. The **rationale** behind metaevaluations is to facilitate learning and accountability with respect to MFA's evaluation function. Thus, they lay the foundation to further enhance evaluation capacities, improve interventions, and increase transparency towards partner countries, development partners, parliamentarians, academia, media, and the public.

All decentralized evaluation reports and corresponding ToRs conducted since the last metaevaluation and until 2020 were subject of this metaevaluation.

The Development Evaluation Unit (EVA-11) is an independent operational entity of MFA which reports directly to the Under-Secretary of State. Since 2007, EVA-11 has commissioned six metaevaluations of decentralised programme and project evaluations of Finnish development cooperation. Decentralised evaluations cover mid-term reviews, mid-term evaluations and final evaluations of single bilateral, multilateral, or multi-bi projects or programmes commissioned by different regional or thematic MFA units, embassies, or multilateral partners.

All decentralised evaluation reports and corresponding ToRs completed between 2018 and 2020 are within the **scope** of this assignment. Furthermore, reports from 2017 which have not been part of the predecessor metaevaluation have been added.

Comprehensive and strategic evaluations referring to entire policies, country strategies, specific themes or sectors are, according to MFA's evaluation norm 1/2015 (MFA, 2015a), centralised evaluations directly managed by EVA-11; these evaluations are not part of this assignment.

For the first time in MFA's tradition of metaevaluations, perspectives of key stakeholders, i.e., MFA commissioners, Finnish embassies, and intervention implementers, are also taken into consideration. Thus, beyond the desk-based analysis of decentralised evaluation reports, primary data collection on the actual use of decentralised evaluations and their usefulness also falls under the scope of this metaevaluation.

The **purpose** of this assignment is to continue and amend the series of metaevaluation (i) to further learn from successes, setbacks, use and usefulness of decentralised evaluations, (ii) to figure out which developments since the previous analysis took place, and (iii) to see whether the quality of evaluations and development results has further grown. In addition, the metaevaluation is expected (iv) to provide ideas on how to further take advantage of the potential of decentralised evaluations, (v) to exploit findings from a comparison with the predecessor metaevaluation results and (vi) to generate aggregated results to inform the drafting of the Development Policy Results Report 2022 for the parliament.



**Primary users** of the metaevaluation consequently are MFA staff responsible for (i) design and implementation of development projects and programmes, (ii) commissioning evaluations, (iii) development policy, (iv) evaluation capacity development, and (v) drafting the Development Policy Results Report 2022. Secondary users comprise experts conducting evaluations, CSOs and other development partners.

The metaevaluation consists of **three components with the following objectives:**

1. The **metaevaluation of decentralised evaluations** from 2017 to 2020 to **assess the methodological quality** of the **evaluation reports** and corresponding **ToRs**.
2. The **meta-analysis of overall quality and results** of development cooperation based on decentralised evaluations from 2017 to 2020 to extract and to **synthesise reliable evaluation results, recommendations, and issues on Finland's development cooperation** as well as to provide information for MFA's results reporting.
3. The **assessment of actual use and usefulness** of decentralised evaluation reports from 2017 to 2020 and the underlying evaluation processes based on new data collection instruments to **obtain MFA commissioners' and intervention implementers' perspectives**.

**Overarching goal is contributing to enhanced evaluation quality and usefulness and to improved development cooperation.**

The **overarching goal of this metaevaluation** is **delivering concise conclusions and recommendations** to further support MFA in (i) **enhancing the quality of evaluations**, evaluation management practices and overall evaluation capacity development, (ii) **improving its development cooperation** based on aggregated insights from decentralised evaluations, and (iii) **increasing the usefulness** of decentralised evaluations.

## 1.2. Evaluation questions

According to the ToR, MFA's evaluation questions are structured along the three components and overarching recommendations. **Component 1**, the **metaevaluation** itself, is to assess the state of evaluation capacity by answering the following questions:

1. How can MFA's decentralised **evaluation portfolio** be described (e.g., by commissioner, budgets, sector and regional distribution, evaluation type)?
2. How is the **quality and reliability of decentralised evaluation reports**? Is there a difference between the quality of MFA-commissioned evaluation reports and the quality of evaluation reports that are commissioned by MFA's partners? Are there other **differences** between relevant sub-groups (e.g., final vs mid-term evaluations, individual/independent evaluator(s) vs consulting firm/institutes) if applicable?
3. How is the **quality of the ToRs** of decentralised evaluations? Are there **systematic patterns** regarding the **quality of ToRs** and the **quality of corresponding evaluation reports**?
4. What are key differences and possible trends compared to the **previous metaevaluation**? Are there **gaps regarding MFA's evaluation capacity**?



**Component 2**, the **meta-analysis** on projects and programmes of Finnish development cooperation, comprises the following questions:

5. What can be said about the **quality of Finnish development cooperation** according to the **OECD DAC criteria** based on reliable decentralised evaluation reports (i.e., relevance, coherence, effectiveness, efficiency, impact, sustainability)?
6. What can be said about the **consideration of cross-cutting objectives**, the human rights-based approach (**HRBA**) and **policy priority areas** in Finnish development cooperation based on reliable decentralised evaluation reports?
7. What are the **major recommendations and lessons learnt to improve Finnish development cooperation** emerging from reliable decentralised evaluation reports?
8. What can be said about the **overall quality of Finnish development cooperation** based on reliable decentralised evaluation reports? What are the **major strengths and challenges** emerging from reliable decentralised evaluation reports? Are there **typical factors for success or challenges**?
9. What are **key differences and possible trends** compared to the overall results of the **previous metaevaluation**?

And **component 3** consists of the following evaluation questions to assess **the use and the usefulness** of decentralised evaluation reports:

10. How are decentralised **evaluations reports used by key stakeholders** (i.e., MFA commissioners, embassies, implementers)?
11. How **useful** are decentralised evaluations (i.e., **reports and processes**) assessed by key stakeholders? To what extent are decentralised evaluations **timely**? To what extent do decentralised evaluations provide **relevant and realistic recommendations**? To what extent have recommendations been **implemented**?
12. Are there **typical factors facilitating or hampering the usefulness** of decentralised evaluations (e.g., report quality, timing)? What are **major recommendations and lessons learnt to improve the use and usefulness** of decentralised evaluations based on the **perspectives of key stakeholders**?

Finally, the MFA expects **overall recommendations from the assignment**, which are structured by the following evaluation questions:

13. What are the **recommendations to improve the quality of MFA's decentralised evaluation reports** (and processes, if applicable)?
14. What are the **recommendations to improve the quality of Finnish development cooperation**?
15. What are the **recommendations to improve the use and usefulness** of decentralised evaluations?
16. What are the **recommendations to improve the methodology of metaevaluations** in the future?





### 1.3. Previous metaevaluation and MFA's management response

The results of this metaevaluation are put into perspective to the predecessor metaevaluation. Given the MFA's commitment to introduce standardised assessment tools during the previous metaevaluation, a comparison with earlier results on report quality and an assessment against the evaluation criteria developed by the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD DAC) is possible. Thus, paving the way to further enhance the long-term utility of metaevaluations. Earlier metaevaluations applied a completely different assessment methodology which does, unfortunately, not allow a systematic comparison.

**Previous metaevaluation revealed that two-thirds of the evaluation reports and about 60% of ToRs were of satisfactory quality. About 70% of the interventions were of moderate or better quality.**

The **previous metaevaluation showed**, for 51 decentralised evaluation reports from 2015 to 2017, that about **two-thirds of the reports assessed were of satisfactory quality**; however, often based on weak methodologies and little discussion on intervention logics and their underlying assumptions (Silvestrini et al., 2017). The overall **quality of ToRs was assessed as satisfactory for 60%** of the cases with limitations regarding methodological guidance, quality assurance and cross-cutting objectives. Not surprisingly, analyses revealed that **higher quality of ToR is associated with a higher quality of decentralised evaluation reports**. Or put differently, closing evaluation capacity gaps at MFA was expected to uplift the quality of decentralised evaluations. Thus, major recommendations centred around improving the evaluation manual substantially, enhancing knowledge of evaluation methodologies and practices, improving quality assurance, and ensuring methodological expertise when recruiting evaluators. It was recommended to consider improving existing structures with respect to knowledge management systems and better coordination with EVA-11.

Regarding the quality of Finnish development cooperation, the previous metaevaluation of decentralised project and programme evaluations came to the result that **70% of the interventions possess a moderate or better quality**. Interventions have been assessed as strong regarding their relevance, while sustainability was identified as the greatest challenge. Areas for improvement were often related to the intervention fields of planning, scope, management, capacity, and sustainability. Given a lack of an inventory of all interventions eligible for decentralised evaluation, a conclusion of the adequacy of MFA's decentralised evaluation portfolio was not possible. Thus, a key recommendation was to set up such a system.

Overall, recommendations were well received by MFA. Based on **MFA's management response**, a bundle of measurements was introduced to facilitate their implementation (MFA, 2021c). Milestones of this process comprise **comprehensive updates of the evaluation manual**, including a web-based version with video content, conducting evaluation training organised by EVA-11, providing support to regional departments and units carrying out decentralised evaluations under the Evaluation Management Service (EMS) framework agreement, **improving guidelines**, ToRs and recruitment of evaluators with respect to cross-cutting objectives, establishing the role of a critical friend in decentralised evaluation processes, **revise templates and writing instructions**, developing self-evaluation modalities, and creating an EVA-11 internal web-based workspace for improved knowledge management. **The last implementation monitoring report of MFA's management response** took place in **June 2021**.



## 1.4. Structure of this report

The remainder of the report is structured as follows: After this introduction (1), chapter 2 displays the general approach and the underlying methodology of the metaevaluation (2.1), data sources and identification strategies (2.2) and its limitations and corresponding coping strategies (2.3). The analysis grid, methodological details, and the data collection instruments (i.e., report and ToR quality assessment tool, content assessment tool, interview guideline and survey) are all provided from annex 4 to 9. In chapter 3, the metaevaluation is briefly contextualised with Finland's development policy (3.1) and the delivery of Finnish aid (3.2); thus, the sample of evaluation reports is presented in light of the Finnish development context (3.3). A complete list of evaluations under this metaevaluation and their key characteristics are displayed in annex 10. Chapters 4, 5 and 6 form the core of the report; they contain findings regarding the quality of evaluation reports and their ToRs (i.e., the metaevaluation) (4), the quality of interventions (i.e., the summative analysis) (5), and the usefulness of the evaluations (6). Frequency tables, statistical tests, regression analyses, and a list of persons interviewed are presented in annexes 15 and 2. In chapter 7, conclusions are drawn, while chapter 8 provides recommendations. Other annexes contain the ToR for this assignment (Annex 1), a list of documents consulted (Annex 3), information on this assignment and the evaluation team and the work plan.





## 2.1. General approach and methodology

Despite EVA-11's long commitment to metaevaluations, this is the first metaevaluation **building on an earlier developed methodology** to allow sound comparison with previous results. We **replicated a two-stage approach**. The **first stage** of the analysis provides insights for all evaluation reports and available ToRs. It focused on **methodological quality** (component 1). In the **second stage**, we delved deeper into detail and concentrated on findings regarding the **interventions' quality** against the OECD DAC criteria (component 2).

We applied a **multi-method approach** with **quantitative and qualitative data collection and analysis methods** for decentralised evaluation reports and primary data from key stakeholders. This enabled us (i) to provide insights on the quality of decentralised evaluations, (ii) to summarise results of the Finnish development cooperation emanating from these reports and (iii) to shed light on their actual use and usefulness (component 3). To adjust to recent advancements in (i) evaluating development interventions and (ii) in Finnish development cooperation, both tools have been **amended to include the new OECD DAC criterion on coherence**

and **stronger consideration of the enhanced cross-cutting objectives** of Finnish development cooperation. As an additional new feature, Finnish **policy priority areas** have been **integrated**.

To allow **comparability between recent and earlier metaevaluation**, we conducted a **static analysis** comparing recent and previous quality scores wherein the new coherence criterion is further excluded. To **acknowledge increased evaluation quality standards**, we further performed a **dynamic analysis**, including the coherence criterion in overall quality scores for reports under the recent metaevaluation, but not for those under the previous assignment.

A two-stage multi-method approach with standardised assessments and qualitative content analysis was replicated.

The **quality assessment tool** is built on a wide range of single aspects related to methodological quality, which were, whenever appropriate and sufficient, checked against yes/no answer options. Otherwise, a four-step scale with the categories “good or very good”, “satisfactory”, “need for improvement”, and “inadequate” was applied in a **standardised assessment** to avoid oversimplification and the well-proved human tendency to centrality. Stronger differentiation at the lower end of this scale turns particularly beneficial for deriving recommendations on evaluation capacity development. Sets of single aspects were weighted and summarised to one sub-section (graded on a four-step scale), sub-sections were similarly weighted and summarised to sections, and sections to an overall assessment correspondingly. This stepwise procedure yielded to summary results tables to gain concise information regarding the quality of decentralised evaluation reports and the identification of general trends against previous metaevaluation results. Details on the development, content and application of the quality assessment tool and subsequent aggregated analysis are provided in Annex 6: Quality Assessment Tool for Reports and ToR.

A **minimal methodological quality threshold** was used to ensure that only valid, objective and reliable decentralised evaluation reports were considered **for** the second stage, i.e., the **summative content analysis**. Minimal methodological quality is, however, not understood according to scientific standards but in the light of good evaluation practice.

In the **content assessment**, we followed a similar stepwise procedure for single aspects, sub-criteria/topics, and criteria/topics. It was analysed if and to which extent evaluators assessed aspects of the interventions as fulfilled. Assessments provided in the decentralised evaluation reports were transferred by the metaevaluation team into standardised ratings according to a four-step scale with the answer options “no”, “rather no”, “rather yes”, and “yes”.



In addition, influencing factors determining the assessments presented in the decentralised evaluation reports were captured, thus going beyond **standardised assessments**. This means such **factors were coded** with the qualitative data analysis software MAXQDA. This allows summarising results and identification of emerging issues of the fraction of Finnish development cooperation within the scope of this metaevaluation. Details on development, content and application of the content assessment tool and subsequent analysis are provided in Annex 7: Content Assessment Tool.

After report-wise quality and content assessment, **descriptive summary statistics** and **regression analysis** were conducted with the statistical software package R. In addition, **qualitative content analysis** was performed to aggregate report-level findings and derive systematic results. As a new feature, **quantitative keyword analysis** was run, providing further hints regarding the cross-cutting objectives of Finnish development cooperation.

While quality and content assessment shed light on the quality of decentralised evaluation reports and the results reported by the evaluators, this **two-stage desk-based analysis remains limited to** the reports themselves and, thus, to the **perspectives of the evaluators** and their methodological quality of work. To figure out more regarding the **actual use of decentralised evaluations and their usefulness** for key stakeholders, we broadened the methodological approach of the predecessor metaevaluation and **added a third layer**.

With a mixed-methods comprehensive primary data collection, we **integrated the perspectives of MFA commissioners (i.e., from headquarters and Finnish embassies)** entrusted with the development interventions and their decentralised evaluations under the scope of this assignment **and (inter)national implementers** who managed the interventions reported in the decentralised evaluations under this metaevaluation.

As new feature mixed-methods primary data collection allowed integration of MFA commissioners' and intervention implementers' perspectives.

Informed by explorative interviews and discussions with the reference group and EVA-11, a guideline for **semi-structured key informant interviews** with MFA headquarters and Finnish embassies staff and a **survey** for all three key stakeholder groups were developed. The interview guideline and the survey questions for the use and usefulness assessment are provided in Annexes 8 and 9.

The data analysis included **qualitative content analysis and descriptive statistics**. Whenever possible, **quantitative data were triangulated with qualitative data**. A focus on **evaluation case-specific information** further allowed to **contextualise information on use and usefulness with methodological quality assessments and summative analysis** conducted by the metaevaluation team. The analysis grid, in Annex 4, shows in detail data sources and data analysis methods used to reply to each evaluation question.

Our **participatory approach** fostered exchange with the MFA during all stages of the assignment. **The high quality was assured** by extensive pre-tests, continuous inter-team coordination, cross-analysis of 10% randomly selected reports, joint analysis, intensive technical backstopping by the team leader and a strong quality assurance role of the deputy team leader. Finally, we used resources efficiently and followed good evaluation practice. More information on practical implementation and quality assurance are displayed at the end of the report.



## 2.2. Data sources and identification strategies

The **sample** under this metaevaluation contains all available decentralised evaluations published between 2017 and 2020 and was not part of the earlier metaevaluation. **80 decentralised evaluations** and **70 ToRs** have been identified from MFA's evaluation plan and were withdrawn from MFA's electronic archive application AHA and MFA's website or were requested by commissioning units with the support of EVA-11. In the absence of an inventory of interventions and corresponding evaluations, sample composition turned out cumbersome. Thus, it is possible that single decentralised evaluations which were eligible for this assignment could not be incorporated into our sample. Further, this explains the lack of 10 ToRs for the metaevaluation: They have neither been attached to the reports by the evaluators nor archived by MFA.

The sample comprises **mainly multi-, bi-, and multi-bilateral projects and programmes** which MFA completely or partially funds. In addition to the earlier metaevaluation, MFA added **decentralised evaluations on trust funds and institutions** to reflect their increasing importance in Finnish development cooperation.

**The sample includes multi-, bi- and multi-bilateral interventions funded by MFA or its partners. For the first time decentralised evaluations on trust funds and institutions were included.**

The administration of interventions and corresponding evaluations under this metaevaluation was done either directly by MFA or by a partner organisation. In the latter case, MFA participated in commenting ToRs and evaluation reports but has not been the commissioner of the evaluation. The inclusion of reports from 2017 which have not been captured under the previous metaevaluation was agreed with MFA to ensure continuity to the earlier metaevaluation whose scope ended in September 2017. The **intervention level** was chosen **as the unit of analysis**; thus, for multiple project evaluations, each evaluation referring to a particular intervention was counted as a single case within our sample. The quality assessment tool for component 1 on the report and ToR quality was applied to 80 decentralised evaluation reports and 70 ToRs.

In total, **eight reports did not comply with minimal methodological standards** and therefore have been **excluded from the content assessment**. This reduced the **sample size** for component 2, the **summative analysis, to 72**. **Data quality for the content assessment** is assessed as of **reasonable quality** because passing the minimal methodological standards suggests at large reliable evaluation results. A complete list of evaluations under this metaevaluation and their key characteristics is provided in Annex 10.

For **primary data collection**, different identification strategies were applied. Responsible **MFA commissioners in headquarters and embassies** were identified jointly with EVA-11 supported by the relevant units and the reference group. All persons identified were invited per email for a **semi-structured key informant interview**. Tracing of headquarters' and embassies' staff caused several attempts and dead ends as knowledgeable MFA staff could often not be correctly identified. We undertook a total of 241 contact attempts at the MFA headquarter and embassy level and finally carried out **26 key informant interviews**. As a larger share of identified MFA staff was involved in several evaluations, this generated, despite challenges, a sound body of qualitative insights on the evaluations at hand. The quality of the interview data is assessed as adequate.

In addition, a **survey** was **distributed among all identified knowledgeable MFA staff** at headquarters and in the embassies to capture as many voices as possible. The **identification of intervention implementers relied on MFA commissioners**. Thus, whenever an MFA contact person was successfully identified s/he was asked to share the implementer's contact details. The latter, in turn, were then also invited to participate in the survey, which yielded a total



of 65 invitations. To keep data collection as efficient as possible, we applied several filters to pose tailor-made questions to gain the perspectives of the three stakeholder groups (i.e., MFA headquarters, embassies, implementers).

For each project, we tried to cover as many of the three perspectives (MFA headquarters, embassies, implementers) as possible in the survey. For the 80 projects, this yielded a **potential set of 80 x 3 = 240 different perspectives in the survey**. In many cases, in which more than one person was identified for one project and perspective, all these persons were contacted. **Each identified person** was then invited to fill out **one survey for each evaluation project** s/he was involved in. Therefore, a survey link was generated for each potential respondent asking the survey questions as MFA headquarter/ embassy staff/ intervention implementer for that specific project. To avoid any mismatches, the title, year, and intervention code of the evaluation project was provided along with the survey link. Potential respondents received as many survey links as projects they were associated with except for a few cases in which this would have resulted in an extreme amount of survey links (e.g., > 10). In these cases, the evaluators picked a subset of survey links that were sent out to the potential respondent. Survey respondents filled out the survey links, providing answers to one or several evaluation projects. **In total, 119 survey responses** were obtained, reflecting 92 out of a maximum of 240 perspectives (i.e., for some perspectives of a single case, several individuals replied). This corresponds to a moderate sample coverage rate of 38%. However, we received survey responses for 57 evaluations out of our sample of 80 evaluations resulting in a rather high overall **sample coverage of 71%**. More details are presented in chapter 6.1.

The **quality of the survey data** is assessed as **moderate** given the relatively long recall period for evaluations published in 2017. Memory gaps, selective recall, and socially desirable responses may have led to biased survey responses. A further threat to data quality may be an upward bias of results emanating from a potential self-selection bias of respondents. Or put differently, it is plausible that survey participation for commissioners who assess an evaluation as (rather) useful is higher than the participation of commissioners assessing it as (rather) useless. To cope with this potential bias, we also provide lower boundary estimates when assuming that missing assessments of evaluation cases would be negative and thus counting them accordingly. Survey data is provided in annex 9.

**Primary data comprised 119 survey responses and 26 key informant interviews.**

## 2.3. Limitations and coping strategies

A **minor limitation** of the meta-analysis is the limited **scope of underlying secondary data**. Solely evaluation reports have been considered to assess the interventions' quality. Yet, detailed project documents (e.g., project proposals, progress, and final reports) were not consulted for this assignment. However, most evaluation reports disclose underlying validation. In 70% of the reports (57 out of 80), evaluators state that **evaluation results were validated** with stakeholders, in 77% (62 out of 80) with commissioners. This suggests that **assessments** provided by the evaluators **emanated at least partially** from an earlier **triangulation process**.

Moreover, for a considerable number of cases, a self-assessment of the overall quality of the intervention was captured from implementers in the survey. A comparison reveals that implementers assess their interventions on average comparable or slightly better. This is in line with experience from other assignments and does not provide hints for concerns regarding data quality.



Regarding primary data collection, a sample coverage rate of 38% (92 out of 240 when considering a maximum of three perspectives on a single intervention) poses a **limitation on case-based triangulation of different perspectives** from MFA headquarters, Finnish embassies, and implementers. Nevertheless, for 37% of the cases (30 out of 80), we got at least two perspectives allowing **triangulation among survey responses** on the same interventions in more than **one-third of the cases**. Further, the case coverage rate of 71% (57 out of 80) underlines that we were able to collect **secondary and primary data for the vast majority of interventions**. This is a clear advantage over earlier metaevaluations which were exclusively grounded on secondary data. Further qualitative key informant interviews yielded a valuable data reservoir to triangulate quantitative survey data against their limitations (presented in chapter 2.2).

As explained in chapter 2.2, **self-selection into the survey** may have fostered **upwards biased results** on the assessment of use and usefulness of the evaluation reports. As a coping strategy, a **lower boundary estimate** was calculated, and **within-case triangulation** was applied if possible. In contrast, the **selection of interviewees** was due to personal interaction, with the attitude to also learn from challenging experiences, **much less biased** than the survey. Hence **data quality** is assessed as **good**, and given the fact that many interviewees were in charge of several evaluations, the 26 key informant interviews conducted captured generalised experiences on a high share of our sample. In addition, it can be regarded as a minor limitation that no primary data from the evaluators were captured. Although their perspective on the interventions' quality is reflected in their reports, their feedback on the quality of the ToRs was thus not captured under this assignment.

With available information sample representativeness cannot be assessed.

Another limitation refers to the **not fully known whole population of Finnish bi-, multi- and multi-bilateral interventions**. Thus, we **cannot assess** to which extent this **sample is representative** of the portfolio of Finnish development cooperation. As the geographical scope, sectoral distribution of the interventions and their budgets vary, and as we look at different interventions themselves, **direct comparability of the quality of Finnish bi-, multi-, and multi-bilateral interventions** as assessed in the **earlier metaevaluation against actual findings** are **strongly limited**.

However, the previous and recent metaevaluations are comparable regarding their shares on mid-term vs final evaluations, MFA vs non-MFA commissioner, independent evaluators vs consulting firms/ institutions, and their regional spread. Thus, **comparisons on the overall quality of the evaluations** of the previous and recent metaevaluation are **possible**. Nevertheless, **several limitations** remain: as evaluation budgets and intervention-specific characteristics (as mentioned above) vary, the evaluability of the interventions might have been very heterogeneous.

These limitations are **partially mitigated by simplifications undertaken** when assessment tools were designed. Moreover, the **equal weighting of small and large interventions** may support coping with challenges in this regard. However, **both aspects**, simplification and equal weighting, pose **at the same time strong limitations** on the metaevaluation. They were acknowledged during results interpretation.

**To reduce the risk of subjective assessments** or different understandings of specifications by different metaevaluation team members, **specifications of four-scale answering options** in the assessment tools were as exact as possible given **precise guidance**. To avoid biased results, **many ratings** have been **limited to detailed yes/no questions**, and aspects have been





whenever possible split-up until yes/no answers became possible. **Aggregation** at all levels was **standardised**, leaving no room for subjective assessments at higher levels.

To ensure **inter-subjective comparability in the content analysis**, the metaevaluation team was trained to exclude their own expert judgement on single interventions. Instead, the content assessment **strictly followed the tools and protocol** presented above. Only at an aggregated level during joint analysis expert judgements were considered and triangulated among metaevaluation team members.

Whenever it comes to **weightings** throughout the aggregation process, however, results are **based on heavy expert judgements**. To minimise subjectivity, **weights** that were in the **predecessor** metaevaluation were discussed intensively and **remained the same** in this assignment. **For additional news aspects** which have been inserted reflecting the new OECD-DAC regime and Finnish policies, **weights were introduced upon consensus** in the metaevaluation. Appreciating this limitation, we **refrain from overemphasising overall scores** also presenting insights on lower levels. Overall scores serve only for the sake of linking different assessment tools and performing an economically efficient regression analysis and comparisons of sub-sample groups.

**Yes/no-answers and clearly defined four-step scales reduced room for biased assessments considerably.**

Finally, **given the nature of the metaevaluation**, this assignment cannot be understood as a re-evaluation of single interventions. **Interpretation of assessments at individual project or programme level is not valid**. Therefore, we **strictly limit results interpretation to the aggregated level**.



## 3. Context Analysis

### 3.1. Finland's Development Policies

The main goal of Finland's human-rights based development policy and development cooperation is to eradicate poverty and reduce inequalities. Finland has launched development policies corresponding to each new parliamentary period in 2004, 2007, 2012, and 2016 (MFA, 2004, 2007, 2012, 2016). The **2016 Development Policy Programme** saw a new structuring of the overall global vision to four specific policy priority areas in line with the **Sustainable Development Goals (SDGs)**:

- 1. Strengthening the status and rights of women and girls**, with an emphasis on sexual and reproductive health and rights.
- 2. Strengthening the economic base of developing countries and creating jobs**, with an emphasis on innovations and the role of women in the economy and female entrepreneurship.
- 3. Education<sup>2</sup>, well-functioning societies and democracy**, with an emphasis on high-quality education, improved tax systems and support for democracy and the rule of law.
- 4. Climate change and natural resources**, with an emphasis on strengthening adaptation alongside mitigation of climate change, food security and water, meteorology and disaster risk prevention, forests and safeguarding biodiversity.

The government that stepped into power in June 2019 endorsed the framework created by its predecessor by publishing a document titled "Theories of Change and Aggregate Indicators for Finland's Development Policy 2020" (MFA, 2020b). This publication draws together the work implemented from 2016 through 2019 on defining the activities, outputs, outcomes, and impacts of the four policy priority areas, including their corresponding indicators. Although with less detail, "Programme of Prime Minister Sanna Marin's Government 2019 - Inclusive and competent Finland - a socially, economically and ecologically sustainable society" also includes provisions on Finland's overall development policy (Finnish Government, 2019).

The original 2016 Development Policy Programme was placed in a long-term context with the formulation of the "Report on Development Policy Extending Across Parliamentary Terms" (MFA, 2021b). The strategy aims to consolidate the longstanding nature and coherence of Finland's development policy and coordination. However, by the end of 2021, the Parliament has not yet formally approved the report.

---

<sup>2</sup> Emphasis on education was brought in with the "Theories of Change and Aggregate Indicators for Finland's Development Policy 2020" discussed in this section of the report.



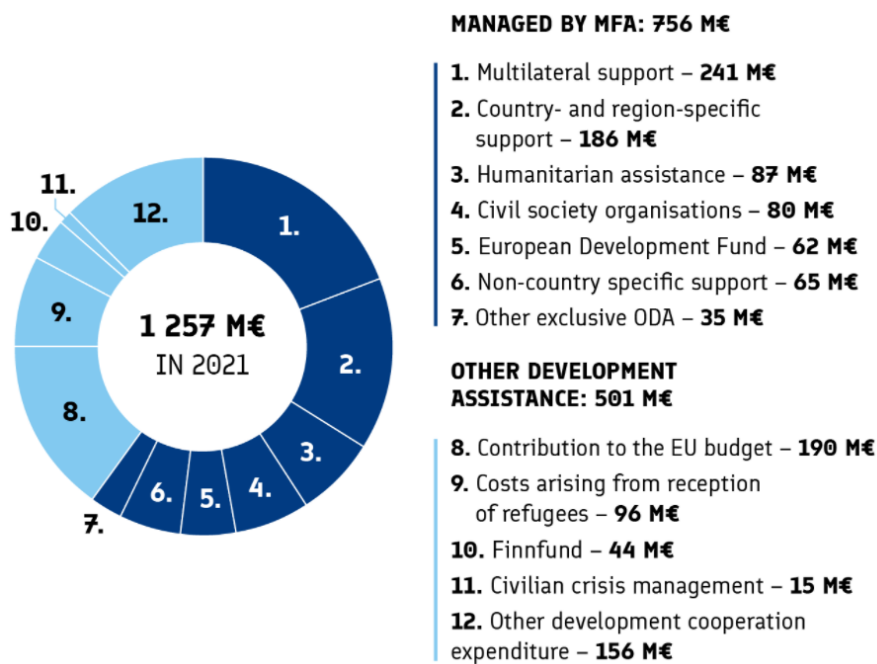
Furthermore, and in parallel with the clarification of Finland’s Development Policy Programme 2016, MFA also developed a Guideline for the Cross-Cutting Objectives in 2020 (MFA, 2020a). While gender equality, reduction of inequalities, and climate change have been cross-cutting **priorities for Finland** over the years, the new Guideline provides a more specific definition of the themes.

The Cross-Cutting Objectives (CCOs) defined in 2020 are (1) **Gender equality**, (2) **Non-discrimination** with a focus on persons with disabilities, and (3) **Climate resilience and low emission development**. Furthermore, the Report on Development Policy across Parliamentary Terms published by the Finnish Government in May 2021 includes the protection of the environment, especially safeguarding biodiversity, as a new CCO (MFA, 2021a). In the framework of this metaevaluation, the third category was broadened to encompass topics broadly under the concept of climate change and the environment. The reasoning behind the approach is to ensure that the analysis captures widely any valuable information and lessons learnt that can guide the MFA in its future work.

### 3.2. Delivery of Finnish aid

Finland’s development aid has been around 1,000 million EUR per year since 2008, corresponding to approximately 0.4 or 0.5 per cent of the Gross National Income (GNI). The amount peaked in 2014; however, budget cuts introduced in 2015-2016 reduced the annual budget gradually to 0.36 % of GNI by 2018. Since then, a positive trend can be observed again; in 2020, Finland’s contribution had reached 0.47% of GNI (MFA, 2021a). Finland’s development assistance is delivered through various channels, many of which involve multilateral organisations, including the EU (see Figure 3: Channels of Finnish aid).

Figure 3: Channels of Finnish aid



Source: MFA (2021a)



The aid modalities included in this metaevaluation covered mainly project-based support and thematic funding to multilateral organisations, bilateral and regional projects/programmes, civil society interventions, and projects implemented under the Institutional Cooperation Instrument. A few cases also covered Fund for Local Cooperation, budget support, and concessional credits. However, it should be noted that the majority of CSO evaluations are commissioned by the CSOs themselves under the project and programme-based support modality. Those evaluations were not part of this metaevaluation's scope.

### 3.3. Evaluation reports in light of the Finnish development context (EQ1)

This metaevaluation covers evaluation assignments carried out between 2017 and 2020. Thus, all evaluations are implemented under the framework of the Development Evaluation Norm established in 2015, which provides the definition and the legal basis for the evaluation of development policy and cooperation.

Development evaluation serves a dual purpose in the MFA, accountability and organisation-wide learning. In terms of accountability, provisions on MFA's responsibilities are set in the State Budget Act and State Budget Decree (MFA, 2015a). The learning aspect aims at constant improvement of quality of development cooperation by providing independent and impartial knowledge on the activities. Evaluations carried out by the MFA are also guided by the Evaluation Manual (MFA, 2018a, 2021c), which sets out the key contents and quality standards of both decentralised and centralised evaluations.

**Results of this meta-evaluation reflect only a fraction of Finland's development cooperation.**

A series of other guidance documents set out standards that are relevant from the perspective of evaluation: the Guideline for the Cross-Cutting Objectives in the Finnish Development Policy and Cooperation (MFA, 2020a), the Manual for Bilateral Cooperation originally published in 2012 and updated in 2016 and 2018 (MFA, 2018b), and the Guidance Note on the Human Rights Based Approach in Finland's Development Cooperation (MFA, 2015b). Additionally, a report template with detailed information on the content of the different sections is handed out to the evaluators to write their reports accordingly.

The interventions evaluated under this metaevaluation cover implementation periods falling between 2008 and 2020. Thus, with the exception of three cases (starting in 2002, 2004, 2007), interventions fall under the validity of the 2007-2012 Finnish Development Policy, their successor 2012-2015 or the 2016 Finnish Development Policy Programme.

Regional units commission most project and programme evaluations while some of them are covered by the Development Policy Department, Political Department, Communications Department, Department for International Trade and also the Legal Service. However, certain aid modalities, such as humanitarian aid or influencing work in multilateral organisations, are evaluated through centralised evaluations commissioned directly by EVA-11. Humanitarian interventions are also evaluated by those organisations that deliver assistance on the ground. Therefore, **the conclusions and emerging issues that will be identified as a result of the metaevaluation reflect only a fraction of Finland's development cooperation portfolio.**



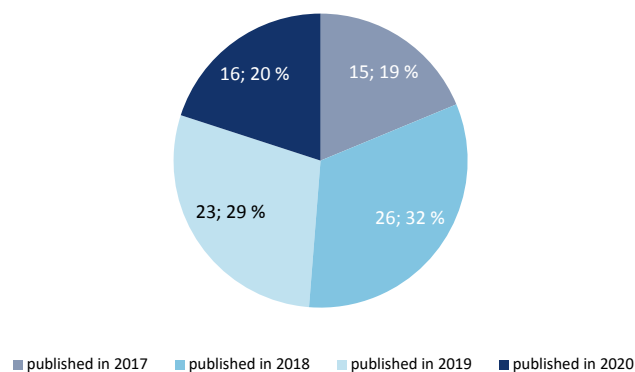
In addition, the sample of projects that have been evaluated by the regional units do not necessarily represent the whole portfolio of interventions; therefore, the conclusions must be understood in this limited context. In the absence of a systematic inventory on all bi-, multi-, and multi-bilateral interventions with their key characteristics (e.g., geographical scope, budget range, nature of the intervention, implementation dates etc.), the representativeness of the sample of evaluation reports cannot be assessed. Nevertheless, according to MFA, the available sample is considered as nearly complete and fairly illustrative of the whole portfolio of bi-, multi-, and multi-bilateral interventions of Finnish development cooperation.

In response to the first evaluation question EQ1 “How can MFA’s decentralised evaluation portfolio be described?” different key characteristics of the evaluation and the underlying interventions are described.

### Key characteristics of the sample

Figure 4 shows that out of 80 evaluation reports, 15 (19%) were published in 2017, 26 (32%) in 2018, another 23 (29%) in 2019 and the remaining 16 (20%) in 2020. Unless stated otherwise, the sample size referred to throughout this report is the total sample of 80 reports for the meta-evaluation, 70 for the ToRs assessment, and 72 reports for summative analysis). When referring to different sample sizes, the actual number will be indicated at the beginning of the paragraph or will be included in brackets (e.g. 18 out of 60, 30%).

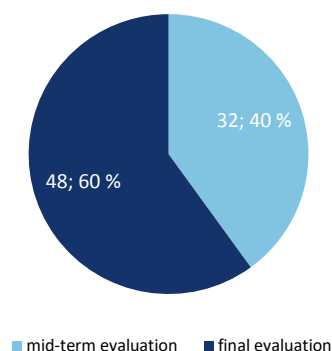
**Figure 4: Year of publication of the evaluation report (n=80)**



Source: own statistics based on analysis of reports

Figure 5 illustrates that the sample is composed of 32 (40%) mid-term evaluations and 48 (60%) final evaluations.

**Figure 5: Nature of the evaluation (n=80)**

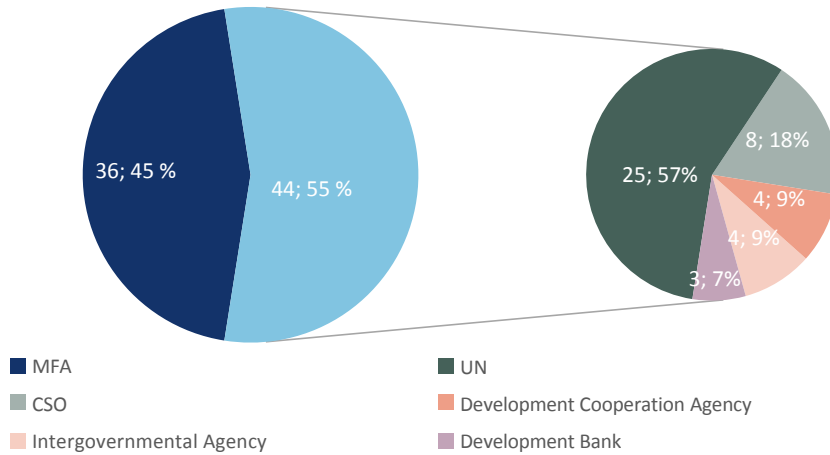




Source: own statistics based on analysis of reports

The commissioner of the evaluation is shown in Figure 6, wherein a bit less than half (36, 45%) of the sampled evaluations were commissioned by the MFA. Out of the 44 (55%) other commissioners, more than half were multilateral international organisations under the United Nations (25, 57%), 8 (18%) were identified as CSOs, including local or international NGOs, and the remaining quarter is scattered among the World Bank, intergovernmental agencies (e.g., OECD and Council of Europe), government agencies and development cooperation agencies from other countries.

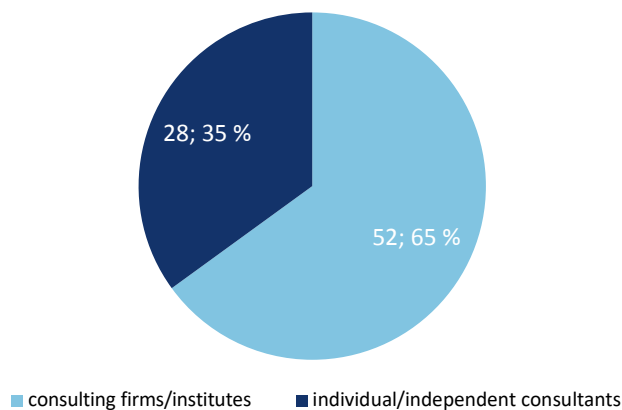
**Figure 6: Commissioner of the evaluation (n=80)**



Source: own statistics based on analysis of reports

Figure 7 shows that 52 (65%) of the sample evaluations were implemented by a consulting firm or an institution, while the remaining 28 (35%) were conducted by independent consultants. A single person hired to conduct the evaluation or a team of two independent consultants recruited are considered here as independent consultants.

**Figure 7. The implementer of the evaluation (n=80)**

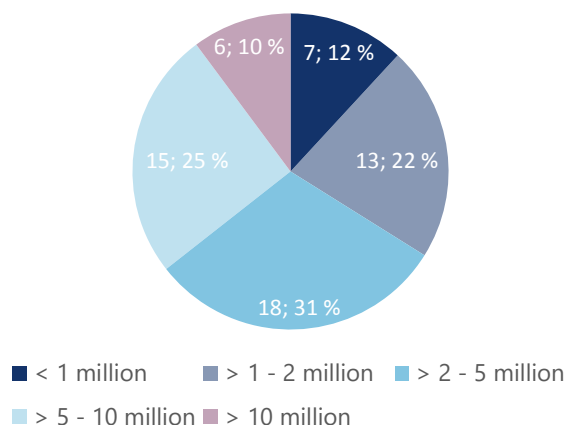


Source: own statistics based on analysis of reports



59 out of the 80 reports had information regarding the intervention's budget contributed by Finland. The smallest amount is 120,000 EUR, and the biggest amount is at 32.7 EUR million (with a mean of roughly 5.3 million and a median of 4 million). Figure 8 further specify that a little more than three-quarters of the sample fall between the budget range of 1-10 million EUR, while budgets under one million and over 10 million are less common for bi-, multi- or multi-bilateral interventions in Finnish development cooperation.

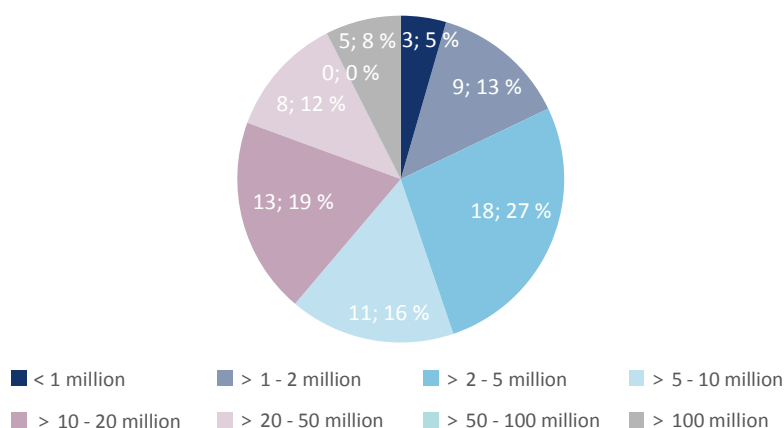
**Figure 8: Finland's budget of the intervention (n=59)**



Source: own statistics based on analysis of reports

Information on the overall intervention budget is available for 67 cases, as depicted in Figure 9. More than 90% of the interventions have overall budgets falling within the range between 2-20 million EUR. The presence of interventions with budgets over 100 million EUR (3 projects in the billion range) is characterised by Finland's participation in large-scale, multilateral efforts.

**Figure 9: Overall budget of the intervention (n=67)**

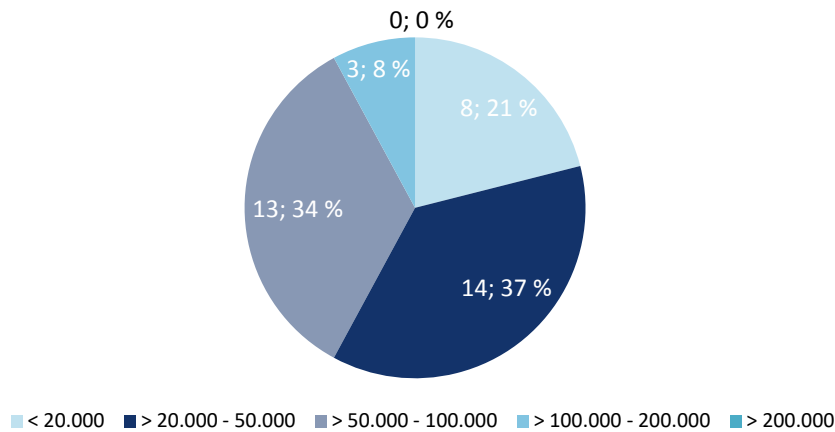


Source: own statistics based on analysis of reports

Only 38 of the sample contained information on the net evaluation budget. As shown in Figure 10, the evaluation budget ranges from 4,000 to 130,000 EUR (with a median of 50,000). Almost three-quarters of the 38 interventions have an evaluation budget ranging from 20,000 to 100,000 EUR.



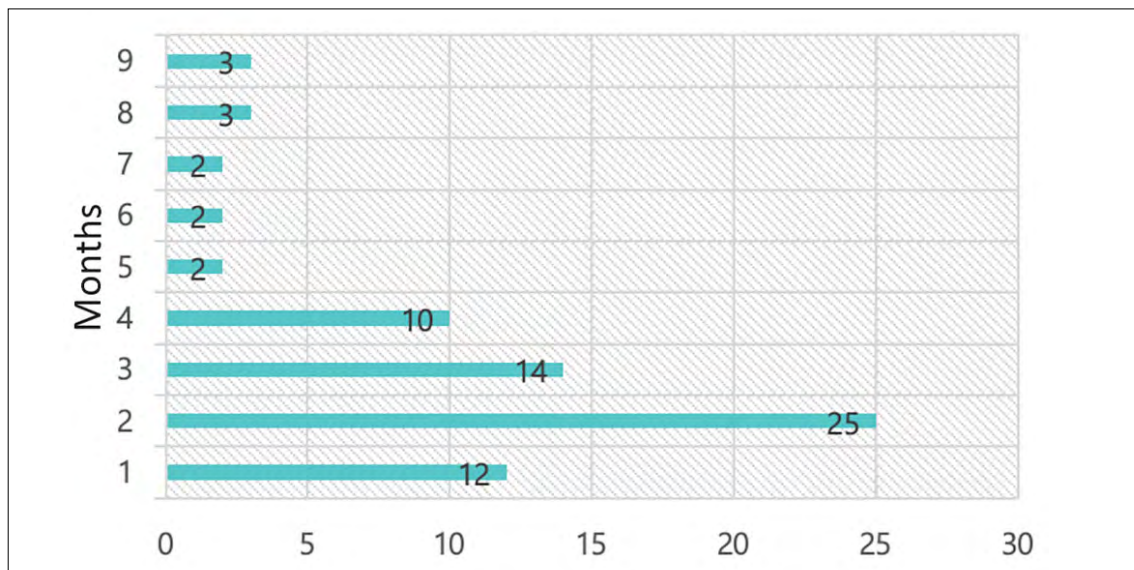
**Figure 10: Net evaluation budget of the intervention (n=38)**



Source: own statistics based on analysis of reports

Meanwhile, Figure 11 summarises the duration of evaluations (actual evaluation time, excluding any MFA internal planning) found in 73 of the interventions. More than half of the 72 evaluations were conducted within 2 to 3 months, with 1 and 2 months each with about 15%.

**Figure 11: Duration of the evaluation (n=73)**



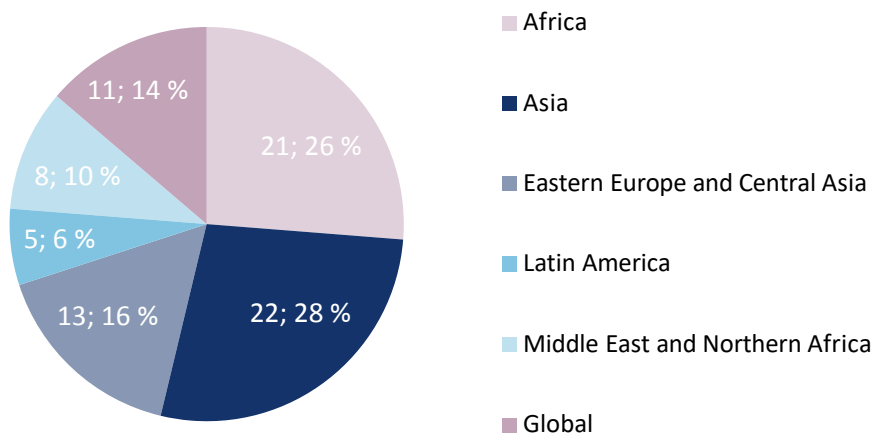
Source: own statistics based on analysis of reports

Figure 12 illustrates the regional distribution of the interventions, with more than half focusing on two regions, Asia (22, 28%) and Africa (21, 26%). 11 (15%) of the interventions are implemented at a global scale, and clearly, the least number of interventions (5, 6%) were implemented in Latin America.





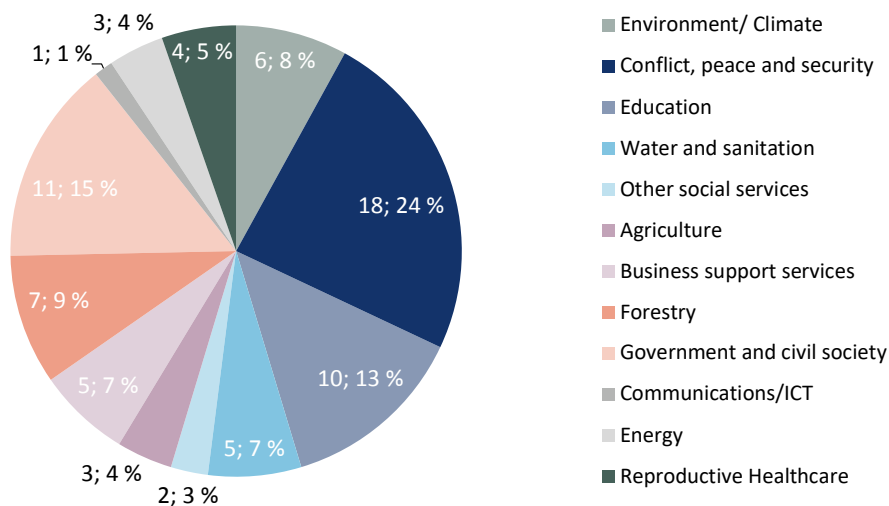
**Figure 12: Regional distribution of interventions (n=80)**



Source: own statistics based on analysis of reports

Regarding sectoral distribution, Figure 13 displays the different thematic areas of the sample interventions. 18 (24%) of the interventions fall under ‘Conflict, peace and security’ coming in second with 11 (15%) is ‘Government and civil society’ with ‘Education’ in a close third at 10 (13%). The remaining half of the sample is distributed over nine other sectors, including ‘Environment/Climate,’ ‘Forestry,’ ‘Water and sanitation,’ and ‘Reproductive Health.’

**Figure 13: Sectorial distribution of interventions (n=80)**

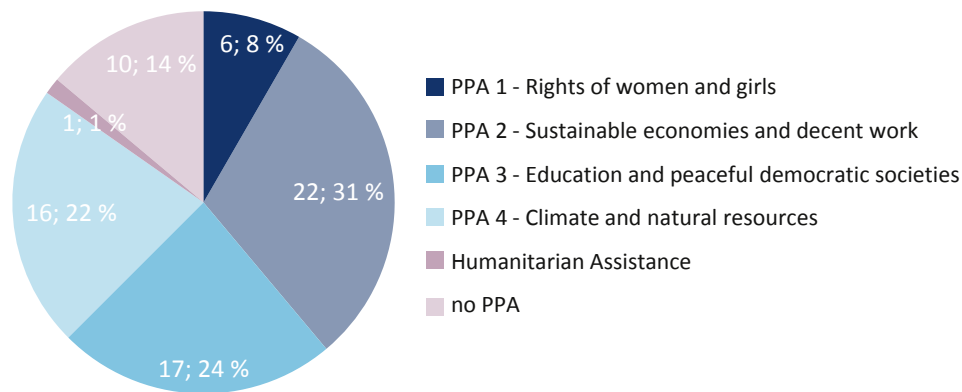


Source: own statistics based on analysis of reports



Policy priority areas (PPAs) are gaining importance in Finnish development cooperation, thus in contrast to earlier metaevaluations, interventions have been categorised per PPA. As displayed in Figure 14, about three-quarters of the interventions in our sample address PPA2 on sustainable economies and decent work (31%, 22), PPA3 on education and peaceful democratic societies (24%, 17) and PPA4 on climate and natural resources (22%, 16). PPA1 on rights of women and girls is only a priority of some interventions (8%, 6). While only one intervention is directed to humanitarian assistance, ten interventions (14%) are not directed to any PPA. This gives an indication that our sample clearly differs from the overall distribution of Finland's development cooperation, where PPA 1 comprises the highest shares in terms of the number of activities and total spending, followed by PPA3, PPA4 and PPA2 (MFA, n.d.).

**Figure 14: Distributions of PPAs (n=72)**



Source: own statistics based on analysis of reports

### Comparison to the sample of the previous metaevaluation

Comparing the above-presented characteristics of the sample with that of the previous metaevaluation (Silvestrini et al., 2017), there are noticeably more similarities than differences. Almost identical is the nature of evaluation, commissioner of evaluation, the implementer of evaluation and geographical scope. They basically display the same ratio of characteristics. Furthermore, not much has changed in terms of Finland's budget contribution; in both samples, the largest share is within >1 million and <10 million and the budget below 1 million and over 10 million are exceptions.

However, a comparison of overall budgets, net evaluation budgets, and sectorial distribution reveals major differences. While roughly half of the overall budget is still occupied by budgets ranging from >5 million to 50 million in both metaevaluation samples, other budget ranges, however, have either doubled or clearly declined in share. A similar trend is observed for the net evaluation budget, with approximately one-third of the pie occupied by >50 thousand to 100 thousand while all other budget ranges are disparate. Only 'Education' maintained its ranking in terms of sectoral



distribution of the interventions. 'Environment/Climate' used to have the most interventions in the previous metaevaluation but fell to about one third in the current sample. Another significant difference can be seen in the sector 'Government and Civil society' wherein the current sample shows more than double that of the quantity from the previous metaevaluation.

This variance among sample characteristics poses serious limitations to the comparability of intervention quality. However, due to methodological simplifications as described in chapter 2.2, a comparison of the quality of the evaluation reports under the previous metaevaluation with those under the recent metaevaluation is feasible.

**Variance among sample characteristics does not allow a comparison of intervention quality. A comparison of evaluation reports' quality is, however, possible.**



## 4. Findings of the Metaevaluation

This chapter presents the findings of the quality assessment of 80 evaluation reports and 70 corresponding ToRs, i.e., component 1. In chapter 4.1, we first reply to **EQ3** on the **quality of ToRs** of decentralised evaluations (4.1.1) and on **systematic patterns** between the **quality of ToRs** and the **quality of** corresponding **evaluation findings** (4.1.2). Our regression analysis also provides answers to the second part of **EQ2** on potential quality **differences between MFA-commissioned evaluation reports and reports commissioned by others** and on potential **differences among other sub-groups** (e.g., final vs mid-term evaluations, individual/independent evaluator(s) vs consulting firm/institutes).

In chapter 4.2, we provide an overall answer to **EQ2** on the **quality and reliability of decentralised evaluation reports** when presenting a summary assessment on the overall quality of evaluation reports. Further, we disentangle overall quality when reporting on the quality of single sections (i.e., introductions and context analyses, evaluation methodologies, evaluation findings inclusive of appropriate capturing of OECD DAC criteria, conclusions and recommendations, acknowledging cross-cutting objectives (CCOs) and human rights-based approaches (HRBA), and executive summaries. In Annex 12: Interventions' Quality on Single OECD DAC Criteria, detailed quality assessments on underlying aspects of single sections are provided. By doing so, we lay the foundation to answer EQ4 on gaps regarding MFA's evaluation capacity.

Finally, in chapter 4.3, we address **EQ4** on **key differences and possible** trends in comparison with the **previous metaevaluation** (4.3.1) and summarise on **gaps** identified **regarding MFA's evaluation capacity** (4.3.2).



## 4.1. Quality of underlying ToRs and their influence on the quality of evaluations (EQ3)

### 4.1.1. Quality of ToRs

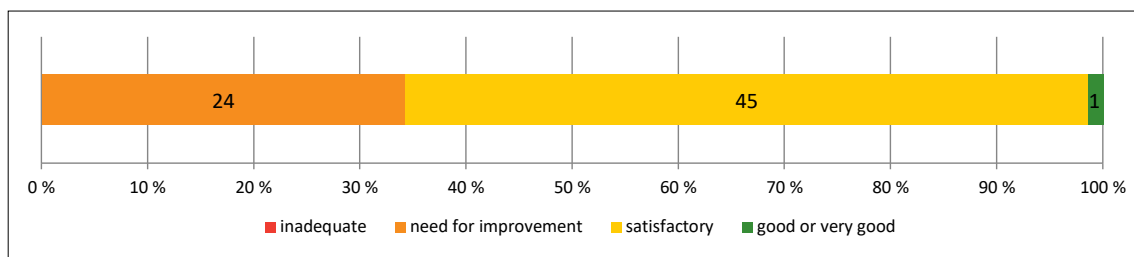
#### Highlights of the section:

- About two-thirds of the ToR are of satisfactory quality; about one third discloses a need for improvement.
- Only one ToR is rated as (very) good; none is inadequate.
- More than half of the ToRs need improvement or are even inadequate regarding methodological guidance and consideration of cross-cutting objectives (CCO).
- No major changes in the overall quality of ToRs between the previous and the recent metaevaluation can be observed.
- However, evaluation criteria are more frequently and better addressed in the ToRs than in the previous meta.



The ToRs lay the foundation for the implementation of evaluations. They also provide formal and structural guidance for the evaluation reports. In total, from the 80 reports, 70 ToRs were available for analysis. As displayed by Figure 15, out of the 70 available ToRs, 45 were of overall satisfactory quality (64%), 24 were identified with the need for improvement (34%), only one ToR was assessed as (very) good, but none was inadequate.

Figure 15: Overall quality of ToRs (n=70)



Source: own statistics based on analysis of reports

More details on the quality of the ToRs can be found in Figure 16, which shows the seven sub-sections of overall ToR quality, including (i) the **description of the intervention**, (ii) **evaluation objectives, purpose and scope**, (iii) **evaluation questions**, (iv) **evaluation criteria**, (v) **methodology**, (vi) **evaluation process and quality assurance**, and (vii) **CCOs and HRBA** (see Annex 6: Quality Assessment Tool for Reports and ToR for details on how the data was aggregated).

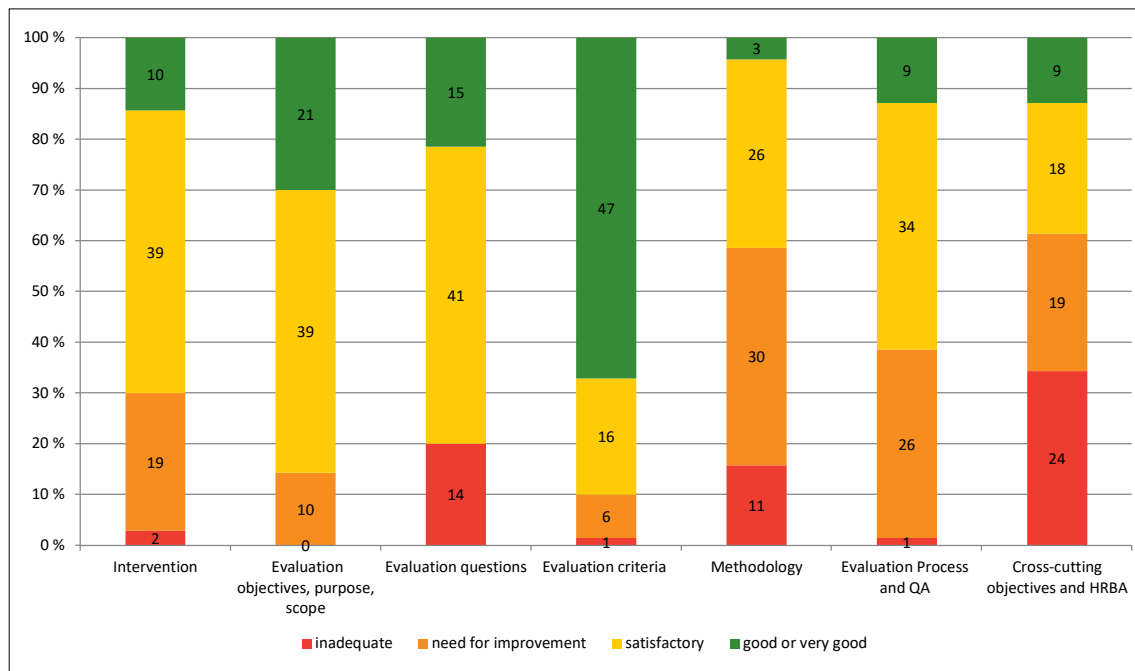


The sub-sections with the **largest room for improvement**, that is, the sections with the smallest numbers of (very) good assessments and high numbers of inadequate assessments, are **methodology** and **cross-cutting objectives**. Specifically, regarding methodology, 11 ToRs were assessed as inadequate (16%), 30 as in need for improvement (43%), 26 as satisfactory (37%), and 3 as good or very good (4%). With respect to CCOs, 24 ToRs were assessed as inadequate (34%), 19 as in need for improvement (27%), 18 as satisfactory (26%), and 9 as (very) good (13%).

On the other hand, the sub-section of the ToR assessment that displays the **largest share of high-quality scores** is on the **evaluation criteria**. 47 ToRs were assessed as being of good or very good quality (67%), and 16 were assessed as satisfactory (23%), while there was a need for improvement in 6 cases (9%), and only one case was inadequate.

The other sub-sections were, to a vast extent, moderately satisfactory or better assessed. In the section on presenting the intervention, 10 ToRs were assessed as good or very good (14%), 39 as satisfactory (56%), 19 as in need for improvement (27%), and 2 as inadequate (3%). Displaying evaluation objectives, purpose, and scope of 21 ToR were assessed as good or very good (30%), 39 were satisfactory (56%), and 10 needed improvement (14%). Finally, the evaluation process and quality assurance were assessed as good or very good for 9 ToR (13%), for 34 as satisfactory (49%), for 26 as in need for improvement (37%), and for one case as inadequate.

**Figure 16: Quality of ToR (n=70)**



Source: own statistics based on analysis of reports

The overall quality of ToR did not change much between the previous and the recent meta-evaluation. Yet, the good rating on the evaluation criteria is an improvement to the previous metaevaluation, at least when comparing the numbers of ToRs of (very) good, satisfactory, need for improvement, and inadequate quality.

## 4.1.2. Linkages between the quality of ToRs and quality of reports

### Highlights of the section:

- The quality of ToRs is a driver of the 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).
- No statistically significant differences in quality of evidence among sub-groups (i.e., MFA vs non-MFA commissioner, individual vs associated evaluators) can be observed.



We tested whether the **ToRs are a driver of report quality** by conducting a series of linear regression models with overall report quality and single sub-sections, including six potential predictors of report quality. These are overall ToR quality and five control variables (see annex 15 for further details). We did this to examine whether **overall ToR quality makes a unique contribution to overall report quality** even after taking other potential influences into account, such as evaluation budgets, evaluation durations, whether the evaluation was conducted by independent consultants (vs consulting firms/institute), whether the evaluation was commissioned by the MFA (vs others), and the geographical scope covered by the evaluation.

Table 1 provides the results and shows that the quality of ToRs is a significant predictor of the **quality of evidence** while controlling for other potential influences. Note that the regression analysis includes only 33 of the overall sample of 80 evaluation reports due to missing data but that the model is robust to different model specifications. The results can be interpreted as follows: a one-unit increase of ToR quality is associated with nearly a one-unit increase in the quality of evidence, provided that all other influences do not change (0.942\*\*, Table 1).



**Table 1: OLS Regression Analysis**

	OLS
(Intercept)	-0.277
	(1.310)
ToR quality	0.942 ***
	(0.290)
Evaluation budget (log)	0.086
	(0.099)
Evaluation duration (months)	0.043
	(0.031)
Independent consultant(s) (dummy)	-0.063
	(0.195)
MFA commissioned (dummy)	-0.306
	(0.180)
Ordered geographical scope	-0.146
	(0.084)
<i>N</i>	33
Model Fit:	
F (6, 26)	3.459
<i>p</i>	0.012**
<i>R</i> <sup>2</sup>	0.444
Adjusted <i>R</i> <sup>2</sup>	0.316
Note: Standard errors are heteroskedasticity robust. Ordered geographical scope: (sub-)national, national, regional and international interventions. *** $p < 0.01$ ; ** $p < 0.05$ ; * $p < 0.1$ . Source: own analysis of evaluation reports and ToRs.	

Source: own statistics based on analysis of reports

As further displayed by Table 1, none of the other variables was significantly linked to the quality of evidence in this analysis. Thus, **evaluation quality only significantly varies with ToR quality**. No differences for different commissioners or regarding other sub-groups could be observed.





## 4.2. Overall quality and reliability of decentralised evaluation reports (EQ2)

### Highlights of the section:

- Nearly half of the reports are of satisfactory quality, while none is (very) good.
- Nearly another half needs improvement, and 10% are inadequate; most of them are however atypical evaluations of funds or organisations.
- (Very) good quality is only found in introductions and context analyses, as well as in conclusions and recommendations.
- The quality of findings, underlying methodology, consideration of cross-cutting objectives (CCOs) and human rights-based approach (HRBA), and executive summaries are mostly in need of improvement.
- Nearly all reports are to some extent evidence-based; about half rather link findings to evidence.
- Only a couple of reports received good scores on appropriate capturing of OECD DAC evaluation criteria. Coherence, efficiency, and sustainability are less appropriately captured than other criteria.
- A quarter of the reports are inadequate regarding CCOs, about half in need for improvement.



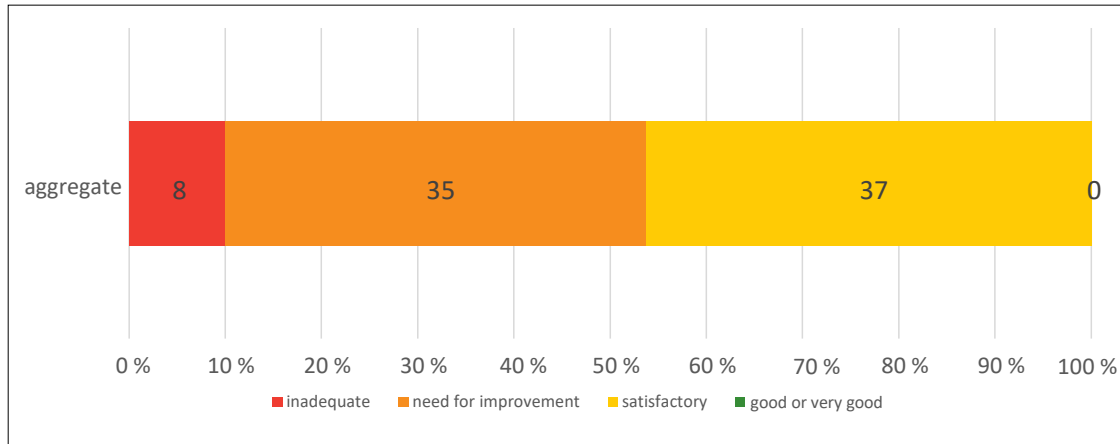
Figure 17 shows the result of the **overall quality aggregate** of decentralised evaluation reports. Out of the 80 reports, 37 are of satisfactory quality (46%), 35 were assessed as in need for improvement (44%). The 8 reports that were rated as inadequate (10%) are mainly atypical cases, referring to the evaluation of funding instruments, supported organisations or complex partnerships.

These evaluation reports are more difficult to capture with the current quality assessment tool that has been designed for programme and project evaluation. This means that atypical cases might have been unintentionally penalised for their different reporting structure, leading potentially to an underestimation of their report quality.

As already suggested in chapter 4.1.2 and further displayed in annex 15, the overall quality of the evaluation reports does not differ between (i) mid-term and final evaluations, as well as between (ii) evaluations conducted by independent evaluators and by consulting firms. Further, neither (iii) budgets, (iv) durations, and (v) geographical scopes of the evaluation nor (vi) the commissioner of the evaluation significantly drives the overall report quality.



**Figure 17: Overall quality of reports (n=80)**

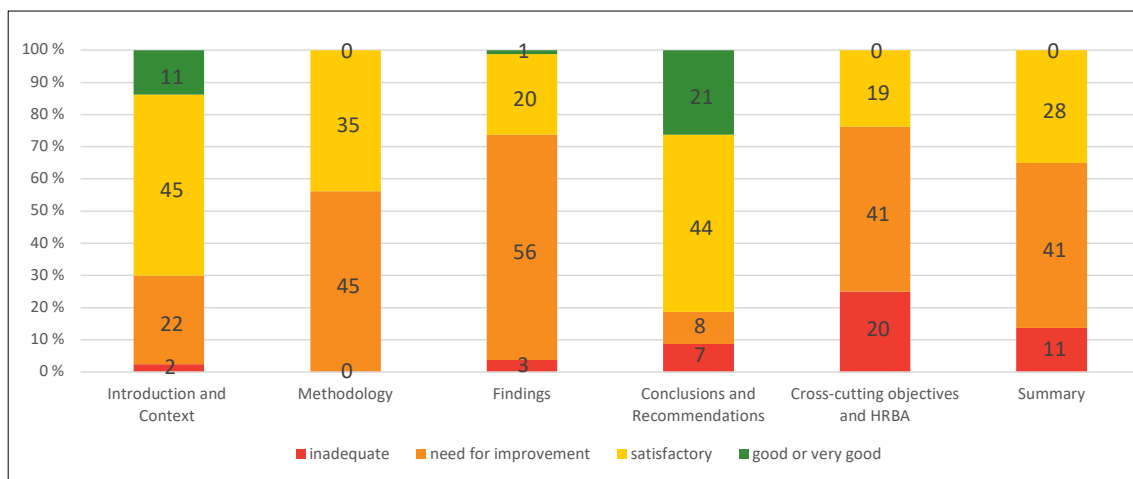


Source: own statistics based on analysis of reports

Figure 18 displays the frequency distributions for the six subsections from which the overall quality aggregate is derived: (i) introduction and context, (ii) methodology, (iii) findings, (iv) cross-cutting objectives, (v) conclusions and recommendations, and (vi) summary. When comparing the bar charts, it becomes clear that the low overall quality assessment mainly stems from low quality in the methodology, findings, CCOs, and summary sections.

The two sections with the **highest quality** assessments are **introduction and context** and **conclusions and recommendations**. Good or very good quality assessments are almost exclusively limited to these two sections (11, 14%; and 21, 26%, respectively). A **strong need for improvement** can be observed on the **methodology** (45, 56%), **findings** (56, 70%), **CCOs and HRBA** (41, 51%), and **summary** (41, 51%) sections with a particularly **large share of inadequate cases** in the **CCOs and HRBA section** (20, 25%).

**Figure 18: Quality Dimensions (n=80)**



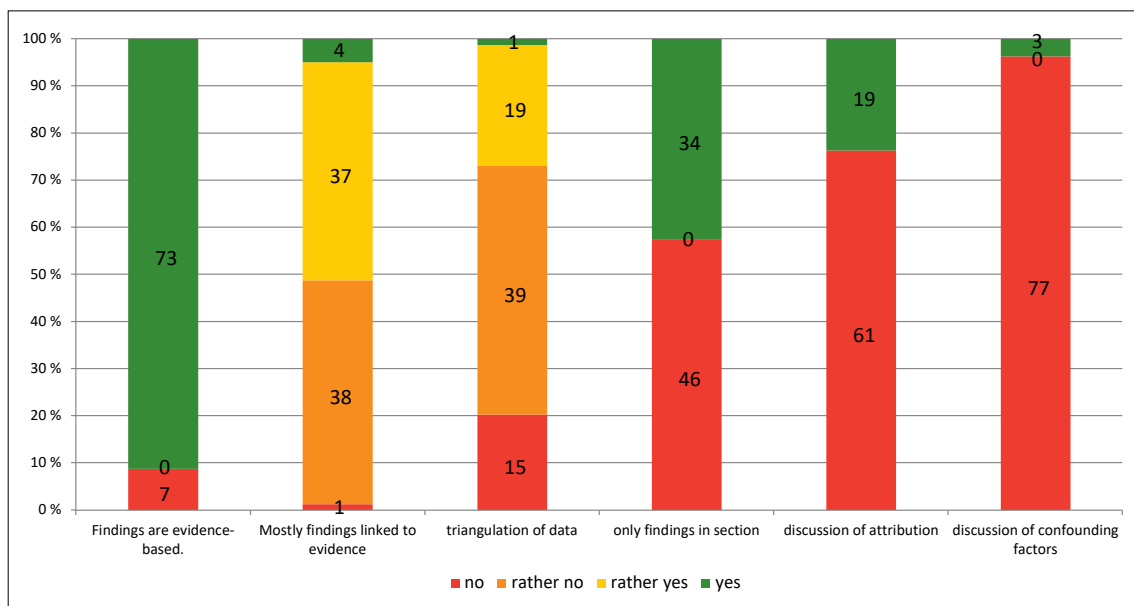
Source: own statistics based on analysis of reports



Although, regarding the **findings** section, most cases fall under the need for improvement (56, 70%), while only a quarter of the cases is considered satisfactory (20, 25%), the overall interpretation on the findings section should not be as negative as Figure 18 suggests. This becomes clear when delving deeper into this important sub-section.

Figure 19 shows key aspects that are included in the findings aggregate, as presented in Figure 18 (bar chart 3). Two key aspects related to the degree to which the findings were evidence-based: the first one with a dichotomous assessment (yes/no, see first bar chart on the left, Figure 19), the second one with a more detailed 4-scale rating (see second bar chart on the left, Figure 19). The findings were generally rated as evidence-based (73, 91%) and also, the more detailed assessment on the quality level of evidence-based findings is moderately positive (37, 46% satisfactory and 4, 5% (very) good). Need for improvement mainly stems from **triangulation of data** (39, 49% need for improvement, 15, 19% inadequate), **presenting only findings** in this section but no conclusions (46, 58% fail), **discussion of attribution** (61, 76% fail), and **discussion of confounding factors** (77, 96% fail) (see Figure 19). Taking this together, we can sum up that **despite the overall limited quality of the evaluation reports, findings are somewhat reliable** as they are at least by far evidence-based and as the quality level of evidenced-based findings is at least satisfactory for half of the reports.

**Figure 19: Quality of findings (n=80)**



Source: own statistics based on analysis of reports

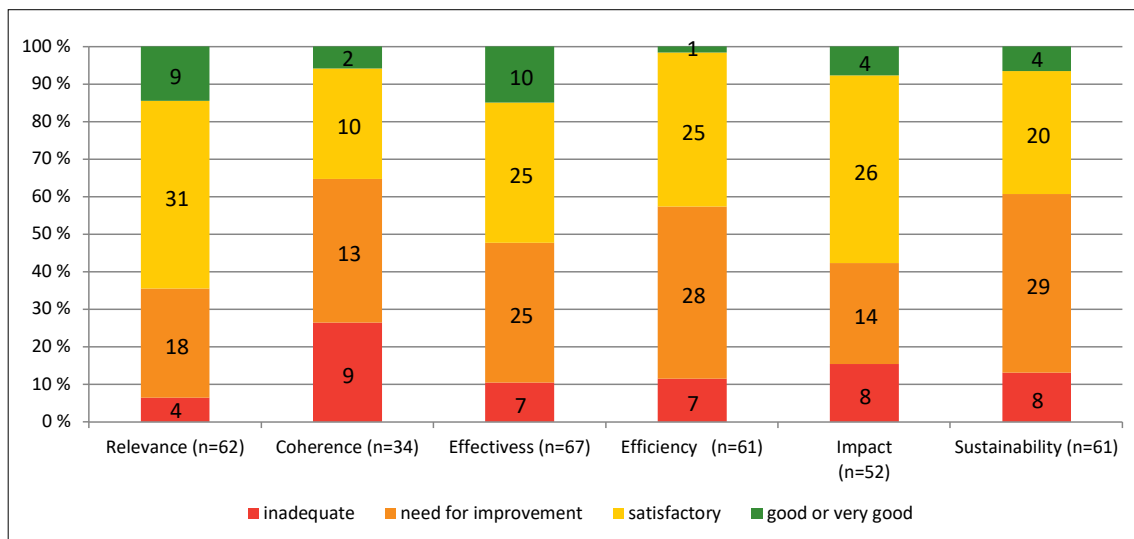


To shed further light on the findings, Figure 20 displays appropriate capturing of OECD DAC criteria, which is also included in the findings aggregate as presented in Figure 18 (bar chart 3, for more details on the aggregate, see Annex 6: Quality Assessment Tool for Reports and ToR). It shows that on the one hand, **relevance** 🎯, **effectiveness** 🚀, and **impact** 📉 are **satisfactorily or better captured in more than 50% of the cases**. However, on the other hand, **coherence** 🗑️, **efficiency** ⚙️, and **sustainability** ♻️ are **less appropriately captured**.

Coherence, being relatively new, is least discussed among the OECD DAC criteria, with less than half of the reports (34, 43%). Among those 34 reports, 22 (65%) are rated less than satisfactory. Efficiency and sustainability are both discussed in 61 reports, and approximately 40% of those are rated satisfactory or better.

In contrast, relevance received the best score among the OECD DAC criteria, with half of the 62 reports discussing this criterion (31) rated as satisfactory and 9 (15%) as (very) good. Out of the 52 reports referring to impacts, half (26) received a satisfactory score, and 4 reports (8%) were rated as (very) good. A little more than half of the 67 reports addressing effectiveness (35, 52%) was rated satisfactory or better.

**Figure 20: Appropriate capturing of DAC criteria**



Source: own statistics based on analysis of reports

Further details on all sections of the quality assessment and aspects assessed under each section are provided in Annex 11: Quality Assessment on single sections of the reports.

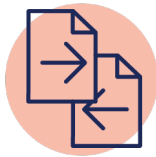


## 4.3. Differences, trends and evaluation capacity gaps (EQ4)

### 4.3.1. Differences and trends to the previous metaevaluation

#### Highlights of the section addressing EQ 4:

- A static comparison, replicating the quality aggregate used in the previous metaevaluation, reveals no differences in the overall report quality between the current and the previous-meta evaluation.
- A dynamic comparison of the overall report quality, acknowledging additional quality aspects, unveils deterioration.
- Likely, evaluators had not yet adapted to increased evaluation standards, yielding to a decline in overall report quality in the dynamic comparison.
- Deterioration may be caused by the inclusion of conceptually different evaluations of funds and organisations into the recent metaevaluation.

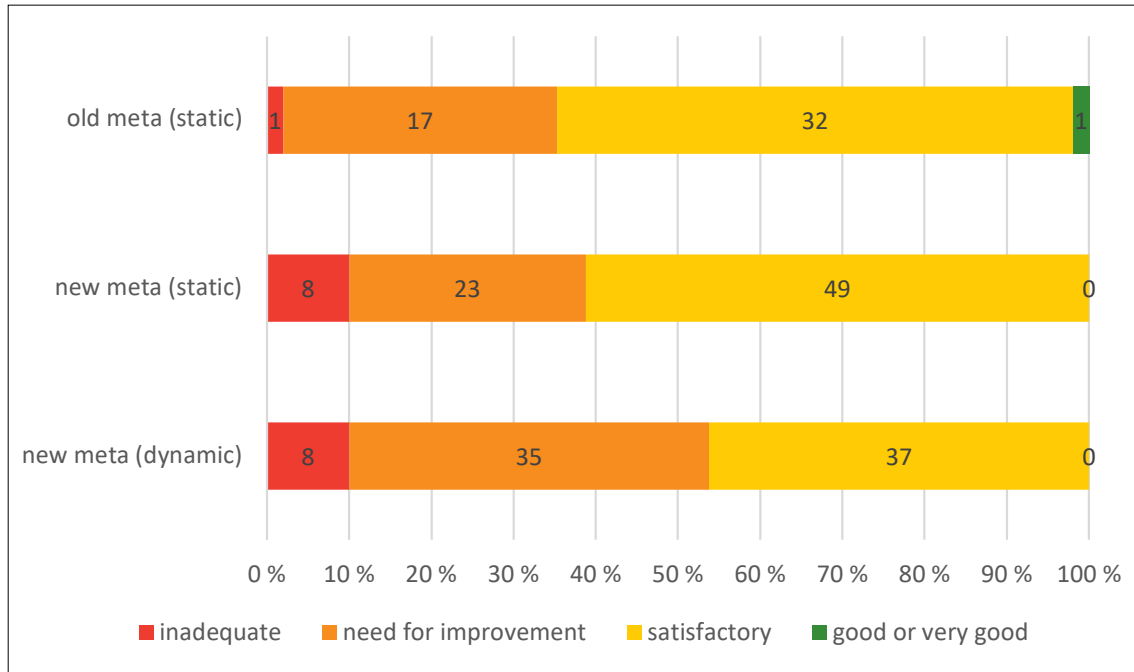


We compared the **overall report quality between the current and previous metaevaluation** to investigate potential quality shifts. In a **static comparison**, the aggregate of overall report quality was calculated the same way as in the previous metaevaluation to maximise comparability. As shown in Figure 20, the distributions over single quality categories closely match each other (see first and second bar chart). Thus, **no difference** between the two metaevaluations is observed.

In contrast, a **dynamic comparison** draws a different picture. The dynamic aggregate additionally reflects newly assessed aspects emanating from increased evaluation standards, for example, coherence and stronger anticipation of CCOs (see chapters 2.1, 4.1, and Annex 5: Methodological details for details). Figure 21 accordingly suggest a **deterioration** of overall report quality under the current metaevaluation (see first and last bar chart). There is a lower share of evaluations of satisfactory quality in the current metaevaluation (37, 46%) as in the previous one (32, 63%), a higher share of evaluations with a need for improvement (current: 35, 44%; previous: 17, 33%), and a higher share of evaluations of inadequate quality (current: 8, 10 %; previous: 1, 2%). Also, in the current metaevaluation no case has been assessed as (very) good, whereas there was one in the previous evaluation.



**Figure 21: Quality aggregates of the previous and current metaevaluation**



Source: own statistics based on analysis of reports

A plausible **reason for the deterioration** is that the quality assessment has been adapted to **raised evaluation standards**, while the evaluators and thus their reports, at least by the time of writing, still lagged behind in these areas. Other factors, like, for example, staff rotation, are not plausible as to the best of our knowledge they have not considerably changed since the previous metaevaluation. Additionally, or alternatively, it is also possible that this downwards shift is driven by the inclusion of atypical cases (i.e., evaluations of funds or organisations), which have been part of the current but not of the previous metaevaluation.



### 4.3.2. Gaps regarding MFA's evaluation capacity

#### Highlights of the section:

- The quality assessment unveiled capacity gaps inside and outside MFA. About one-third of the ToRs and roughly half of the evaluation reports need improvement, while systematic consideration of CCOs and HRBA is mostly missing in both.
- Deterioration of report quality when anticipating increased quality standards provides hints for increasing capacity gaps.










The second part of EQ4 asks about potential **gaps regarding MFA's evaluation capacity**. There is no direct assessment to answer this question, but the following summary of results can be used to derive some insights in this regard. The quality assessment disclosed considerable quality flaws: About a third of the ToRs and roughly half of the evaluation reports are assessed as in need for improvement (see chapters 4.1.1 and 4.2). Thereby it turned out that adequate consideration of CCOs and HRBA is a serious challenge for commissioners and evaluators. Moreover, the trends between the previous and current metaevaluation point to lower overall report quality when taking increased evaluation quality standards into account (see chapter 4.3.1).




Those weaknesses suggest that the evaluation capacity inside and outside MFA does not yet tap its full potential. At the evaluators' level, the findings underline capacity gaps, particularly in the fields of methodological rigour, provision of high-quality, evidence-based findings, appropriate capturing of OECD DAC criteria and adequate consideration of CCOs and HRBA. At the MFA level, the findings unveil capacity gaps regarding the drafting of high-quality ToRs and regarding the enforcement of its own guidelines to avoid approval of low-quality evaluation reports (both further elaborated in chapters 7 and 8).




## 5. Findings of the Summative Analysis

After assessing the quality of 80 evaluation reports and 70 associated ToRs, the summative analysis focuses on the content of the evaluation reports and, thus, on the **quality of the interventions**. We aggregate the assessments provided in each evaluation report which passed minimal quality standards (i.e., 72 reports). Note that we **did not re-evaluate** the interventions but rather **synthesise the findings by the evaluators** as presented in their evaluation reports. Hence, a fraction of Finland's development cooperation comprising selected bi-, multi- and multi-bilateral interventions is assessed based on reliable decentralised mid-term and final evaluation reports.

In response to **EQ8**, chapter 5.1 offers an assessment of the **overall quality of Finnish development cooperation**  based on reliable decentralised evaluation reports and points to **major strengths and challenges**. Further, it touches upon summary findings in response to **EQ5** on the quality of Finnish development cooperation according to the OECD DAC criteria relevance , coherence , effectiveness , efficiency , impact  and sustainability . In Annex 12: Interventions' Quality on Single OECD DAC Criteria, we delve more into detail and provide findings on the quantitative assessment of single sub-aspects under each criterion and on aggregated results of the qualitative content analysis on underlying reasons for their assessment provided by the evaluators.

Chapter 5.2 responds to **EQ6** on the consideration of the **cross-cutting objectives (CCOs)** gender equality , non-discrimination  and climate sustainability , the **human rights-based approach (HRBA)**, and **policy priority areas (PPAs)** as overarching aspects.

In chapter 5.3, we synthesise **lessons learnt** (5.3.1) and **recommendations**  (5.3.2) **drawn by the evaluators** to reply to **EQ7** on major recommendations and lessons learnt to improve Finnish development cooperation.

In the end, chapter 5.4 responds to **EQ9** on key **differences** and possible **trends compared to** the overall results of the **previous metaevaluation** .





## 5.1. Overall quality, strengths and weaknesses of the interventions (EQ8 & EQ5)

### Highlights of the section:

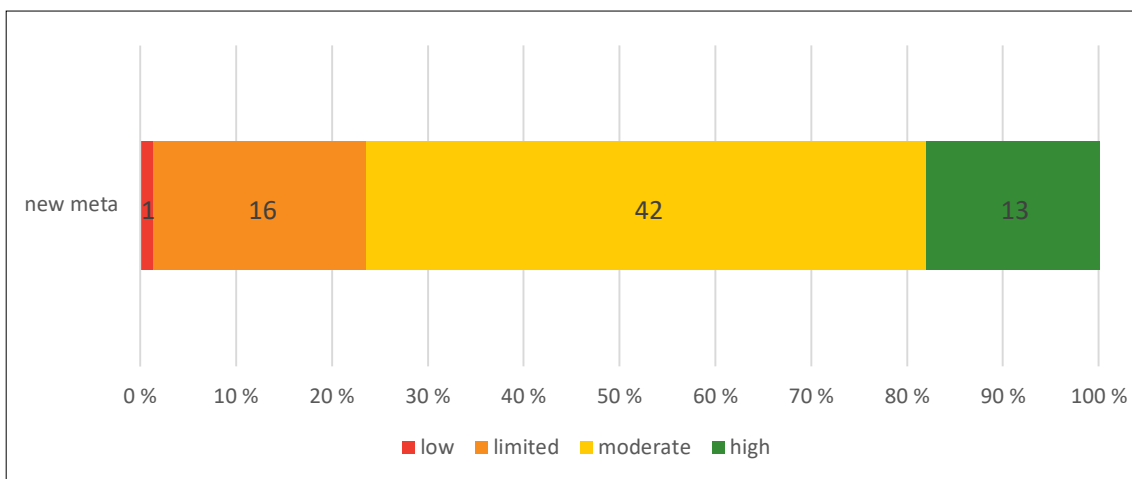
- Note: The metaevaluation is not a re-evaluation but transferred and aggregated evaluators' individual assessments.
- Overall, the quality of bi-, multi- and multi-bilateral interventions under consideration is quite good.
- About 20% of the interventions possess high, roughly 60% moderate and around 20% limited quality.
- Relevance is a strength of Finnish development cooperation, followed by effectiveness and efficiency.
- Sustainability leaves the greatest room for improvement, followed by coherence and impact.



In the course of the summative content analysis, the **overall quality of intervention** was calculated as the sum of the assessments of all OECD DAC criteria captured in the evaluation report, divided by the total number of OECD DAC criteria covered.




As Figure 22 shows, the distribution of the overall intervention quality is **quite positive**, with a large share of interventions that have been assessed as good or very good by the evaluators (13, 18%) or satisfactory (42, 58%). Some interventions are of limited overall quality (16, 22%), and only one case is considered inadequate.

Figure 22: Quality of the intervention (n=72)



Source: own statistics based on analysis of reports



To delve into detail, the quality of the interventions is **disaggregated at the level of the six single OECD DAC criteria** as presented in Figure 23. When comparing the bar charts of the quality of single DAC criteria, **relevance**  is the criterion with the **largest share of high-quality assessments**. Out of the 61 reports that have assessed the relevance of the intervention, 36 show high relevance (59%), 23 moderate relevance (38%), 2 limited relevance (3%), and none of low relevance. This is **followed by effectiveness**  **and efficiency** , which are largely assessed as being of moderate quality (35 out of 66, 53% and 29 out of 54, 54%, respectively) to high quality (11 out of 66, 17% and 8 out of 54, 15%, respectively).




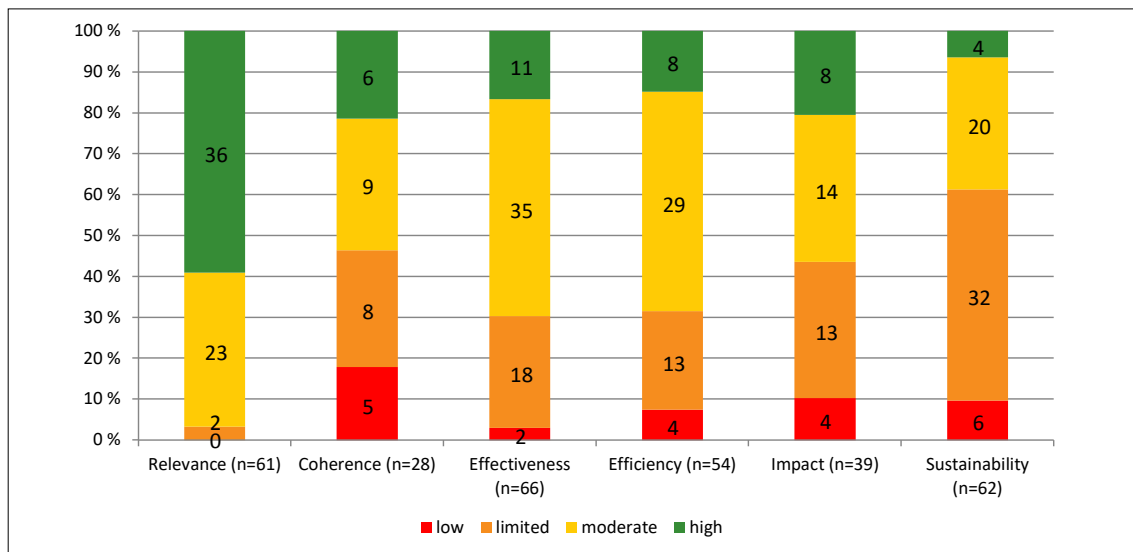
By contrast, **sustainability**  has the **largest share of assessments indicating limited quality** (32 out of 62, 52%) and low quality (6 out of 62, 10%), while fewer cases are assessed as moderately sustainable (20 out of 62, 32%), or highly sustainable (4 out of 62, 6%). Moreover, **coherence**  **and impact**  **are also often constraints** as only about half of the cases are being assessed as of moderate quality (9 out of 28, 32% and 14 out of 39 and 36%, respectively) or high quality (6 out of 28, 21% and 8 out of 39, 21%, respectively), whereas the other half is of low quality (5 out of 28, 18% and 4 out of 39, 10%, respectively) or limited quality (8 out of 28, 29% and 13 out of 39, 33%, respectively). More details on in-depth quantitative and qualitative content analyses of the intervention's quality on single OECD DAC criteria are provided in Annex 12: Interventions' Quality on Single OECD DAC Criteria.

Figure 23: Quality on single DAC criteria



Source: own statistics based on analysis of reports



Taken together, we can summarise that **relevance is a strength** of Finnish development cooperation, followed by larger distance by effectiveness and efficiency. Relevance is high or moderate for 97% of the interventions (59 out of 61). Mostly positive results were also found for effectiveness and efficiency, which are assessed moderate or high in roughly two-thirds of the interventions (46 out of 66, (70%) and 37 out 54, (65%), respectively).



Regarding the **challenges** emerging from the summative analysis, the **greatest room for improvement is found for sustainability**, followed by impact and coherence. Nearly two-thirds of the interventions were assessed as of limited or low sustainability (38 out of 62, 61%) and almost half of the interventions assessed revealed limited or low coherence (13 out of 28, 46%) and



impact (17 out of 39, 44%). However, it should be noted that impact and coherence were often not assessed; thus, data availability poses a limitation on sample representativeness of results.

With respect to typical factors for success and challenges, the metaevaluation cannot draw a concise picture. In annex 12, a qualitative content analysis of underlying reasons for positive and negative assessments of the evaluators is provided. It shows that there are multiple factors facilitating or hampering interventions' success. It is not surprising that interventions seem to be highly context-specific and that identifying a blueprint yielding to high-quality interventions is impossible. However, the findings presented in annex 12 lay the foundation for further in-depth analysis on single OECD-DAC to gain a better understanding of potentially crucial aspects.

Overall, about three-quarters of the 72 interventions considered were found to be of moderate or high quality. In turn, the overall quality of the bi- and multilateral interventions is found to be limited in one-quarter of the interventions disclosing that some interventions are in their entirety weak.

## 5.2. Consideration of cross-cutting objectives (CCOs), human rights-based approach (HRBA) and policy priority areas (PPAs) (EQ6)

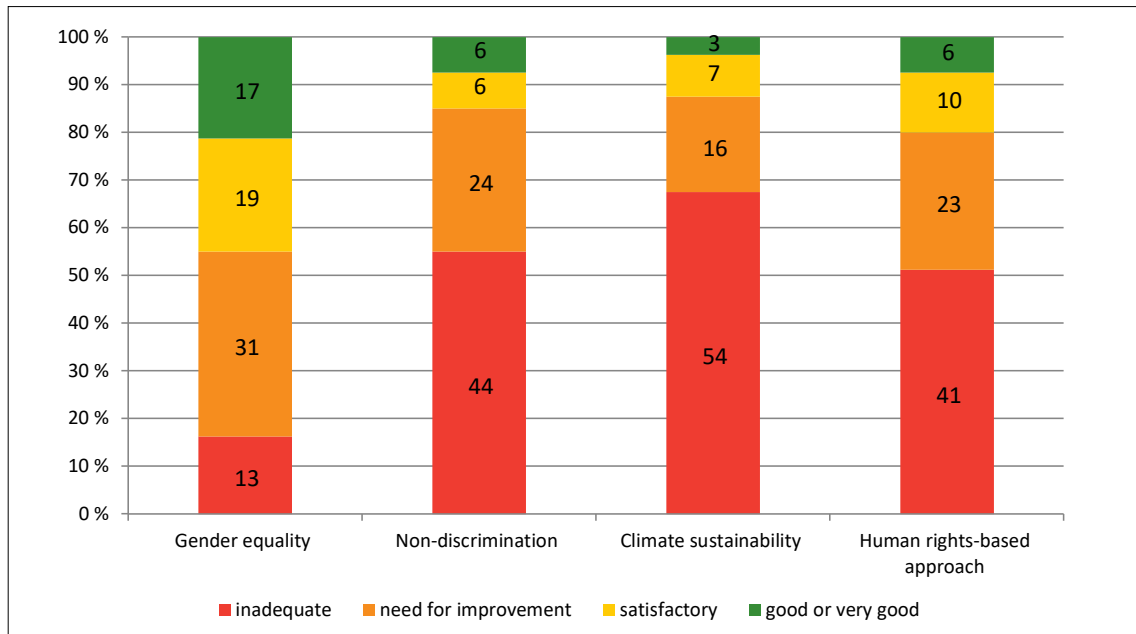
### Highlights of the section:

- Gender equality is better mainstreamed than other CCOs and HRBA.
- Non-discrimination, climate sustainability and HRBA are rarely considered by evaluators. Thus, interventions' quality in this regard is not assessable.
- Several interventions yield to promising gender results comprising gender policy, women's economic and social empowerment, women's leadership, women's health, and women in peace and security.
- Two reports disclose adverse effects on women and girls.
- Gender equality emerges as a relative strength of Finnish development cooperation. Evidence provides first hints that other CCOs and HRBA may be a weakness.
- The overall quality of the interventions does not differ among different policy priority areas (PPAs).

Cross-cutting objectives (CCOs) and the human-rights based approach (HRBA) are deeply rooted in Finnish development policy (see chapter 3.1). Thus, evaluators are expected to acknowledge them in their analyses, particularly in the findings, conclusions, and recommendations sections of the reports. Despite this, from the quality assessment as displayed in Figure 24, we know that the **majority of evaluators failed** to do so for non-discrimination (55%, 44), climate sustainability (68%, 54) and HRBA (51%, 31). As we cannot differentiate whether the interventions themselves were not mainstreamed or whether the evaluators simply failed to pay attention to those CCOs and HRBA in their analysis, we must **refrain from the summative analysis**.



**Figure 24: Integration of CCOs and HRBA (n=80)**



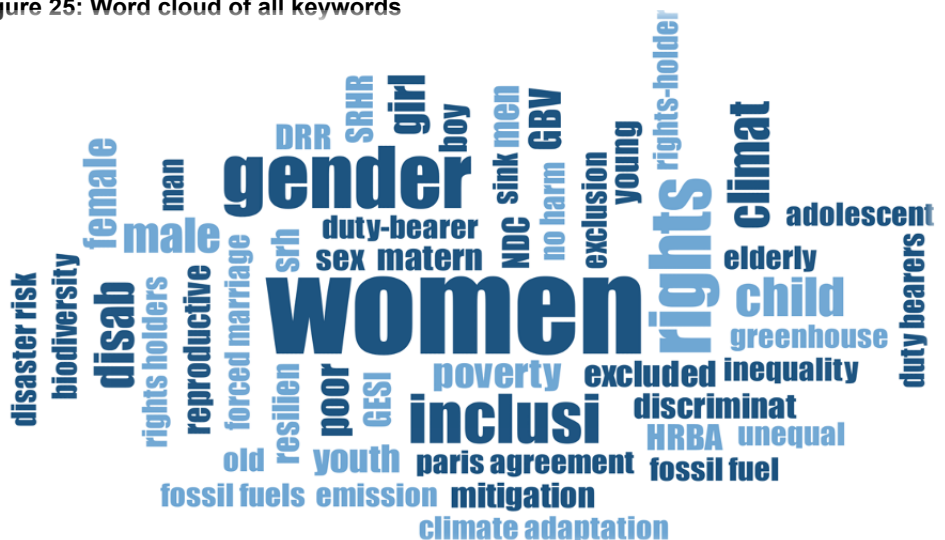
Source: own statistics based on analysis of reports



Figure 24, however, also draws a different picture with respect to **gender equality**. Nearly half of the reports acknowledge this cross-cutting objective satisfactorily (24%, 19) or (very) good (17, 21%). The share of inadequate recognition of gender equality (16%, 13) is considerably lower. This is plausible given the importance and the longstanding tradition of gender equality within Finnish development cooperation. Yet less than half of the reports contain an in-depth analysis.

Quantitative keyword analysis underlines this observation. The word cloud of all keywords in Figure 25 illustrates a high variance among and within CCOs while highlighting gender equality being clearly **at the forefront**.

**Figure 25: Word cloud of all keywords**



Source: own analysis of evaluation reports



Table 2 further discloses that evaluators sometimes do not even mention any of the keywords related to a particular CCO. This holds for 22% (14) of the reports regarding climate sustainability, for 17% (11) regarding HRBA and for 3% (2) regarding non-discrimination. At least a complete neglect of gender-related keywords was not observed.



When keywords are mentioned, their frequency varies substantially. For example, for gender, it ranges between 3 and 601 per report, with half of the reports mentioning keywords 54 times (i.e., median). The share of reports having a higher-than-average frequency of keywords per report (80 hits) is about one third (32%, 21) for gender equality. The share of higher-than-average frequency of keywords per report is about a quarter for non-discrimination (26%, 17), climate change (25%, 16), and HRBA (23%, 16) lower. The same holds true for the average number of hits per report: non-discrimination (23), climate change (13), and HRBA (18). For further details on the keyword analysis and underlying dictionaries, please refer to Annex 13.

**Table 2: Results on the frequencies of keyword groups in evaluation reports (n=72)**

Cross-cutting objective (group of keywords)	# of reports that did not mention any of the keywords	% of reports that do not mention any of the keywords	Smallest # of hits per report	Highest # of hits per report	Median	Average number of hits per report	# of reports above the average per report	% of reports above the average per report
Gender equality	0	0%	3	601	54	80	21	32%
Non-discrimination	2	3%	0	231	8	23	17	26%
Climate change	14	22%	0	132	3	13	16	25%
HRBA	11	17%	0	248	7	18	15	23%
Other	3	5%	0	150	12	21	16	25%
All keywords	0	0%	7	694	117	155	26	40%

Source: own analysis of findings, conclusions, recommendations sections of evaluation reports



Delving deeper into gender equality reveals that **positive gender results** in the following fields were obtained:

- Influencing a gender policy or law and gender in national governance in general (seven cases)
- Improved women's economic empowerment or jobs (seven cases)
- Increased women's participation in political decision-making or social empowerment (six cases)
- Strengthened women's leadership (three cases)
- Improved health, family planning, and reproductive health (three cases)
- Women in peace and security (three cases)

**Overall intervention quality does not differ among policy priority areas (PPAs).**

Other topics described – but fewer cases or with less detail provided by the evaluators – included gender mainstreaming in general terms, services for vulnerable women (such as shelters or support for women with disabilities), gender-based violence, awareness-raising, and training to women, influencing UN Women, and land rights. Negative gender results were rare and only reported in two evaluations. One case referred to the construction of toilet facilities that were inadequate or even dangerous and did not meet the needs of adolescent girls. In the other example, a district-heating project was criticised for a general lack of gender sensitivity; however, the analysis lacks depth and granularity.

As mentioned earlier, several evaluation reports did not discuss gender results at all. Many of them mentioned the topic only so briefly that it was not possible to extract concrete achievements. Thus, gender results must be interpreted with caution as they are incomplete given the lack of (adequate) assessments by the evaluators.

Beyond CCOs and HRBA, a look at the **policy priority areas (PPAs)** disclose that 14% of the interventions (10) within our sample fall thematically outside any specific PPA. Further, a comparison of overall intervention quality did **not reveal any quality differences among different PPAs** (see statistical annex 15 for Mann-Whitney tests).

### 5.3. Differences and trends to the previous metaevaluation (EQ9)

**Highlights of the section:**

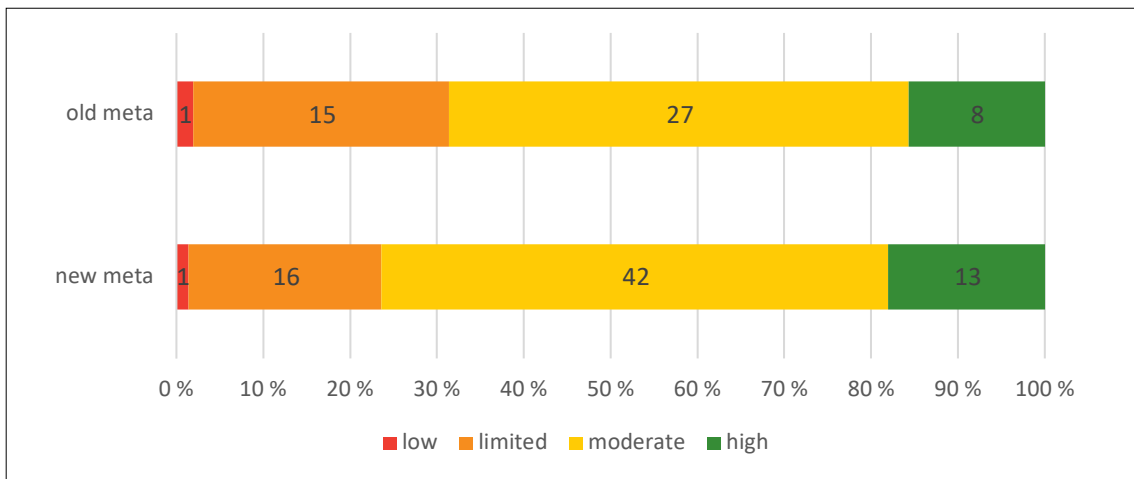
- The overall quality of the interventions did not change at large from the previous metaevaluation.
- However, direct comparisons at the intervention level are not possible as both metaevaluations looked at different interventions and the sample composition varied with respect to intervention budgets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



With respect to key differences and possible trends between the overall quality of the recent and the previous metaevaluation, some limitations should be noted: First, it is not possible to directly compare the results at the intervention level as the previous metaevaluation included a different set of interventions. Second, the sample composition varies between both metaevaluation in terms of intervention budgets and thematic sectors of the interventions. Therefore, we can only put the results of the above-presented summative analysis into perspective with earlier findings.

Figure 26 shows the bar charts for overall intervention quality under the previous and the recent metaevaluation to carefully identify possible trends. Figure 26 hints at a slight difference between the shares of interventions of limited quality: However, as direct comparisons are not valid, we conservatively wrap up, this suggests that the overall level of intervention quality observed under the previous metaevaluation was at large maintained.

**Figure 26: Possible trends on overall intervention quality**



Source: own statistics based on analysis of reports



## 5.4. Lessons learnt and recommendations by evaluators (EQ7)

### 5.4.1. Lessons learnt presented in the evaluation reports

#### Highlights of the section:

- 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.



A total number of 171 **lessons learnt** were identified among the sample of 72 evaluation reports. 38 reports (53% of the sample) contain lessons learnt, whereas 34 reports (47%) do not feature them. As displayed in methodological Annex 5: Methodological details, we only considered lessons that were either formulated according to the OECD DAC definition or lessons formulated in a way that generalisation with a reasonable level of expert judgement was possible. So-called lessons that required arbitrary interpretation to go beyond intervention-level recommendations were not considered.

Table 3 gives an overview of the number of reports containing lessons learnt, sorted by theme. It shows a large variance regarding the distribution of frequencies. The **most frequent themes are capacity and planning**, appearing in 14 reports (19%) each, followed by coherence in 12 (17%). The table also reveals that no single intervention field was predominantly prevalent appearing in more than half of the reports. Thus, **no typical lessons learnt** could be derived **from the majority of reports**.

The themes addressed by lessons learnt which have been identified in at least 10% of the reports, were further analysed and are presented as follows.





**Table 3: Number of reports including lessons learnt categorised under different themes**

LESSONS LEARNT	NUMBER OF REPORTS	IN % OF ALL REPORTS	NUMBER OF LESSONS
Capacity	14	19%	16
Planning	14	19%	19
Coherence	12	17%	18
Financial aspects	9	13%	9
Others	9	13%	14
Participation	9	13%	11
Sustainability	8	11%	10
Aid effectiveness	6	8%	7
Communication	6	8%	7
Gender	6	8%	8
M&E	6	8%	12
Management	6	8%	7
Relevance	6	8%	8
Personnel	4	6%	4
Time	4	6%	7
Efficiency	3	4%	3
Equipment	3	4%	3
Impact	3	4%	4
Scope	2	3%	3
Effectiveness	1	1%	1
<b>Total number of reports</b>	<b>72</b>	<b>100%</b>	-
<b>Reports with lessons learned</b>	<b>38</b>	<b>53%</b>	-
<b>Reports without lessons learned</b>	<b>34</b>	<b>47%</b>	-
<b>Total Number of lessons learned</b>			<b>171</b>

Note: Other lessons learnt refer to specific technical aspects which could be aggregated at a general level.

Source: own analysis of evaluation reports

Lessons referring to the **capacity** of target groups and beneficiaries point mainly to

- the mere importance of capacity development for the success, impact and sustainability of interventions, often with a particular focus on the capacities of local partners and government institutions,
- the need of a comprehensive and systematic capacity development approach, based on a needs assessment and performance-oriented monitoring, or
- the success of specific capacity development approaches, such as peer-to-peer training, mentorship, or the inclusion of diaspora experts in the training of local partners.

A selection of such lessons learnt is presented in Box 1.



### Box 1: Examples for lessons referring to the theme capacity

*“Capacity building [is] seen as crucial: Counterparts of both projects, SUFORD-SU and SNGS/- EP [Strengthening National Geographic Services in Lao PDR/-Extension Phase], emphasised repeatedly that Finland’s TA support to building staff capacity and installing systems at all levels was crucial to the achievements of the projects. The same was expressed by EMSP [Environment Management Support Programme] stakeholders at MoNRE [Ministry of Natural Resources and Environment]. At all involved institutions, DoF, DoFI [Department of Forest Inspection], DEQP [Department of Environmental Quality Promotion] and NGD [National Geographic Department], Finland is generally held in high regard for the many trainings it provided to GoL [Gov. of Lao] staff. The focus on capacity building at GoL institutions was an invaluable support to general development and potentially lets Finland cooperation stands out positively among the donor community.” (Report No. 17a,b, p. 143)*

*“Capacity development is a challenging, sensitive and for some stakeholders still vague process, which requires a consistent and coherent and systematic approach. Therefore, it should benefit of 1) complex capacity needs assessment, 2) institutional/organizational capacity development plan, 3) adequate human and financial resources allocation, 4) involve a clear step-by-step implementation and, finally 5) performance-oriented and systematic monitoring.” (Report No. 37, p. 36)*

*“CF [Community Forest] works and has created great interest among many partners in the country. Adequate time, human resource to provide necessary technical capacity and guidance is needed to ensure sustainability.” (Report No. 8, p. 59)*

Lessons learnt addressing the theme **planning** typically encourage

- adapting projects to local situations, e.g., by conducting a proper context analysis and stakeholder mapping,
- conducting a comprehensive risk analysis that covers all relevant types of risks and developing a project design acknowledging the implications derived from the risk analysis,
- flexible arrangements which allow adaptive management,
- a thorough selection of the right partner organisations in terms of capacity, networks, contextual knowledge, and local acceptance, and
- formulating a theory of change with clear outcomes and impacts.

Some examples of such lessons learnt are presented in Box 2.



## Box 2: Examples for lessons learnt in the theme planning

*“The important lessons learnt is [...] adapting interventions to the local situation in the country and plan them in early stage in the process together with constituents. To some extent, the DW II Project showed that it allows for experimentation and exploration, followed only then by implementation. At this point in time, this is a rather rare strength of a development project. Here, the basic assumption is that if labour management systems are facilitated to change from within, rather than through a number of direct and distorting interventions, better and more sustainable results can be achieved.” (Report No. 47, p. 47)*

*“Usage of well-chosen partners is a strength that could be tighter implemented. For the implementation of the Eid bi Eid programme UN Women has been able to identify partner organizations that are well-placed and well-fitted for the implementation of the different sub-components. However, UN Women has left much leeway for the single partner organization to implement the components they oversee. This has meant that the programme has not utilized a harmonized definition of central programme components such as vulnerability definitions which has meant lack of cohesion across programme components and an opportunity lost to influence these organizations to develop a more sophisticated selection criterion.” (Report No. 28, p. 52)*

*“Risk analysis not for itself: The Project Appraisal Document of SUFORD-SU has a section on risk rating and analysis in. While the section is there, it has not really touched upon the critical risks regarding counterproductive GoL [Gov. of Lao] policies or pressures created by non-forest sectors and actors on sustainable forest management and project outcomes. Project designs will benefit from a risk assessment that is comprehensive and analyses all potential risks. This will contribute to improved risk management and mitigation strategies that obviously would need to be well reflected in the project strategy” (Report No. 17a, p. 143)*

*“Decide a desired future — it is important to identify the desired long-term goals of the policy and strategy capacity building intervention and then work back from these to identify all the conditions (outcomes) that must be in place for the goals to occur. Future projects demand a Theory of Change (ToC) as starting point. The ToC should lead to a clear and fully endorsed project log-frame or results-based management framework. Country-level interventions should also incorporate such a ToC, log-frame and Monitoring and Evaluation framework and be accepted and used as key administrative and compliance tools.” (Report No. 58, p. 70)*



Lessons learnt in the theme **coherence** refer to

- the importance of decent coordination between donors, implementers, and partners to support successful, efficient, and complementary project implementation in the same region or sector, and to avoid unclear responsibilities and duplications of activities,
- the crucial role of networks when projects are encouraged to (i) either choose partner organisations with sound networks in the region, (ii) to become themselves part of existing regional/national policy or sectoral networks, or (iii) creating networks through the establishment of new partnerships or the organisation of exchange platforms when they are still lacking, and



- the need to align project concepts with national or international policies and strategies such as national development plans, the Agenda 2030 or UNFCCC.

A selection of such lessons learnt is presented in Box 3.

**Box 3: Examples for lessons in the theme coherence**

*“Donor-Implementor-Coordinator relations are very important for the success of any project. Poor relations will create stress on implementation, and results cannot be achieved. Therefore, strong coordination, communication systems, relationship development among stakeholders is very important.” (Report No. 4, p. 44)*

*“In some countries there is a number of donors available that are funding different interventions. In this case, careful donor mapping and coordination would ensure that the TA is efficiently aligned with capital investments in order to deliver maximal impact. In countries with limited external funding opportunities (i.e. Zimbabwe), WCO [World Customs Organization] and even yet the ESA [Eastern and Southern Africa] II could help develop a pipeline of capital/modernization projects that would then be ready for funding either from domestic or external sources.” (Report No. 12, p. 36)*

*“Engaging regional structures as the League of Arab States with civil society organisations and structures (such as JNCW) is a positive step in fostering coordination at the regional level. Further efforts and attention are required for this strategy to lead to concrete achievements.” (Report No. 29, p. 45)*

*“There are many opportunities for the MIDA FINNSOM Health and Education project to link with on-going initiatives in the health and education sector by regular attendance of aid coordination meetings and building existing partnerships both in Somalia and Finland. Improved coordination in the health and education sector forums is important, because other non-MIDA health and education programmes are supporting organizational development to build self-sustaining institutions.” (Report No. 32, p. 60)*

Lessons learnt in the theme of **participation** relate to

- identifying ways of increasing the participation of marginalised groups,
- the importance of involving all relevant stakeholders in project planning (e.g., in the development of the Theory of Change) and during implementation (e.g., through a steering group),
- acknowledging the crucial role of participation of the target group and the final beneficiaries for the creation of ownership, and
- fostering the use of local languages instead of English to support the participation of beneficiaries and marginalized groups.

Exemplary lessons learnt are provided in Box 4.



#### Box 4: Examples for lessons in the theme participation

*“Strong involvement of key stakeholders from government, civil society, traditional leaders and local communities.” (Report No. 8c, p. 51)*

*“Ideally, the final project design is done in a stakeholder workshop (ideally before project approval but alternatively during the inception phase) where a Theory of Change (or problem tree or similar approach) is developed, from which a logical framework can then be derived jointly by all stakeholders. Such a workshop will require a good facilitator with a strong expertise in Theories of Change and log frames.” (Report No. 24, p. 59)*

*“Equity as a development issue – integration of equity issues in the policy and strategy development is critical. However, as this evaluation showed, the equity issues need to go beyond gender – which also needs to be better understood. It should include people with disabilities and those with resources limitation to effectively partake in the benefits of digital economies. This implies that the participation of the marginalized groups should be seen both from development and equity angles. While ICT access is an important equity equation, other aspects such as skills, use and social appropriation by marginalized groups should be given attention. (Report No. 58, p. 70)*

The lessons in the theme **financial aspects** often refer to

- the funding of local partner organisations, pointing to the requirement of funding on a long-term basis rather than on short-term,
- the benefit of funds that are not bound to a certain task or delivery but can be used by partners according to their own priorities to sustain long-term institutional development.
- the importance of transparency and accountability regarding the use of funds or introducing particular innovative funding models.

Some illustrative examples are presented in Box 5.



#### Box 5: Examples for lessons in the theme financial aspects

*“Financial support to independent research and similar institutions should not be earmarked. It should be institutional/core-support, to enable the organisation to develop its own (and owned) agenda. Usual grant making processes (with minutiae’s control) would have killed IESE [Institute of Social and Economic Studies] at the start. Untied support is not easy for many donors as it certainly brings financial and political risks that must be accepted.” (Report No. 54, p. 34)*

*“The most effective support to environmental CSOs is broad organisational development support, that allows them to build their capacities and resources for implementation of activities that are defined by themselves rather than by detailed donor guidelines.” (Report No. 8b, p. 59)*

*“Fiscal accountability: All projects like this carry an inherent risk in financial management, whether as a result of inadvertent mismanagement or corruption. Provision by FI [Finland] to TVLA [Tanzania Veterinary Laboratory Agency] of training on standard accounting processes could mitigate this risk.” (Report No. 50, p. 7)*



Lessons learnt in the theme **sustainability** commonly referred to several important factors that ensure the sustainability of interventions, including

- the close cooperation with national partners who have the capacity to continue the project activities independently to consolidate results over time,
- the institutionalisation of results, e.g., through institutional development and the support of new laws and regulations,
- sufficient time to implement preconditions for sustainability, and
- financial sustainability: it is mentioned that the willingness or capacity of partners to allocate sufficient resources to the maintenance and continuation of achievements is critical for sustainability.

A selection of such lessons learnt is presented in Box 6 below.



#### Box 6: Examples for lessons in the theme of sustainability

*“Conducive government policies but lack of implementation budget: Examples from both evaluated projects have shown that projects contributed to the development of good and conducive policies (SUFORD-SU), to the improvement of the legal frameworks (SNGS [Strengthening National Geographic Services in Lao PDR], surveying decree), and to the development of viable models and systems (SUFORD VLDG or payment for work; SNGS mapping standards). Due to either a lack of Government budget or conflicting priorities, recurrent budgets have not been made available by the Government to support these models and systems in the future, thereby calling the financial sustainability of the achieved results into question. The lesson learned from both SUFORD-SU and SNGS/-EP [-Extension Phase] is that it would be necessary to assess partner government ownership to a project also through the commitment to provide recurrent funding to key activities after project closes down.” (Report No. 17a,b, p. 143)*

*“Institutional development is critical to sustainability of achieved results: If there is a key lesson to be learned from the SNGS and SNGS-EP, it is that technical inputs are important, but alone cannot solve development problems. The neglect of institutional aspects in the design of the SNGS has created a situation where the very good and well appreciated results may not be sustained for much longer. In evaluating the SNGS results it has become evident that institutional aspects are critical to the sustainability of project results. Seen from the door-perspective this signifies that investment in technology-oriented institutions can well pay off in the long term, but only if other aspects such as sustainable funding and institutional arrangements are properly addressed early on. Without sufficient support to and development of the counterpart institutions, the development input remains a one-off contribution with the country not being able to renew or replicate these inputs in the future. (Report No. 17a,b, p. 143)*

*For sustainable project results, it is crucial to include, right from the conception, close cooperation with national partners (such as Forestry Institutes) to ensure continued support to project beneficiaries once the project has phased out. In the case of Nicaragua, for example, the strategic partnership with the national forestry institute (INAFOR) provided for (1) good field-level knowledge allowing to identify target groups and beneficiaries; (2) opportunities to ensure access rights and long-term benefits through the formal registration of resources and assets (Eucalyptus plantations, biomass-supplying farms); and (3) continued advisory services to project beneficiaries, after end-of-project, with support from other donors. (Report No. 35a, p. 22)*



## 5.4.2. Recommendations identified in the evaluation reports

### Highlights of the section:

- 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.



Recommendations are much more common than lessons learnt and are present in almost all of the reports (96%, 69). In total, 1,123 recommendations were drawn by the evaluators. Table 4 provides an overview of the number of reports containing recommendations and their total number sorted by theme.

In **around half or more of the reports**, recommendations address the following themes: **M&E, coherence, planning, sustainability, management, financial aspects, and/or capacity**. Thus, those themes comprise typical recommendations which were further analysed.

**Table 4: Number of reports including recommendations categorised under different themes**

RECOMMENDATION	NUMBER OF REPORTS	IN % OF ALL REPORTS	NUMBER OF RECOMMENDATIONS
M&E	49	68%	105
Coherence	48	67%	117
Planning	45	63%	109
Sustainability	39	54%	64
Management	37	51%	108
Financial Aspects	35	49%	65
Capacity	35	49%	76
Communication	32	44%	72
Gender	29	40%	52
Personnel	24	33%	58
Efficiency	22	31%	43
Technical/Context specific	22	31%	67
Participation	19	26%	24
Relevance	19	26%	31
Effectiveness	16	22%	69
Impact	13	18%	18
Aid effectiveness	9	13%	15
Scope	9	13%	11
Time	8	11%	9
Equipment	3	4%	3
Others (not captured above)	2	3%	7
Number of reports with recommendations	69	96%	-
Number of reports without recommendations	3	4%	-
Total number of reports	72	100%	-
Total number of recommendations			1,123

Source: own analysis of evaluation reports





105 recommendations on **monitoring and evaluation (M&E)** are presented in 49 out of 72 reports (68%). When reviewing those recommendations, two major categories appear: (i) those pointing at the establishment of an M&E system and (ii) those focussing on the improvement of existing systems.

When no M&E system exists, evaluators frequently stress the importance of such a system for adaptive and evidence-based project management, learning and accountability. When an M&E system exists, in many cases, the evaluators recommended **improving the quality of indicators** (sometimes referring to the SMART criteria, i.e., specific, measurable, attainable, realistic and time-bound) or **increasing the coverage** of the M&E system by introducing indicators for each intervention objectives comprehensively. Some reports also suggest **revising the underlying results model**, for instance, by differentiating several results levels (i.e., output-outcome-impact) or shifting the focus **from output delivery to the achievement of outcomes**.

Some recommendations suggest **increasing M&E specific knowledge** and capacity among the project staff or hiring **M&E experts**. Furthermore, several recommendations point to the improvement of data quality through **increased efficiency of data** collection or the **improvement of internal processes**. A recurrent aspect in this respect is the **improvement of templates and documents** (such as log-frames or data collection plans), the revision of **data collection instruments** and the **collection of baseline data**.

In several cases, the evaluators recommend **involving partners** (such as government intuitions or research institutes) in the monitoring to **increase ownership and sustainability**. This also involves **building M&E capacity among those partner** institutions and providing them with methodological and technical advice. **Harmonising M&E-related activities with other donors'** interventions is also recommended for some of the interventions.

Few reports also recommend **commissioning external formative or ex-post evaluations** to complement internal monitoring. Finally, a frequent recommendation is that **management and planning of new** (follow-on) projects should be **based on** and follow implications and conclusions drawn from previous **M&E results**.

Some examples of recommendations in the theme M&E are provided in Box 7.



### Box 7: Examples for recommendations on M&E

*“It is recommended that the project management and M&E team uses a three levelled approach (goal, objectives, outputs,) and develop appropriate indicators for each level to ensure an ease of project management and monitoring. It is a high priority recommendation that Phase II of the project revises the results-framework for the project to ensure logic, coherence and accurate representation of project achievement as soon as possible.” (Report No. 01, p. 32)*

*“Improve supervision, monitoring and evaluation of the quality of SLLC [Second Level Land Certification], PIA [Public Information and Awareness] activities and intermediate results: • M&E process and data quality should be audited in the field. • Monitoring and evaluation of the quality of the SLLC and PIA activities and intermediate SLLC results should be strengthened through field cross checking. • Based on audit results M&E capacity on project level may need to be strengthened.” (Report No. 5, p. 46)*

*“Monitoring systems should be given more attention. They should be rigorous, based on results-based principles and should allow for effective adaptive project management. Thorough baseline- and end line-surveys should be part of these systems including a template and procedure for responding to a MTR.” (Report No. 24, p. 53)*

*“An ex-post evaluation should be considered in 1 or 2 years’ time. It should assess in particular if the capacity building support provided by FMI [Finnish Meteorological Society] has led to sustainable results, and what lessons can be learned, given the current Finnish development policy which states that “Finnish institutions active in various fields can play a significant role in improving local competences in e.g. the natural resource sector and in climate change mitigation.” (Report No. 24, p. 56)*

*“The project should strengthen the Monitoring and Evaluation System of DW II Project to better manage the process of monitoring, analysing, evaluating and reporting progress toward achieving the project’s objectives. It should be done through development of a Performance Monitoring Plan (PMP) for the Phase II and preparation of the standardised templates of data collection tools like trainers’ reports, feedback questionnaires, etc. PMP in comparison with the Log frame should contain not only the indicators, baselines, targets and the methods of data collection, but also point out the frequency of data collection for each indicator, specify responsible parties for collecting the data and how the collected data is planned to be used. All that will allow to ensure timely and efficient generation, storage and use of strategic information by the DW II Project.” (Report No. 47, p. 44)*



117 recommendations on **coherence** are presented in 48 out of 72 reports (67%). More common recommendations on this topic refer to the **intervention’s external coherence**, emphasising the importance of **strengthening complementarity with other interventions** in a similar context. Most recommendations also highlight that **aligning strategy** will strengthen existing collaborations and cooperation and provide a better overview to minimise duplication of efforts and ensure a sustainable and comprehensive strategy.

Another bundle of recommendations is related to **internal coherence** that suggests promoting synergies between programme components and better-coordinated activities. This also includes



**synergies with government and donor policies.** Finally, recommendations from reports on **multi-country interventions** point to the opportunity to promote **stronger regional collaboration** between different countries to leverage performance.

Furthermore, a majority of the recommendations presented point to **developing and expanding collaborations** or **strengthening existing linkages** with different stakeholders. However, most of these recommendations raised the importance of establishing guidelines or systematic mechanisms to create platforms that ensure effective but realistic multi-stakeholder partnerships. This means **scheduling regular meetings**, having a **clear structure, assigning responsibilities** and proper communication and documentation.

Moreover, exchanging **information and learning from each other** are important aspects of coherence that evaluators frequently recommend. Experiences and **lessons learnt** should be properly **documented and shared** internally or with relevant networks, including donors. On the one hand, it was suggested by some evaluators that better coherence helps facilitate proper channels to share information. On the other hand, some evaluators argue that more sharing of information and learning from each other helps strengthen collaboration and improve coherence.

Box 8 provides exemplarily such recommendations in the theme coherence.



#### Box 8: Examples for recommendations on coherence

*“Develop linkages with other projects, other agencies, and other donors for the development of a long-term strategy for employment generation/job creation. UNDP can support working on a long-term job-creating strategy for Afghanistan in association with all the actors/cooperating partners in development.” (Report No. 04, p. 43)*

*“To optimise synergies with other projects, it is recommended not to implement projects in isolation but as part of a larger effort to electrify rural areas. This includes a need to know what other projects are doing and in particular to partner with livelihood programs that encourage a productive use of electricity that can increase net revenues.” (Report No. 13, p. 39)*

*“Adopt an enhanced coherence model that centres around WASH as entry and livelihoods as follow-on, with the use of other infrastructure and interventions limited to solving bottlenecks that prevent intended beneficiaries, especially women, poor and excluded groups, to effectively participate in WASH and livelihoods.” (Report No. 16, p. 62)*

*“Establish a mechanism through which organisations involved in the CF/REDD [Community Forest/Reduced Emissions from Deforestation and Forest Degradation] approach can learn from another and can coordinate their efforts (e.g. exchanging experience on baseline assessments in line with international standards and avoiding that the same communities and forests are targeted by different organisations).” (Report No. 8b, p. 52)*

*“Sharing the experience and learning from each other could help to avoid problems and improve the performance. This could also apply to enhancing the promotion of the cross-cutting objectives, relevant Finnish bilateral strategies and commercial interests. It is recommended that the project managers/coordinators of ICI [Institutional Cooperation Instrument] projects in Vietnam (or other countries as well) would meet each other, say four times a year. Also, the ICI consultant and the Desk Officer(s) from MFA and, through video links, also relevant embassies could attend these meetings. Similarly, it would be beneficial if the coordinators of ICI projects could receive copies of project documents and annual and final reports of other ICI interventions.” (Report No. 30a, p. 53)*

109 recommendations **on planning** are presented in 45 out of 72 reports (63%). Evaluators either suggest (i) revising the project concept of ongoing interventions or (ii) making suggestions for improved planning of upcoming projects. Frequently, recommendations point to improving **the project design** and the underlying **Theory of Change**. This includes for instance, a revision of unrealistic or unclear objectives, the right choice of activities and instruments for the achievement of the intended objectives and the revision of inconsistent results hypotheses.

Moreover, evaluators often recommend **grounding project planning on evidence-based needs and/or context assessments**. In this regard, the adaption of project concepts to the local context is frequent. It includes the suggestion that planning should entail a **solid risk analysis**. Further, recommendations also point to the importance of **involving partners, stakeholders, and final beneficiaries** in planning processes. While this may ensure that interventions fit their needs and the conditions on the ground, the participation of partners and beneficiaries in planning may also increase ownership and responsibility. **A number of recommendations also mention improving participation and ownership by inviting stakeholders and partners to become members of an advisory board or steering committee.**



On a different note, some recommendations point at **integrating cross-cutting objectives**, such as gender equality, human rights or environmental protection, and climate change mitigation in the planning process, such as conducting a gender analysis or an environmental impact assessment or by conducting an environmental impact assessment **disaggregating indicators and monitoring data** by gender.

Few recommendations also refer to the **time frame or duration of interventions being too short** to reach the intended outcomes and impacts.

Exemplary evidence on the theme planning is presented in Box 9.

#### **Box 9: Examples for recommendations on planning**

*“Focus on integrated service delivery: The ultimate objective of policy-making and programming should go beyond improving the quality of school WASH services to a broader goal such as reducing the incidence of water-borne disease or of girl drop-outs due to poor WASH. This could ensure that programme efforts not just ensure that every school has water supplies, toilets and hand-washing stations, but that these work effectively to impact the health of school students.” (Report No. 72, p. 97)*

*“More attention should be given to project design and the design of the logical framework. A well designed logical framework (ideally based on a Theory of Change) will be able to guide the implementation phase, and will serve as the main reference framework for monitoring and for external evaluations.” (Report No. 24, p. 50)*

*“The future bilateral programmes of MFA must be subject to a human-rights, gender equality, and environmental issues analysis, which should inform the formulation of the programme document, and subsequently, the implementation of the inception phase of the programme. Interventions should also earmark funds specifically to gender equality and HRBA related activities to be implemented throughout the project or programme, as was partly done in WSPST [Water and Sanitation Program for Small Towns in Vietnam]. The analysis of environmental aspects should focus on environmental impact of the intervention, but should also include climate change from the perspective of adaptation and mitigation when relevant.” (Report No. 18, p. 76)*

*“In order to create stronger ownership at higher level (parent) organisations and among other stakeholders – public and private – they should be involved in the projects from the very beginning. Senior officials/managers or their representatives of organisations that are important for the implementation/replication/upscaling of project outputs should be invited in advisory committees/boards. The identification of the key players would require more thorough stakeholder analyses at the preparatory phase and, probably, more country-specific support.” (Report No. 30a, p. 52)*



64 recommendations on **sustainability** are presented in 39 out of 72 reports (64%). Some of the recommendations merely suggest **extending the intervention phase** to continue with vital activities, plan a follow-on project, or provide at least some follow-up support and supervision to partners. Frequently, however, evaluators point to the importance of a **solid exit strategy** to ensure the sustainability of project results. Some also recommend develop-

ing this exit strategy from an assessment of existing threats to the sustainability of the intervention's achievements.

Recommendations further entail aspects of **financial sustainability**, pointing to the problem that partners often lack funding to continue and maintain project activities once the projects withdraw. In this regard, some recommendations propose to **leverage other funding sources** through the involvement of the **private sector** or to convince **national governments** to engage in the long-term funding of activities and structures. The latter issue is connected to the importance of national governments' ownership and identification with interventions' achievements which is a prerequisite for their willingness to provide long-term funding.

Another frequent recommendation puts a focus on building **sufficient capacity of partner institutions** to enable them to continue independently with activities formerly carried out by (or with the support of) the project. The **capacity to maintain equipment, systems or infrastructure** is also regarded as crucial in this respect.

In some cases, evaluators also recommend increasing the focus on the **institutionalisation of approaches and processes**. Further, some recommendations highlight the importance of complementing interventions with political advice to support the **integration** of the intervention's achievements **into national regulations or policy frameworks**.

Some recommendations in the theme of sustainability are provided in Box 10.



#### Box 10: Examples for recommendations on sustainability

*“Most importantly, MoHD [Ministry of Health and Development] needs to do a thorough preparedness strategy for the exit of the project especially in relation to bridging the financial gap when MIDA FINNSOM phases out. MoHD should mobilize resources in a more sustainable manner by engaging the private sector, diaspora nationals worldwide for example by establishing a diaspora trust fund whereby diaspora can contribute financially to support the health system in their country. Moreover, MoHD needs to approach and engage various international and national partners that can support the project financially, materially or technically.” (Report No. 2, p. 43)*

*“Adequate political support and funding for the maintenance, annual updating, and training for the FORMIS [Forest Sector Management Information System] system is needed to keep the platform relevant and useful for the forest sector. [...] MARD [Ministry of Agriculture and Rural Development (Vietnam)] and VNFOREST [Vietnam Forest Administration] further analyse and address the existing threats to the sustainability of the FORMIS platform and its applications.” (Report No. 21a, p. 74)*

*“The AIHRC [Afghanistan Independent Human Rights Commission] should enhance the sustainability of its programme to promote human rights through even greater strategic emphasis on institutionalization and empowerment. Planning of training and awareness raising programmes should be focused on achieving sustainable outcomes.” (Report No. 25, p. 10)*

*“To ensure better sustainability of the work undertaken under SPF [Social Protection Floors] in Kyrgyzstan, ILO should ensure that the results of ABND [Assessment-Based National Dialogue] exercise will be incorporated in the new Social Protection Development Program for the Population of the Kyrgyzstan to be developed during 2017 as the current one is active only for the period of 2015-2017. Consider the provision of support for the design of specific social protection floor schemes or the reform of existing schemes based on the ABND recommendations. (Report No. 47, p. 45)*

*“The project should develop sustainability plan for DW II Project which should both outlines the steps that should be taken throughout the rest of implementation period to ensure sustainability and describe how tripartite partners intend to carry forward project results. This type of sustainability plan can then serve as a good practice for future similar projects of the DWT/CO [Decent Work Team and Country] Office in Moscow. It is also recommended that in future, issues of sustainability be part of the funding and agreement with partners at all levels.” (Report No. 47, p. 46)*

*“ECDC [Eden Centre for Disabled Children] should provide follow up support and supervision to community volunteers so as to improve skills and sustain their early intervention services to CWDs [Children with Disabilities] within the community after the ECID project closure.” (Report No. 64, p. 22) “Provide support to strengthen the capacity of the programme beneficiaries to continue application of learning through community initiatives. (Report No. 66, p. 45)*



108 recommendations on **management** are presented in 37 out of 72 reports (51%). A large share of the recommendations refers to the **organisational structure** of the interventions suggesting, for instance, to **shift competencies** between staff positions or offices, **clarify communication lines and responsibilities**, create new management positions or units, or **introduce a steering committee** for general oversight.

In some cases, the evaluators also recommend making changes in the intervention documentation, the **reporting structures** or frequencies or they suggest revising or introducing various kinds of **guidelines or work plans**. Some recommendations also refer to the change or introduction of **new technical platforms** for communication and internal work organisation.

Few reports entail recommendations referring to a **culture of learning and accountability** and encourage the management to exchange and draw on good and successful practices and lessons learnt derived from stakeholder feedback as well as from monitoring and evaluation results.

Box 11 displays a few examples of recommendations in theme management.





### Box 11: Examples for recommendations on management

*“The Steering Committee should set up a Management Team to support the Chief Technical Advisor in planning, monitoring and reporting about TA support activities and achievements and to serve as a platform for information sharing.” (Report No. 3, p. 30)*

*“Establish a clear work plan between regular activities (data collection, analysis and reporting) and developmental activities (additional linkage, backstopping), and anticipate the need for additional staff/TA support. Based on this work plan, reinforce staffing at central level.” (Report No. 6, p. 144)*

*“ICJ [International Commission of Jurists] should strive to enhance learning and promote a culture of results-oriented thinking. To improve interaction and integration, the organisation should explore technical platforms allowing for faster communication and exchange among staff members. Additionally, ICJ should organise regular events for staff members to reflect and draw conclusions and lessons learned in order to improve future project and management decision-making and reduce potential risks (Report No. 7, p. 45)*

*“Programme partners should consider a simpler management structure for a possible Second Phase. This could be based on a common fund managed by DEEF [Directorate of Economic and Financial Studies], which will promote national ownership, but which also depends on Government being able to re-build trust with donors. Alternatively, the common fund could be managed by UNU-WIDER [United Nations University - World Institute for Development Economics Research] or another fund manager.” (Report No. 10, p. 52)*

*“We recommend that Regional Programme Managers be given more influence over their programmes so they are more aligned with local needs and priorities to ensure efficiency and relevance. Within the framework of the Strategic Plan, Managers should be given discretion to adjust programmes up to a certain level (say 5-10% of the budget) to meet changing circumstances. Managers should also be involved as a matter of course in recruitment decisions for their offices, except in cases of a conflict of interest.” (Report No. 25, p. 10)*

*“Clarify and agree on coordination structures: There are several layers of bureaucracy in the project based on the number of implementing agencies involved and a second phase of the project should look into modifying the existing coordination structure for cost-effectiveness and ease of project implementation. The reality in the country is one, where greater representation is expected to be seen in Somalia, and increasingly more meetings and workshops take place in Somalia. Ultimately, decision on changes in coordination and management structures should be discussed between IOM Finland, IOM Somalia and Nairobi Office and the donor.” (Report No. 32, p. 64)*

*“A strong follow up system should be developed and implemented to address lapses in the quality control system. Key issues such as infection prevention, patient consent and privacy require attention. A comprehensive training plan, with clear priorities are already in place, however, it is recommended that is further strengthened so that clinic staff are empowered being the main business drivers.” (Report No. 42, p. 9) “While several members expressed concern about the Network’s lack of strategic plan and too close relationship with FCA [Finn Church Aid], and a lack of clarity of their roles vis-à-vis the Network, they were also unable to come up with solutions to this problem, particularly while still benefiting from the funds the Network obtains from the MFA and FCA. Thus, for the moment, the structure should stay within the FCA framework, but the strategic plan could consider alternatives for 2020.” (Report No. 49, p. 26)*



65 recommendations regarding **financial aspects** are presented in 35 out of 72 reports (49%). The theme is closely linked with management and efficiency as many of its recommendations refer to the financial management and to cost efficiency of projects. In this regard, some recommendations refer to changes or improvements in the **financial administration** of interventions. In a few cases, reports also recommend **increasing accountability** and oversight of the financial management to prevent corruption and fraud.

Evaluators frequently recommend **allocating funds to particular equipment, services or activities** assessed as crucial for the success of the intervention or to **stop funding** of items or components seen as **dispensable** or counterproductive. Some suggestions are also made on how to use remaining funds or to **shift funds** from one component or activity to another.

Recommendations to **leverage additional sources of funding** from other donors, international organisations, national governments, or the private sector appear quite frequent as well. Moreover, some recommendations are addressed to MFA or other donors pointing out underfunding of components, activities or management units. Thus, recommending the **provision of a sufficient budget** for their implementation in the future.

Few recommendations refer to **financing instruments** that are bound to specific contexts or donor set-ups of interventions. This included recommendations on **public-private-partnership** instruments, the establishment of **trust funds**, the funding of **local civil society organisations** or a set of sub-projects through tendering procedures.

Box 12 shows a few recommendations addressing financial aspects.



## Box 12: Examples for recommendations on financial aspects

*“The fund will stay in VDB [Vietnam Development Bank] to provide loans to WSCs [Water Supply Company] with the new terms. Under this scenario, VDB is required to do better business appraisals (which according to them they would do) whereas MFA must require details of the governance arrangements of the fund and insist on an annual report to MOF/MFA as the fund shareholders. Although the size of the FRWF [Finnish Revolving Water Fund] is small (and will further be reduced by inflation), it will support investment in water sector in small towns alongside other funding sources, as was the original purpose of the fund. Validity of the recommendation depends on its legality under Finnish law.” (Report No. 18, p. 75)*

*“Budgeting to include medical equipment that is not available in hospitals in Somaliland. At the moment MIDA FINNSOM does not provide a budget for supplying medical equipment. There is a strong belief that thorough skill transfers are playing a major role in changing access and quality of health services, there is a need to equip facilities with adequate tools and the equipment required. Budget should also be availed for hiring biomedical engineers that can install medical equipment, maintain and repair machines that have been out of service and train local staff.” (Report No. 02, p. 42)*

*“It is recommended that MFA considers parallel financing as an option for its multi-bi partnerships. In traditional multi-bi interventions, all project funds are commonly managed by the project partner or deposited into a trust fund managed by the donor partner, opportunities for MFA/Embassy involvement in, for example, in influencing policy dialogue exist in a limited fashion. In parallel financing MFA manages its own contribution. Therefore, it is an arrangement that requires more MFA and Embassy involvement than a traditional multi-bi intervention. However, the significant benefit of parallel financing is that it provides potential for MFA to add value to the partnership beyond mere financial inputs. This can be achieved, for example, by bringing insights from Finland and Finnish institutions or experiences and lessons learned through other MFA-funded interventions directly into the partnership. MFA could position itself as a valued and trusted partner in providing high-level TA expertise to multilateral projects. (Report No. 17a, p. 145)*

*“Projects should have sufficient budget and resources for project management and a specific budget line for M&E (including budget for baseline and endline surveys). (Report No. 24, p. 53)*

*“The Evaluation recommends that the SPM [Support to Payroll Management] Project incorporate enhanced “anti-corruption” outputs and activities (and/or indicators) in its programming and that the MOIA [Ministry of Interior Affairs] take all necessary steps –per MOIA Donor Condition B12—to ensure that the MA findings pointing to possible fraud or corruption within the MOIA or ANP [Afghan National Police]—specifically related to and involving the payroll process—are investigated by the Inspector General Office (IGO) and violators referred to the Attorney General of Afghanistan for prosecution. Any SPM outputs in this regard should be closely coordinated with MPD [MOIA and Police Development] Project outputs and activities for the IGO to bolster MOIA internal audit capacity and capability and the capacity of the anti-corruption cell within the IGO.” (Report No. 26, p. 66) “More resources are recommended to be allocated to the preparation and mobilisation of new cooperation between Finnish and partner institutions. The institutions might need more time and resources and, especially, additional country-specific support and advice should be made available to them, for example by the Embassy of Finland and/or eligible consultancy.” (Report No 30a, p. 52)*



*“In order to better operationalise Finland’s transition strategy for Vietnam in 2016–2020, more innovative use of available financing instruments, e.g., the Public Sector Investment Facility (PIF) that replaced concessional credits and the NGO window, is recommended. Finnish institutions can promote Finnish know-how and technology under ICI [Institutional Cooperation Instrument] projects but for that purpose they need to be requested/advised to do so, and ICI norms and manual should allow and facilitate this.” (Report No. 30a, p. 53)*

76 recommendations with regard to the **capacity** of implementing partners and beneficiaries are made in 35 out of 72 reports (49%). They often point to the lack of capacity and know-how among partner organisations, government institutions or groups of final beneficiaries and recommend introducing or **scale-up capacity development**.

Evaluators also frequently recommend **improving existing capacity development components**, for instance, through **needs assessments, customised training for different groups** or the development of **strategic capacity development plans**. Likewise, evaluators often point out topics or skills that should be integrated into the capacity training.

Moreover, some recommendations also underline the **importance of institutional capacity** and suggest shifting the focus from capacity development of individuals to institutional development in a broader sense.

Some recommendations related to capacity building are displayed in Box 13.

#### **Box 13: Examples for recommendations on capacity development**

*“Future projects should invest on THL [Finnish National Institute for Health and Welfare] and scale up the institutional capacity building it started in the four crucial pillars. It’s strongly believed that institutional building and skills transfer centered strategy towards MoHD [Ministry of Health and Development] would help sustain and multiply the project outcomes and impacts achieved. Though a number of capacity building and technical assistance missions have been conducted, it needs strong follow up, supervision and monitoring to ensure the gains of the phase IV can be strengthened and scaled up.” (Report No. 2, p. 42)*

*The TA support could be more efficiently used for the development and delivery of customised training for different groups and development of strategic capacity development plans. The opportunity to develop alternative, sustainable measures for capacity development using modern technology could be explored (Report No. 03, p. 30)*

*The second phase of the project should focus more on capacity strengthening at the individual and institutional levels. The Human resources deployed to support labour migration, policy implementation, enforcement of labour laws, and contracting with the private sector should be trained to take up the challenges of implementation (Report No. 04, p. 42)*

*Output two should be purely a skill development output. The activities under this output should include assessment of available skills in the market; required skills for the industry and business, liaison with industry and business, development of linkages with prospective employers; and placement services, liaison with universities and colleges for educational enhancement programmes, liaison with embassies for educational loans and scholarships; (Report No. 04, p. 42)*



*DEEF/MEF [Directorate of Economic and Financial Studies / Ministry of Economy and Finance] and the donors should assess the alternative option of developing capacity for economic policy analysis and research on inclusive growth in an autonomous government institution outside MEF during the next phase of the programme (Report No. 10, p. 51)*

*If the objective for the NGD [National Geographic Department] is to become a provider of national information services, it has to develop from a mapping and surveying administration into a service-oriented organisation. NGD needs institutional development to generate the required roles in service development and management, as well as the underlying support roles such as communication and IT (Report No. 17b, p. 146)*

*Special WASH training for teachers and principals: In contrast to the general notion that WASH does not require any special training (since everybody 'knows' how to drink water, use a toilet and to wash hands), specialized WASH training must be part of the curricula of all regular induction training, teacher training programmes and refresher trainings. The focus here must not only be on the critical importance of WASH practices (water hygiene, food hygiene and personal hygiene – and how to practice these correctly – in order to break the faecal-oral chain of infection) but also on the special and innovative techniques necessary (and possible) to make WASH trainings interesting, relevant and therefore useful and effective for school children of different ages. Building a cadre of good-quality professional WASH trainers nation-wide, and province-specific, would be a logical first step in training teachers to train children properly. (Report No. 72, p. 98)*



## 6. Findings of the Use and Usefulness Assessment

After the (methodological) quality assessment of the evaluation reports (chapter 4, component 1) and the summative analysis on the quality of the underlying interventions based on reliable evaluation reports (chapter 5, component 2), the use and usefulness assessment (component 3) focus on the **valorisation of the evaluations by key stakeholders**. We exploited survey data from three key stakeholder groups (i.e., staff at MFA headquarters and Finnish embassies and implementers) and triangulated it with 26 key informant interviews among staff at MFA who were often commissioning multiple interventions from our sample. Chapter 6.1 provides further details on the **underlying data structure and triangulation**.

Chapter 6.2, in response to **EQ10** reports on the **different uses** 📄 of **decentralised evaluations** and identifies which uses are mentioned most frequently in our sample.

In chapter 6.3, we reply to **EQ11** and show **how useful** ⚙️ decentralised evaluations are overall assessed by the three key stakeholder groups 6.2.1. Moreover, we disclose whether they are perceived as **timely**, to what extent they lead to **relevant and realistic recommendations** 💡, and to what extent these recommendations are **implemented** (6.2.2).

Chapter 6.4 addresses **EQ12** and provides insights on typical **facilitating** 🏠 and **hampering factors** for useful decentralised evaluations based on survey responses (6.4.1). And finally, it points to **major recommendations** 💬 and **lessons learnt** to improve the use and usefulness by interviewed **MFA stakeholders** (6.4.2).

### 6.1. Underlying data for the assessment

**Highlights of the section:**

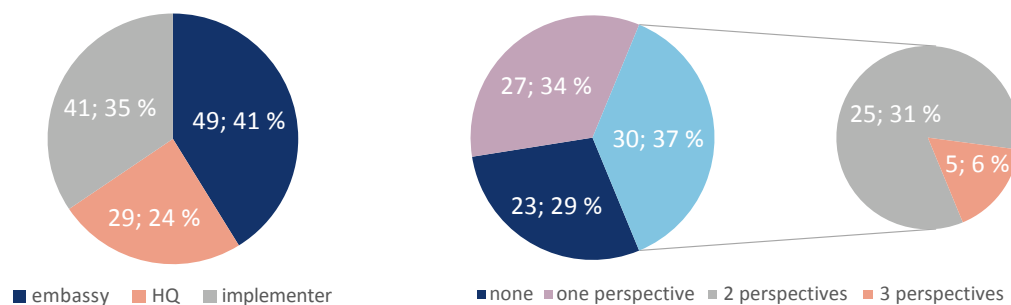
- Survey data provides insights on 71% of the cases (57 out of 80).
- However, only 38% of the maximum attainable unique perspectives (92) from MFA staff at headquarters and embassies and implementers (3 per case) limit sample representativeness.
- 26 key informant interviews with MFA staff provided a nuanced picture and were used for triangulation.
- The use and usefulness assessment adds a unique perspective to the metaevaluation.

As shown in Figure 27 (left pie chart), **in total, 119 survey responses** were obtained: 29 from MFA headquarters' staff (24%), 49 from embassies' staff (41%), and 41 from implementers (35%).



We averaged multiple responses received for one evaluation case to reduce potential biases if they fall under the same of the three possible key stakeholder perspectives (i.e., MFA headquarters, embassies, or implementers). This yielded **92 unique survey responses**. Out of all possible perspectives for the 80 cases in our sample (N = 240), these 92 reflect a total **coverage of 38%**. However, the 92 different perspectives from the survey cover a **total of 57 cases** and thus possess a **case coverage of 71%**. Figure 25 (right pie chart) displays that for about a third of the cases (27, 34%), we received one perspective, for about another third (30, 37%) two or three perspectives; for the remainder (23, 29%) no perspective could be attained and only for 5 out of 80 projects (6%) received a complete set of survey responses (i.e., perspectives of MFA headquarters, embassies, and implementers) on the same evaluation.

**Figure 27: Survey responses (n=119) and share of perspectives per case (n=80)**



Source: own statistics based on survey

The **low overall coverage rate ultimately limited triangulation** of the survey data at case level. We tested whether the survey responses differed between stakeholders for the sub-sample of 30 cases in which we had more than one perspective per case. There were no statistically significant differences, so we decided to use the case-level aggregated data (57 cases) for our analyses (see details in the statistical annex 15).

**Triangulation with the interview data** of 26 key informant interviews allowed deeper insights into the use and usefulness than the survey data alone. The interviewees provided their perspective on a specific evaluation project from the sample but also answered more general questions about their overall view on the usefulness of evaluations. The interviewees appeared to be open and provided quite frankly a more nuanced picture of their view of the usefulness of evaluations than displayed by the quantitative data.

Further, quantitative data obtained from the quality assessment (component 1) was put in context to the survey results to investigate how data relates to each other. For 53 out of the 80 cases under this metaevaluation it was possible to calculate the **correlation between the overall quality** of the evaluation report as **assessed by the key stakeholders** in the survey and the **overall quality** of the report as assessed in the quality assessment **by the metaevaluation team** (component 1). The two different quality assessments were very weakly related, as they shared less than 1% of the common variance, and the correlation was **insignificant** (see statistical annex 15 for details).

The counter-intuitive weak relationship is **plausible** given the **different nature** of the **assessment** and the **assessors**. While the quality assessment under the metaevaluation stems from a standardised and weighted assessment with a strong focus on aspects of methodological quality, completeness and attainment to good evaluation practice, the quality assessment in the survey



reflects the personal assessment of key stakeholders' response to a single question on overall quality. Differing overall quality assessments are further plausible, as the metaevaluation team undertook an independent assessment from a distance whereas key stakeholders were locked-in to the evaluation process, the underlying interventions, and the use of evaluation results. Thus, their assessments are expected to ground on much more elements than the methodologically focused quality assessment of the pure evaluation reports by the metaevaluation team.

Overall, the analyses of the relationship between the different assessments indicate that component 3 adds a unique perspective to the metaevaluation of decentralised evaluations. Further assessments of different quality aspects by the survey respondents are provided in annex 15).

## 6.2. Use of decentralised evaluations (EQ10)

### Highlights of the section:

- Learning in teams was the most frequent use, followed by decision making, planning, and management of ongoing and follow-up interventions.
- More strategic usages for Finnish development cooperation were less frequent, but this is not negligible as only applicable for MFA headquarters and embassy staff and not for implementers.



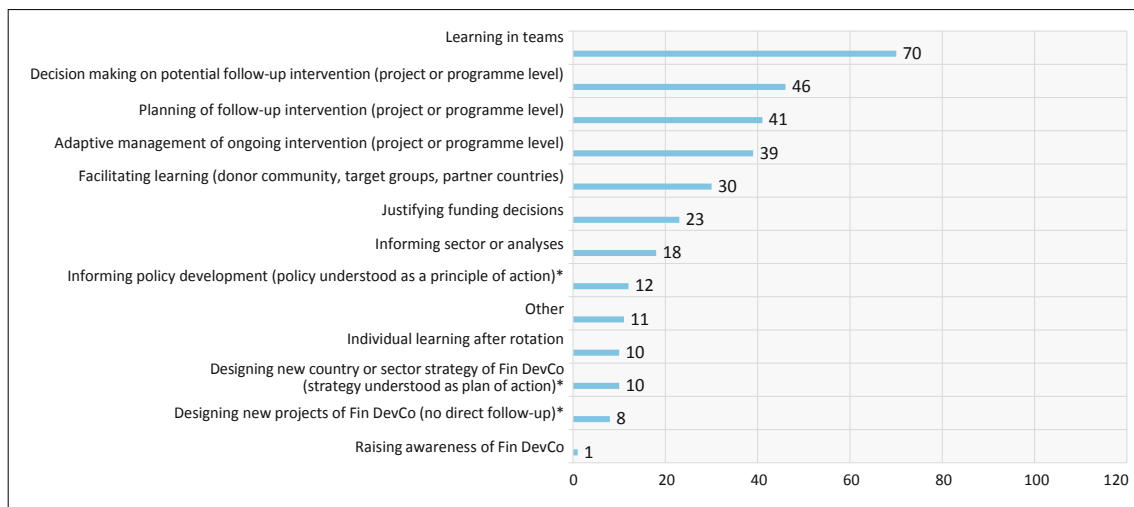
The usage of the decentralised evaluation was assessed using a checkbox list of **19 potential usages** (14 for implementers). Each participant was allowed multiple usages for each evaluation report and to use a free text field to add other usages not included in the list. Further, a residual option allowed participants to indicate if they did not use the evaluation report at all.

In the total sample, including all 119 survey responses, **319 usages were selected overall**. To ease interpretation of the following findings, similar usages were aggregated to one category, e.g., learning in teams which was originally assessed as “learning in the country team”, “learning in the sector team”, and “learning in the project team”. Figure 28 displays how often different usages were mentioned. Usages that did not appear in the implementer questionnaire are marked with an asterisk. The **most frequently** mentioned usages were for **learning in teams** (project, country, sector; 70), followed by intervention-specific usages, such as **decision making on potential follow-up** (46), **planning of follow-up of the intervention** (41), and **adaptive management of an ongoing intervention** (39). More strategic usages in the context of the Finnish development corporation, such as designing new country strategies, sector strategies, or projects, were mentioned less frequently (10 and 8, respectively). However, it should not be interpreted that these usages were negligible. As they are only applicable for MFA staff, they were not presented as answer options to the implementers.





**Figure 28: Usage of evaluations (multiple answers, 319 usages out of 199 survey responses)**



Source: own statistics based on survey

## 6.3. Usefulness of evaluations (EQ11)

### 6.3.1. Overall usefulness of the evaluations

#### Highlights of the section:

- Almost half of the surveyed evaluations are assessed as (very) useful, another half as satisfactory.
- Only in single cases usefulness is assessed as in need for improvement or inadequate.
- 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.



**The overall usefulness of decentralised evaluations** was rated by the key stakeholders on a 4-point scale ranging from 1 = “inadequate” to 4 = “(very) good”. As Figure 29 shows, the 57 evaluation reports assessed in the survey were perceived as **highly useful**. The usefulness of nearly half of the evaluations (48%, 27) was assessed as satisfactory, while it was very good for nearly the other half (45%, 25) of the evaluations. Only 5% of the survey responses (3) indicated the need for improvement, one case was considered “inadequate”, and for one case, a respondent did not provide an assessment.

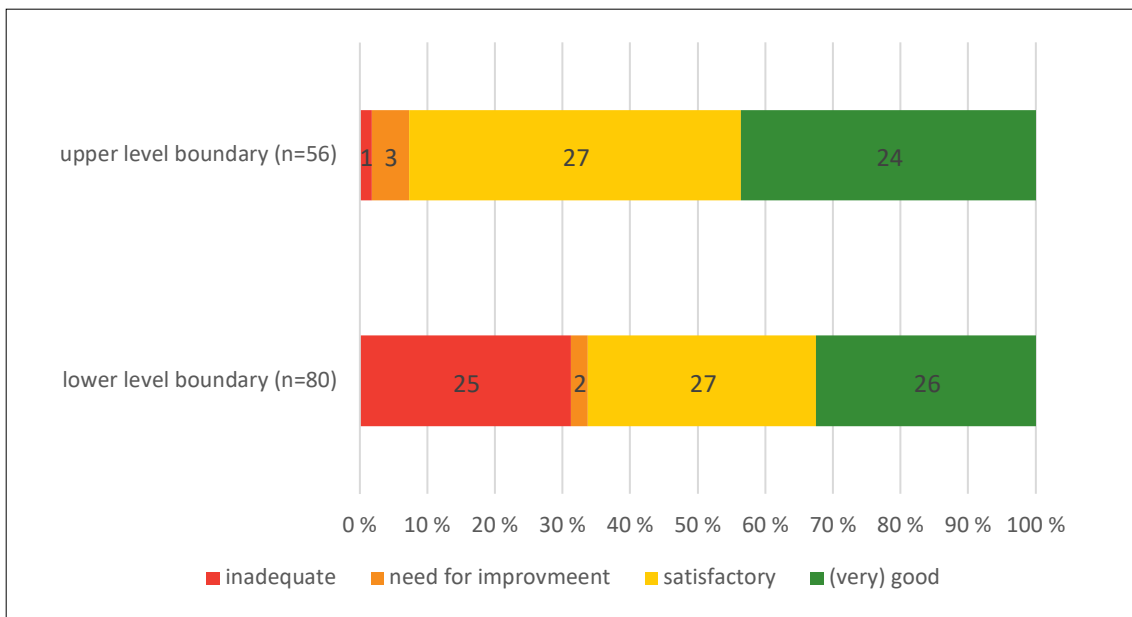
We tested whether this overall very positive picture changes when excluding implementer responses. This was because implementers may be less critical of their interventions, which may have



also led to more optimistic assessments on evaluations' usefulness. Five cases less were assessed as (very) useful when excluding implementer responses, but there was no major shift in the overall perceived usefulness. This suggests that there is **no major upward bias from implementers' responses**.

However, the above-mentioned figures can be seen as an **upper-level boundary of the perceived actual usefulness** by key stakeholders in our sample. As perceived actual usefulness covers only the evaluations for which survey responses were received, probably a positive selection bias comes here into effect. Or, put differently, key stakeholders involved in interventions with an evaluation with very limited usefulness may have been more reluctant to fill out the survey, leading to missing data for less useful cases. As also displayed in Figure 29, when assuming the usefulness of all 24 cases for which we did not receive any survey responses were inadequate, we receive a **lower-boundary estimate of the usefulness** of decentralised evaluations for the complete sample of 80. This estimate is very conservative as missing responses are unlikely to be exclusively limited to useless evaluation reports. Nevertheless, it is **still quite encouraging** that more than half of the 80 evaluations achieved at least satisfactory usefulness.

**Figure 29: Overall usefulness of evaluations (case level, ranges)**



Source: own statistics based on survey



### 6.3.2. Usefulness of timing and recommendations

#### Highlights of the section:

- The timing of the evaluations, as well as the timeliness of formal and informal results delivery, are overall positively assessed.
- The same holds true for the relevance of the recommendations, the extent of recommendations being realistic, the learnings from the evaluation and the implementation of recommendations.
- Most surveyed evaluations achieve at least satisfactory or even (very) good assessments.
- Only few cases are assessed with in need for improvement and single cases as inadequate.



Figure 30 shows the results of seven survey questions. Overall, the usefulness was **(very) positive** as for **more than 90 %** of interventions, the **timing and recommendations** were being assessed as **at least satisfactory**.

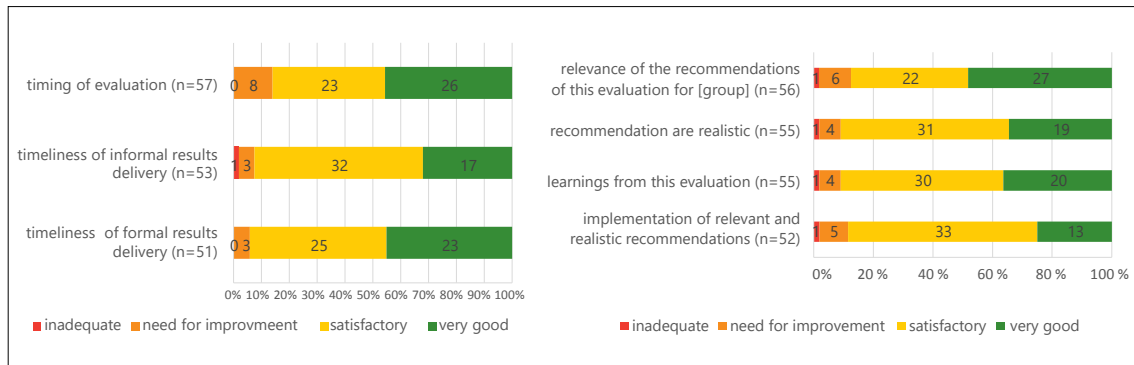
The left part of Figure 30 presents the detailed assessments concerning the **timing of the evaluation** (57 cases), the **timeliness of informal results delivery** (53 cases), and the **timeliness of formal results delivery** (51 cases). For all three aspects, **positive assessments dominate**. A large share is assessed as (very) good (26, 46%, 17, 32%, and 23, 45%, respectively). A larger share was perceived as satisfactory (23, 40%, 32, 60%, and 25, 49%, respectively). Need for improvement was only reported for a few cases (8, 14%, 3, 6% and 3, 6%, respectively), and only in one case, the timeliness of informal results delivery was assessed as inadequate.

The right part of Figure 30 displays the findings related to the **usefulness of the recommendations**. The overall pattern is very similar to timeliness and timing, with **mostly positive assessments**. The **recommendations** of 27 out of 56 cases were perceived as (very) **relevant** for the key stakeholder groups (48%), 22 cases were perceived as being of satisfactory relevance (40%), while the relevance of 6 cases was perceived as in need for improvement (11%), and only one case assessed as inadequate. In 19 out of 55 cases, the **recommendations** were further perceived as (very) **realistic** (35%), in 32 cases as satisfactory realistic (56%), in four cases as in need for improvement (7%), and in one case as inadequate (2%).

The right part of Figure 30 further reveals almost the same pattern for the **learnings from the specific evaluation** as the earlier presented aspect. Finally, the **implementation of relevant and realistic recommendations** was slightly more critical. Out of 52 cases, 13 were assessed (very) good (25%), 33 cases as satisfactory (63%), 5 cases as in need for improvement (10%), and one case as inadequate.



**Figure 30: Timing and timeliness (case level) and Recommendations (case level)**



Source: own statistics based on survey

## 6.4. Drivers of usefulness (EQ12)

### 6.4.1. Facilitating and hampering factors for usefulness

#### Highlights of the section:

- Typical factors facilitating the usefulness of evaluations are selecting the adequate nature of the evaluation (mid-term vs final), setting-up appropriate ToRs, follow-up on the management responses, producing management responses and technical and methodological competence of the evaluation team.
- 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 vs 116).



To identify typical factors facilitating or hampering the usefulness of decentralised evaluations, survey respondents received two checkbox lists on potentially facilitating (16) and hampering factors (15)<sup>3</sup>. It was possible to select multiple factors and to add factors not included in the list in a free text field. We also allowed participants to select “I don’t know” as a residual option.

Overall, **252 facilitating factors** were selected in a total of 85 survey responses. Figure 31 displays the most often selected factors. The most frequently mentioned facilitating factors related to the **nature of the evaluation**: the evaluation being a final evaluation (39 responses) was mentioned about as often as a facilitating factor as being a mid-term evaluation (34 responses). This suggests that selecting the adequate nature of the evaluation is important and varies among

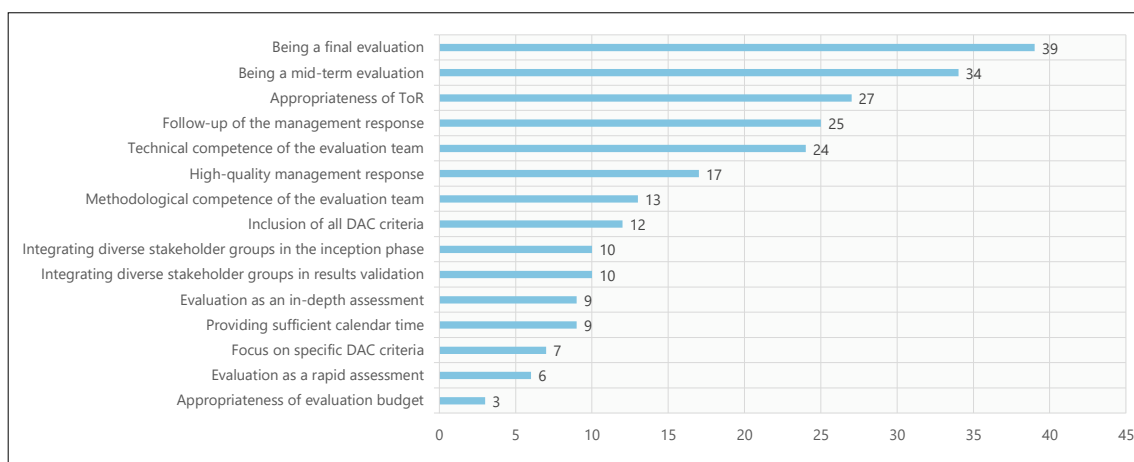
<sup>3</sup> Implementers received a list with only 14 hampering factors as we omitted “Evaluation capacity gaps at MFA level” from the implementer questionnaire.



interventions. Most other frequently selected factors are the **appropriateness of the ToRs** (27 responses), **follow-up of the management responses** (25 responses), and **high quality of management response** (17 responses). Taking this together, the **main facilitating factors** are **under the MFA's sphere of influence**. Furthermore, the **technical competence** of the **evaluation team** (24 responses) and its **methodological competence** (13 responses) were frequently mentioned.

Factors that were rarely mentioned included integrating diverse stakeholder groups, including all OECD DAC criteria, focusing on specific OECD DAC criteria, the evaluation being an in-depth assessment, the evaluation being a rapid assessment, providing sufficient calendar time, and the appropriateness of the evaluation budgets.

**Figure 31: Main facilitating factors**



Source: own statistics based on survey

With respect to factors that typically hamper the usefulness of decentralised evaluations, **116 hampering factors** were mentioned in 64 survey responses. No typical hampering factors emanate from the data. Factors mentioned more than ten times but not sufficiently frequent to be considered typical were insufficient time resources, lack of M&E systems, and travel restrictions due to the Covid-19 pandemic. Hampering factors that were rarely mentioned included a lack of coordination between MFA headquarters, embassies, and implementers, a lack of evaluation capacities, a lack of management response and follow-up, insufficient financial resources, a lack of ownership, a lack of high-quality ToRs, and the end of MFA engagement in the country or sector.

A comparison with the overall amount of facilitating factors suggests that the identification of the latter is much easier for respondents. Overall, 252 facilitating factors were identified, among them some typical factors, against only 116 hampering factors without any typical factor.



## 6.4.2. Interviewees' lessons & recommendations

### Highlights of the section:

- Interviewees' recommendations and lessons to improve evaluations' usefulness comprise mainly organisational aspects, timing and timeliness of the evaluations, and aspects of report quality.
- Support to enhance the clarity on the scope and focus of the evaluation, efficiency of learning and results dissemination at an institutional level is desired by MFA staff and expected to enhance usefulness.



Qualitative content analysis of the interview notes from 26 key informant interviews with MFA staff at headquarters and embassy level were firstly inductively coded into categories and secondly summarised. **Interviewees' lessons learnt and recommendations** to enhance the use and usefulness of decentralised evaluations were related to the following seven categories:

- **organisational aspects** including the improved planning process, ensuring commissioning of external and independent evaluators, and fostering the application of specific evaluation tools (16),
- **timing and timeliness of an evaluation**, including effective time management and adequate timing with respect to setting-up an evaluation (as sometimes too late or too early) (14),
- **report quality**, including concise and comprehensive reporting and following good evaluation practice (14),
- **quality of recommendations**, with a focus on being specific and practicable (9),
- **communication and exchange**, pointing to commissioners and evaluators staying in touch throughout the evaluation process (8),
- **evaluation budget**, particularly linking higher budgets to higher-quality evaluations (3), and
- **follow-up exchange**, underlining the importance of discussing evaluation results (3).

On broader lines, interviewees identified improving the **clarity on the scope and focus of the evaluation** as an important driver for useful evaluations. Exemplarily, one interviewee raised that answers to core questions like “What do we really want to find out with this evaluation?” and “What problem do we want to solve?” help set up ToRs in a way that they guide evaluators adequately. Confirming this, several interviewees identified support required to develop high-quality evaluation questions, which they consider a precondition for useful evaluations.



Another broader strain underlines the importance of **communication and exchange**. Some interviews said that discussing **ideas to enhance the efficiency of learning** would be helpful and advice on how to digest evaluation results best for those who are not sufficiently closely involved to possess time resources for reading full long reports. Finally, the efficiency of learning was assessed as being hampered by rotation. This underlines the last broadly acknowledged aspect, the limited exploitation of evaluation results and recommendations due to knowledge gains at the individual level, which were mostly **not yet leveraged to MFA as an institution**.



## 7. Conclusions

### **Opening remark on MFA's decentralised evaluation portfolio (EQ1) and key limitation of the findings of this metaevaluation**

Because of unknown volumes and characteristics of the whole population of bi-, multi- and multi-bilateral interventions, the representativeness of the sample for this portion of Finnish development cooperation cannot be assessed. Thus, conclusions on the adequacy of MFA's decentralised evaluation portfolio are not possible.

Geographical scope, sectorial affiliation, the nature of interventions, commissioners, and implementers, as well as intervention budgets, vary widely (see chapter 3.3). The same holds true for the nature of the evaluations, their commissioner and evaluation budgets. Thus, simplifications like equal weighting for small and large interventions and evaluations were applied (see chapter 2.2).

To put the following conclusions correctly into perspective, these limitations must be kept in mind.

### **7.1. Component 1: Methodological quality assessment of evaluation reports and ToRs (EQ2, EQ3 & EQ4)**

**C1: MFA does often not enforce compliance with its guidelines. Therefore, most evaluation reports are grounded on weak methodologies; nevertheless, findings are somewhat reliable.**



The quality assessment (see findings chapter 4.2 and annex 11) revealed that the vast majority of reports are constrained with respect to methodological rigour, provision of high-quality, evidence-based findings, appropriate capturing of OECD DAC criteria and adequate consideration of CCOs and HRBA. Despite these limitations, many evaluation results are grounded on adequate context analyses and comprehensive data collection, while developed recommendations and conclusions seem to be plausible. In a nutshell, decentralised evaluation reports are somewhat reliable but substantially lack transparency in terms of informing the reader about how evaluators came to their conclusions.

It is alarming that commissioners inside and outside MFA accept reports with considerable quality flaws. For MFA-commissioned evaluations, we conclude that (i) evaluators are regularly not complying with existing guidelines (otherwise, e.g., CCOs and HRBA would have been acknowledged) and that (ii) commissioners systematically fail to enforce compliance (otherwise, e.g., evaluators would have delivered reports with sound methodologies).





**C2: Every third ToR reveals needs for improvement; thus, evaluation capacity inside MFA is still constrained. This marks an important action area for MFA as sound ToRs leverage the quality of evaluation findings.**



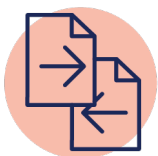
The quality assessment (see findings chapter 4.1.1) disclosed that two-thirds of the ToRs are of satisfactory quality, only one is rated as (very) good, and the remainder is in need of improvement. Thus, inside MFA, a solid fraction of commissioners can formulate adequate ToRs, while for a third of the ToRs, commissioners fail to do so. Hence, an evaluation capacity gap among MFA commissioners is plausible, and the mere existence of internal guidelines was insufficient to mitigate this.

Although some revisions of the guidelines were too late for a fraction of our sample, the result on the remainder still suggests that revising the guidelines did not suffice to uplift the overall quality of the ToRs at a considerable scale since the previous metaevaluation. This yields us to two strains of conclusions: (i) widespread capacity constraints at different levels persist, and (ii) systematic enforcement of improved guidelines inside MFA did not yet occur.



Consequently, ToRs mainly provide (rather) little methodological guidance to evaluators and often lack requirements to analyse CCOs and HRBA. That is cascading to the quality of evaluation reports which are frequently weak in this regard. Regression analysis confirmed that the quality of evaluation findings is significantly driven by the quality of underlying ToRs (see chapter 4.1.2). Thus, MFA's evaluation capacity – as displayed in its ToRs – is linked to and can hamper or foster the quality of its commissioned evaluation reports. We conclude that evaluation capacity development of its staff marks an important action arena for MFA to take influence on the quality of decentralised evaluation reports.

**C3: Acknowledging increased evaluation standards, there is a risk that MFA may fall behind its previously attained level of quality.**



A comparison of the recent quality assessment with the previous metaevaluation (see findings chapter 4.3) reveals that the overall quality of evaluation reports stays constant at a moderate level. Roughly two-thirds of reports possess satisfactory quality, and about one-third features considerable weaknesses. However, having maintained this moderate level is not sufficient for MFA to remain on track.



As soon as increased evaluation quality standards were featured in the recent metaevaluation, overall quality scores decreased. Only a bit less than half of the reports is still assessed as satisfactory. Or put differently, the other half disclosed the need for improvement when anticipating the new coherence criterion, CCOs and HRBA in overall quality scores. It is plausible that evaluators have not yet fully adapted to state-of-the-art evaluation practice. In a dynamic environment where standards of good evaluation practice are steadily rising, we conclude that MFA is endangered to lose track.



## 7.2. Component 2: Summative analysis of the quality of interventions (EQ5, EQ6, EQ7, EQ8 & EQ9)

**C4: Regarding bi-, multi-, and multi-bilateral interventions, MFA is further on track, yet every fifth intervention is of limited quality. Particularly on the OECD DAC criteria sustainability and coherence, considerable room for improvement persists for all developmental partners.**



The content assessment disclosed that the quality of the bi- and multilateral interventions under consideration continues to be assessed quite positively (see findings chapter 5.1, 5.3 and annex 12). Overall, about 20% of the interventions possess high, 60% moderate and around 20% limited quality. That is slightly better than in the previous metaevaluation, where about a third of the interventions was of limited quality. Yet, it should be kept in mind that we are looking at a set of different interventions which vary on sample characteristics like sectorial spread and intervention budgets; thus, direct comparisons are invalid. Nevertheless, we conservatively conclude (i) MFA is further on track with four out of five interventions possessing at least moderate quality while (ii) several interventions of limited quality at the same time call for stronger efforts of Finnish development cooperation. This holds equally true for multi- and multi-bilateral interventions with major contributions by other donors and does not vary between different PPAs.

Moreover, the content assessment further unveiled that the relative strengths and weaknesses of the interventions do not vary among different commissioners and are thus not MFA-specific. While relevance turned out as a particular strength with nearly all interventions assessed being of moderate or high quality, sustainability, coherence and to a lesser extent, impact are relatively weak, with about half of the interventions assessed being of limited or low quality. Consequently, and in line with the previous metaevaluation, limited sustainability is the most serious threat to Finland's development cooperation quality.

**C5: Despite gender being more often mainstreamed than other CCOs and HRBA, Finland's Development Policy is not yet entirely reflected in its bi-, multi-, and multi-bilateral interventions.**



The metaevaluation disclosed that cross-cutting objectives (CCOs) and the human rights-based approach (HRBA) are neither consequently assessed by the evaluators nor – if assessed – generally mainstreamed in Finnish development cooperation (see findings chapter 5.2). Despite the limited data, gained evidence clearly shows that gender equality is better mainstreamed than other CCOs and HRBA. At least partially, this reflects its prominent role and long-standing tradition in Finnish Development Cooperation. We still conclude that systematic consideration of gender equality as a CCO remains below expectations given its strong recognition in Finnish development policy over the past decades.

Furthermore, it is alarming that data show widespread neglect of non-discrimination, climate sustainability and HRBA by most evaluators. That provides at least hints that interventions might frequently lack systematic mainstreaming, as otherwise, evaluators were likely to be informed on CCOs and HRBA by



interventions' key stakeholders in the course of data collection. According to the evaluation reports at hand, we thus conclude that Finnish Development Policy is not yet fully reflected in Finnish bi- and multilateral interventions.

**C6: The vast body of recommendations and lessons learnt points to continuing areas for improvement for MFA. Ill-functioning M&E systems and weak planning, sustainability, coherence, management, capacity development, and financial aspects are not yet at the desired level.**



The qualitative content analysis of evaluators' recommendations and lessons learnt shows that most interventions still lack a functioning M&E system; moreover, planning, sustainability, coherence management financial aspects and capacity development remain areas for improvement (see findings chapter 5.4). In conclusion and in line with the previous metaevaluation, MFA is still confronted with major issues impacting interventions' quality and evaluability.

Again, as we cannot observe differences between MFA- and non-MFA commissioned evaluations, these challenges are not specific to MFA but rather of general nature in development cooperation.

### 7.3. Component 3: Use and usefulness assessment of the evaluations (EQ10, EQ11 & EQ12)

**C7: Decentralised evaluation reports often support learning, decision making, planning and adaptive management inside MFA, but room to enhance usage beyond individual levels and to avoid spending on evaluations of little usefulness remains.**



According to the use and usefulness assessment, decentralised evaluation reports are generally used by MFA staff at headquarters and in embassies (see findings chapter 6.2). A minimum of 70% of the evaluation reports were used by MFA staff and/or by implementers, mainly for learning, decision making and planning on follow-on interventions, and adaptive management. Yet, data suggests that not all reports are used. Thus, we conclude there is still room for

improvement to enhance their usage inside and outside MFA and eradicate spending on useless evaluations in the worst-case.

**C8: Usefulness of evaluation reports often remains at an individual level. A lack of a functioning institutional knowledge management system hampers exploitation for organisational learning inside MFA.**



The analysis further shows that the overall usefulness of decentralised evaluation reports is assessed positively by most survey respondents and some interviewees (see findings chapter 6.3.1). This also holds true for their timing, timeliness, recommendations, and learning (see findings chapter 6.3.2). Evaluation processes and quality of executive summaries are assessed as slightly less useful (see annex 14). MFA staff frequently pointed to the usefulness of



decentralised evaluations for their own work but at the same time, mentioned challenges to spread gained knowledge among colleagues (see findings chapter 6.4.2). Therefore, we conclude that insights from decentralised evaluations are not systematically exploited for organizational learning. In contrast, in the absence of a functional institutional knowledge management system inside MFA, they remain at an individual level or, even worse, get lost due to rotation.

**C9: MFA can enhance the usefulness of its decentralised evaluations by actively making use of its sphere of influence.**

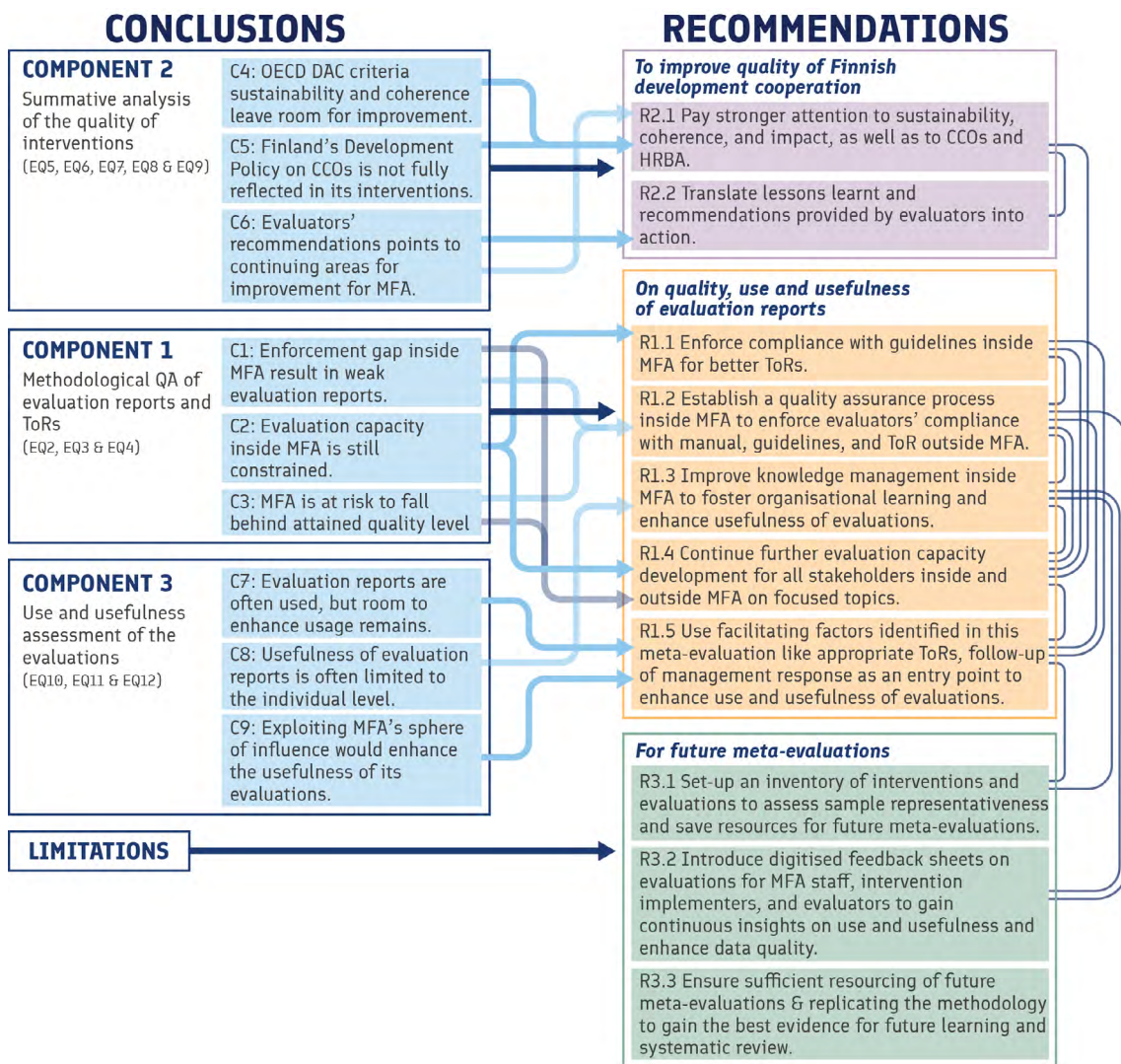


The use and usefulness assessment further revealed that factors facilitating the usefulness of an evaluation often fall under MFA's sphere of influence (see findings chapter 6.4.1). They predominantly comprise choosing the right evaluation type, ensuring the appropriateness of the underlying ToR and following up on the corresponding management response. Hampering factors were more scattered and, in much fewer cases, identified. Amongst others, the lack of a functioning M&E system at the intervention level and insufficient time were most prominently named as a threat to useful evaluations. Given the much easier identification of facilitating factors than detection of hampering factors, we conclude that analysing good practices is a good departure point for MFA to further strive towards more useful evaluations reports.

# 8. Recommendations

Figure 32 presents the section at a glance.

Figure 32 Summary of the section



Source: Evaluation team



The recommendations from the metaevaluation are structured by the following evaluation questions: What are recommendations to improve

- the quality of MFA's decentralised evaluations reports (EQ13),
- the use and the usefulness of decentralised evaluations (EQ15),
- the quality of Finnish development cooperation (EQ14), and
- the methodology of metaevaluations in the future (EQ16)?

They are derived from and linked to the conclusions drawn in chapter 7, which in turn are grounded on the metaevaluation findings on the quality assessment of evaluation report and ToR (component 1), the summative analysis of interventions' quality (component 2), and the use and usefulness assessment of evaluations (component 3) as presented in chapters 4, 5, and 6.

## 8.1. Recommendations on quality, use and usefulness of evaluation reports (EQ13 & 15)

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

The quality assessment shows that one-third of the ToRs feature considerable weaknesses. This underlines that systematic adherence to MFAs internal guidelines is not given. This is alarming as guidelines that have been improved upon recommendations of the previous metaevaluation, thus, cannot exploit their full potential. As high-quality ToRs yield a higher quality of evaluation findings and thus better reports, we highly recommend enforcing compliance with guidelines inside MFA. Adherence to guidelines can be strengthened in the course of an institutionalised evaluation quality assurance process. More details in this regard are provided under recommendation R.1.2. In the absence of such a process, and as an immediate reaction, we recommend that high-level MFA management launches a call to underline the importance of adherence to internal guidelines. This could be, for example, delivered in a speech during a physical meeting or by distributing an official action letter. Offering brown-bag meetings to discuss the guidelines and good practices of ToR-drafting would be promising options for follow-up. Opportunities to do so arise, for example, in the course of the metaevaluation results dissemination or when an – in the light of the new coherence criterion – updated evaluation development norm is communicated. Another interesting possibility would be in the context of launching the upcoming Framework Agreement for Decentralized Evaluations and Reviews (FADER).

*This recommendation is linked to conclusions C2 and findings of chapter 4.1.*

*Main implementation responsibility: high-level MFA staff, EVA-11, regional units*

*Urgency: high*

*Priority: high*



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

On the one hand, the approval of weak evaluation reports, and on the other hand, the existence of improved manuals, web-based learnings and two-thirds of ToRs being of satisfactory quality, points to an additional enforcement problem. MFA does not yet systematically enforce evaluators to adhere to its guidance. We recommend establishing a quality assurance process inside MFA to overcome this challenge. A comprehensive evaluation quality assurance process can be understood as institutionalised workflow commencing with the selection of the right point in time for the evaluation of a pre-selected intervention and lasting to the provision of feedback on the particular evaluation process, or even further to the follow-up on the implementation of the management response.

It comprises all steps required to foster high-quality evaluation reports, brings them into a logical order, displays responsibilities of different actors, calendar time and working time required, points to supporting materials to comply with each step and indicates focal points that can potentially advise coping with upcoming challenges. An evaluation quality assurance process ensures that commissioners are aware of their responsibilities and tasks and know which steps must be undertaken to provide structured and constructive feedback on all deliverables.

As an important feature of an evaluation quality assurance process, we recommend a more detailed annotation for inception and evaluations reports to clearly instruct evaluators. Such annotations would guide evaluators by specifying sub-chapters and providing elaborations on aspects and their level of detail to be covered, like, for example, requests to present and discuss intervention logics, to elaborate on how collected data is analysed, or to provide gender-disaggregated assessments. We suggest involving experienced and highly skilled evaluators as reviewers to assure compliance of inception and evaluation reports with the annotations and further guidance. The latter could step in, for example, if commissioners are faced with resource constraints, and at the same time, they may be a valuable source for exchange on critical evaluations or regarding adaptation if evaluation standards further emerge. Another promising feature inside an evaluation quality assurance process may be establishing an evaluation quality assurance help desk to support commissioners. It can also be considered to mandate such a desk to follow up on the adherence to the evaluation quality assurance process. Further, we suggest linking the quality assurance process to the upcoming FADER.

*This recommendation is mainly linked to conclusions C1, C3 and findings of chapters 4.2, 4.3, annex 11.*

*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*



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

The use and usefulness assessment revealed that the usefulness of decentralised evaluations at the individual level is quite high while exploiting evaluations systematically for organisational learning does not occur. To further uplift evaluations' usefulness, we recommend improving knowledge management inside MFA. A functional institutional knowledge management system ensures that relevant information on interventions and evaluations (like the key characteristics presented in chapter 3.3, intervention implementing status, contact details of responsible MFA, embassy, implementer staff, and evaluators) is regularly fed into a system and that such information is tagged to be accessible to interested users. Anticipating that knowledge management is a challenge for many organisations, we recommend seeking professional support to inquire on the technical functionality and user-friendliness of the existing system, including in the context of the development policy system reform (KeTTU). Furthermore, an expert can 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 the evaluation quality assurance process as recommended by **R1.2**.

*This recommendation is mainly linked to conclusions **C8** and findings of chapter 6.2.*

*Main implementation responsibility: EVA-11 for setting up the system, regional units for archiving evaluation insights*

*Urgency: high, immediately, should be systematically integrated in 2022*

*Priority: high*

**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.

The quality assessment unveiled that one-third of the evaluation reports and ToRs feature considerable quality flaws. Moreover, the quality level decreases as soon as increased evaluation standards were fed into the analysis. We recommend continuing further evaluation capacity development at different levels in response to this. As methodological rigour, provision of high-quality, evidence-based findings, appropriate capturing of OECD DAC criteria, and adequate consideration of CCOs and HRBA has been clearly identified as challenging, we recommend tailoring any measurements undertaken to focus on those aspects with the greatest room for improvement. Thereby we suggest prioritising training regarding the provision of high-quality, evidence-based findings and appropriate capturing of CCOs and HRBA. While the earlier is understood as the backbone of reliable evaluations, the latter reveals the largest evidence gaps to assess the quality of Finnish funded bi-, multi- and multi-bilateral interventions against Finnish Development Policy.

Capacity development formats are recommended for both stakeholders inside and outside MFA (i.e., commissioners of evaluations and evaluators). This could be extended to intervention implementers who lay the foundation for sound evaluations when setting up and running monitoring systems. In relation to recommendations **R1.1** on enforcement of guidelines inside MFA and to recommendation **R1.2** on establishing a quality process to enforce evaluators' compliance with





guidelines, compulsory web-based training on prioritised issues is a promising option to be considered. Successful course participation would then be awarded a certificate that can be made a precondition to award assignments to evaluators or comply with on-the-job training requirements inside MFA. To anticipate dynamic framework conditions with raising evaluation standards, we further recommend considering refresher courses. Finally, setting-up such formats is mostly promising if accompanied by an evaluation capacity specialist and consultation with key stakeholders. It might be promising to also inquire how efforts can be linked or streamlined with the upcoming FADER.

*This recommendation is mainly linked to conclusions C3, C2 and findings of chapters 4.1.1, 4.2 and 4.3.*

*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*

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

Putting together evidence gained from survey respondents and interviewees, the use and usefulness assessment reveals that some room for improvement exists. Therefore, we recommend improving the development of appropriate ToRs, the follow-up on management responses, and the selection of the right evaluation type. They have all been identified as facilitating factors for the usefulness of evaluations and thus mark an important entry to further reduce the number of little useful or even useless evaluations.

To enhance the appropriateness of the ToRs, we refer to **R1.1** on enforcement of compliance with internal guidelines and **R1.4** on tailored evaluation capacity development. Strengthening the follow-up on management responses can be linked to the proposed quality assurance process suggested in **R1.2** when adding two additional steps to the process: first, making the provision of a management response also compulsory (or at least highly recommended) for decentralised evaluations and second, institutionalise follow-up on their implementation. Thereby, we suggest setting clear and mandatory timelines for both steps and ensuring their enforcement. Finally, selecting the right evaluation type may be supported by an easy-to-use and maintain inventory of bi-, multi- and multi-bilateral interventions that display interventions' implementation status as a derivate from the envisaged start and end dates and approved prolongations, if any. For further recommendations and benefits on this, see **R.3.1** below.

*This recommendation is mainly linked to conclusions C9, C7 and findings of chapters 6.2, 6.4 and 6.3.*

*Main implementation responsibility: MFA EVA-11*

*Urgency: medium*

*Priority: medium*



## 8.2. Recommendations to improve the quality of Finnish DevCo (EQ14)

**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.

Despite not being an MFA-specific challenge, the content analysis revealed that interventions' performance on sustainability, coherence and impact is generally weaker than on the other OECD-DAC criteria. Moreover, CCOs, particularly non-discrimination and climate sustainability as well as HRBA, are weakly considered. We recommend paying stronger attention to those aspects inside and outside MFA. For the latter, we recommend actively following and participating in thematic, country-level, regional, and international exchanges to foster peer-to-peer learning for better sustainability, coherence, and impact in the donor community. Inside MFA, we suggest laying stronger foundations ensuring better consideration of CCOs and HRBA already at the planning and designing stage of interventions. One option to do this can be a mandatory focussed capacity development format to obtain a kind of "driving licences for appropriate CCOs and HRBA consideration" by awarding a certificate. The length, mode, and contents of such formats should be well balanced, acknowledging standards set by the Finnish Development Policy and reflected in current guidelines and stakeholders' needs. Given the lack of mainstreaming of CCOs and HRBA in interventions and evaluation reports, we recommend exploiting synergies with respect to the implementation of **R.1.4**.

*This recommendation is mainly linked to conclusions C4, C5 and findings of chapters 5.1, 5.2, 5.4, annex 12.*

*Main implementation responsibility: regional units, project planners, chief technical assistants, M&E officers*

*Urgency: medium*

*Priority: high*

**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.

The qualitative content assessment of lessons learnt and recommendations provided rich evidence on generalisable interventions fields and intervention-specific recommendations to uplift the quality of Finnish development cooperation on bi-, multi-, and multi-bilateral interventions. To make the best use of aggregated results, we recommend ensuring sufficient and appropriate dissemination at individual and institutional levels and fostering uptake. The latter can be facilitated by a smaller group of internal and/or external technical experts delegated to work on translating results of the synthesised lessons learnt and recommendations into specific options for actions. For example, they could provide topic-wise fact sheets or facilitate communities of practice to foster evidence-based learning. They are then disseminated among MFA key stakeholders to support the planning and implementation of Finnish funded development interventions. Potentially



promising dissemination channels comprise feeding them into well-established exchange fora and providing tailored web-based inputs focussing on key topics or more general workshop formats with training inputs and discussion sections. The dissemination channels should also use the Framework Agreement for Planning Bilateral Development Cooperation to facilitate planning for better quality interventions.

As M&E has been identified as one of the major themes, it is recommended to undertake similar efforts of support by a professional M&E specialist. This is of particular importance as a functioning M&E system at the intervention level lays the foundation for methodologically sound and high-level evidence-based evaluations, generating further intervention-specific insights to improve Finnish development cooperation. A departure point would be looking at the quality elements for a functioning M&E system and identifying, with the help of the metaevaluations results, how far and in what aspects the interventions are lagging behind. In the next step, support formats to meet the requirements and needs of stakeholders inside and outside MFA can be developed.

*This recommendation is mainly linked to conclusions C6 and findings of chapters 5.4 and 5.3.*

*Main implementation responsibility: regional units, project planners, chief technical advisors, M&E officers*

*Urgency: medium*

*Priority: high*

### 8.3. Recommendations for future metaevaluation (EQ16)

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

Sound sample representativeness is an important quality criterion to draw valid conclusions from a sample to the whole population. As in the previous metaevaluation, it is not yet possible to calculate the representativeness of the selected evaluations reports on the whole population of Finnish bi-, multi- and multi-bilateral development cooperation. This is a clear limitation of this metaevaluation. Setting up and maintaining an inventory of all interventions and evaluations is thus highly recommended. It not only contributes to enhancing the explanatory power of future metaevaluations but is also expected to save resources. Metaevaluation teams would no longer spend resources identifying and searching for potentially eligible evaluation reports. In addition, an inventory allows the application of efficient sampling strategies. Furthermore, it allows MFA to conduct a stratified sampling of interventions to be evaluated and thus increasing rigour on its evaluation decisions. We recommend a simple inventory with key characteristics on interventions and corresponding evaluations (e.g., budgets, durations, sector, region, nature of the intervention and of the evaluation, commissioner, ToRs). We further recommend keeping the inventory as simple as possible and seeking professional support for its development. To ensure usefulness and maintenance of the inventory, key stakeholders should be consulted during all stages of inventory



development and testing. Advantages of an interface to AHA-systems should be considered, any possible linkages to the development policy system reform (KeTTU) explored, and the system could be started in the context of launching the FADER.

*This recommendation is mainly linked to the opening remarks of the conclusions section.*

*Main implementation responsibility: MFA EVA-11 for set-up, regional units for archiving*

*Urgency: low*

*Priority: high*

**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 metaevaluations and coverage for future metaevaluations.

As discussed in the limitations chapter (2.3), the quality of survey data on the use and usefulness of specific evaluations is likely to be affected by memory gaps, selective recall, and socially desirable. Conducting key informant interviews allowed coping with such limitations at least partially but compromised on the representativeness. Thus, we recommend introducing feedback sheets on decentralised evaluations as a more comprehensive solution. That allows key stakeholders to share their perspectives straight after the closure of an evaluation and thereby enhances the database for future metaevaluations in two ways: It positively impacts data quality and, if properly implemented, enlarges the sample coverage, and thus representativeness of results. Thereby, we recommend also inviting evaluators to share feedback. That would lay the foundation to add their perspectives on the quality of the ToRs and on the evaluation process in future metaevaluations and thus, broaden the evidence base. Beyond serving future metaevaluation, feedback sheets would also benefit MFA in multiple ways: (i) they allow internal immediate quality assessments of evaluations and thus support decision making on future evaluation experts, (ii) they enable just-in-time peer learning from useful evaluation processes and (iii) they can be used by commissioners of evaluations for a self-assessment.

We recommend keeping the feedback sheets short and simple, firmly anchoring their distribution inclusive of friendly reminders in the evaluation process, grounding them on a digital solution, and supporting the process by a professional M&E expert. Linking data with the inventory recommended in R3.1 should be carefully considered but enhances the complexity of the inventory, rather calls for a decentralised system.

*This recommendation is mainly linked to the limitations in chapter 2.3 section.*

*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*



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

This metaevaluation was the first in a series of assignments that could establish a comparison with its predecessor. For the following metaevaluation, this will gain importance, as change requires time and as the period between the recent and the previous metaevaluation was small and only comprised a couple of years for the later produced reports of the sample. Given increased evaluation standards and a plausible time lag for widespread adaptation, future metaevaluations should be grounded on the same methodology to allow inquiring trends over a longer period of time. We recommend ensuring sufficient resourcing of future metaevaluations to leverage insights from a comparison over time. Given the growing body of relevant evidence from earlier endeavours, we recommend either enlarging the budget accordingly or identifying and excluding analyses of lower interest to MFA from future assignments to save resources when budgets remain constant. From the metaevaluations point of view, the qualitative content assessment of underlying reasons for evaluators' assessment on OECD DAC might be a possible field to compromise on insights. As the analysis produced at large comparable results as in the previous metaevaluation while at the same time being very resource-intensive, it might be considered to skip them in one metaevaluation and then look at them again in a subsequent metaevaluation. On a different note, the fact that the database of the recent metaevaluation could not be fully exploited in the course of the assignment calls for enlarging the future budget to gain the best evidence for future learning and systematic reviews.

*This recommendation is mainly linked to implementation challenges in chapter 2.2.*

*Main implementation responsibility: EVA-11*

*Urgency: low*

*Priority: high*



# References

- Finnish Government. (2019). *Programme of Prime Minister Sanna Marin's Government: Inclusive and Competent Finland—A Socially, Economically and Ecologically sustainable society*. Valtioneuvosto. <https://valtioneuvosto.fi/en/marin/government-programme>
- MFA. (n.d.). *Databank on Finland's development cooperation*. OpenAid.fi. <https://openaid.fi/en/>
- MFA. (2004). *Development Policy 2004 [Government Resolution]*. Ministry for Foreign Affairs of Finland.
- MFA. (2007). *Development Policy Programme 2007—Towards a Sustainable and Just World Community*. Ministry for Foreign Affairs of Finland. [https://um.fi/publications/-/asset\\_publisher/TVOLgBmLyZvu/content/kehityspoliittinen-ohjelma-2007](https://um.fi/publications/-/asset_publisher/TVOLgBmLyZvu/content/kehityspoliittinen-ohjelma-2007)
- MFA. (2012a). *Manual for Bilateral Programmes*. Ministry for Foreign Affairs of Finland.
- MFA. (2012b). *Finland's Development Policy Programme [Government Decision in Principle]*. Ministry for Foreign Affairs of Finland (MFA).
- MFA. (2013). *Evaluation Manual*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/384998/385866/evaluation\\_manual/93315abc-02e4-17cf-74f9-566da249552b?t=1528465057886](https://um.fi/documents/384998/385866/evaluation_manual/93315abc-02e4-17cf-74f9-566da249552b?t=1528465057886)
- MFA. (2015a). *Development evaluation norm of the Ministry for Foreign Affairs (MFA) (Norm01/2015)*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/35732/48132/development\\_evaluation\\_norm\\_2015](https://um.fi/documents/35732/48132/development_evaluation_norm_2015)
- MFA. (2015b). *Human Rights based approach in Finland's development cooperation guidance*. [https://um.fi/documents/35732/48132/human\\_rights\\_based\\_approach\\_in\\_finlands\\_development\\_cooperation\\_guidance](https://um.fi/documents/35732/48132/human_rights_based_approach_in_finlands_development_cooperation_guidance)
- MFA. (2016). *Finland's Development Policy: One world, common future—Towards sustainable development*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/35732/48132/government\\_report\\_on\\_development\\_policy\\_2016](https://um.fi/documents/35732/48132/government_report_on_development_policy_2016)
- MFA. (2016b). *Manual for Bilateral Programmes*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/35732/48132/manual\\_for\\_bilateral\\_programmes](https://um.fi/documents/35732/48132/manual_for_bilateral_programmes)
- MFA. (2018a). *Evaluation Manual*. MFA Development Evaluation Unit. <https://eoppiva.zapter.io/evaluationmanual2018>



- MFA. (2018b). *The Development Policy Results Report 2018*. Ministry for Foreign Affairs of Finland. [https://um.fi/publications//asset\\_publisher/TVOLgBmLyZvu/content/kehityspolitiikan-tulosraportti-2018](https://um.fi/publications//asset_publisher/TVOLgBmLyZvu/content/kehityspolitiikan-tulosraportti-2018)
- MFA. (2018c). *Manual for Bilateral Programmes*. Ministry for Foreign Affairs of Finland. [https://um.fi/publication/-/asset\\_publisher/iYk2EknIlmNL/content/manual-for-bilateral-programmes](https://um.fi/publication/-/asset_publisher/iYk2EknIlmNL/content/manual-for-bilateral-programmes)
- MFA. (2020a). *Guideline for the Cross-Cutting Objectives in the Finnish Development Policy and Cooperation*. Ministry of Foreign Affairs of Finland. <https://um.fi/documents/35732/o/Annex+4+++Crosscutting+Objectives+%281%29.pdf/e9e8a940-a382-c3d53c5fdc8e7455576b?t=1596727942405>
- MFA. (2020b). *Theories of Change and Aggregate Indicators for Finland's Development Policy 2020*. Ministry for Foreign Affairs of Finland. [https://um.fi/publications/-/asset\\_publisher/TVOLgBmLyZvu/content/suomen-kehityspolitiikan-tuloskartat-ja-indikaattorit-2020](https://um.fi/publications/-/asset_publisher/TVOLgBmLyZvu/content/suomen-kehityspolitiikan-tuloskartat-ja-indikaattorit-2020)
- MFA. (2021a). *Report on Development Policy Extending Across Parliamentary Terms*. Ministry for Foreign Affairs of Finland. <https://julkaisut.valtioneuvosto.fi/handle/10024/163218>
- MFA. (2021b). *Development cooperation evaluation manual*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/384998/o/Evaluation+Manual\\_2021+%281%29.pdf/6fe6d6a9-86e2-2f5b-4d35-fa232582a703?t=1617079302186](https://um.fi/documents/384998/o/Evaluation+Manual_2021+%281%29.pdf/6fe6d6a9-86e2-2f5b-4d35-fa232582a703?t=1617079302186)
- MFA. (2021c). *Development cooperation appropriations*. Ministry for Foreign Affairs. <https://um.fi/development-cooperation-appropriations>
- OECD. (2010). *Glossary of Key Terms in Evaluation and Results Based Management*. Organisation for Economic Co-operation and Development (OECD). <http://www.oecd.org/development/peer-reviews/2754804.pdf>
- OECD. (2021). *Applying Evaluation Criteria Thoughtfully*. OECD Publishing. [https://www.oecd-ilibrary.org/development/applying-evaluation-criteria-thoughtfully\\_543e84eden?\\_ga=2.236907143.748008699.1628759336-43054910.1628499330](https://www.oecd-ilibrary.org/development/applying-evaluation-criteria-thoughtfully_543e84eden?_ga=2.236907143.748008699.1628759336-43054910.1628499330)
- OECD DAC. (1991). *DAC Criteria for Evaluating Development Assistance*. Organisation for Economic Cooperation and Development (OECD). <http://www.oecd.org/dac/evaluation/49756382.pdf>
- Silvestrini, S. Väh, S.J., Römling C., Mikkolainen, P., Lieckefett, M. (2017). *Metaevaluation of project and programme evaluations 2015-2017*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/384998/385866/meta\\_evaluation\\_report\\_2018](https://um.fi/documents/384998/385866/meta_evaluation_report_2018)



# The Assignment, the Evaluation Team, and the Quality Assurance Process

This metaevaluation was implemented under an Evaluation Management Services (EMS) framework agreement contracted to a consortium by Particip GmbH and NIRAS. The assignment started with the provision of a concept note for the metaevaluation by EVA-11. After approval of the team leader by EVA-11, drafting the ToR, proposing evaluation team members and budgeting started. After revisions of the draft ToR by the team leader according to EVA-11's oral and written comments, the consortium submitted the draft final ToR and the draft final budget. With their approval by EVA-11, the second service order, i.e., the standard evaluation procedure, began. The EMS Service Coordinator facilitated both service orders and served as interlocutor and quality assurance expert between the parties. Quality was further ensured by the consortium's internal quality assurance as well as by a reference group established by EVA-11. The overall budget was 286,000 EUR.

Results were validated by EVA-11 and stakeholders who were at the same time members of the reference group. They shared feedback during the findings, conclusions, and recommendations (FCR) workshop and commented on the draft metaevaluation report.

The assignment was conducted by a team of six persons led by Dr Susanne Johanna V  th as Team Leader and Dr Stefan Silvestrini as Deputy Team Leader. Dr Hansj  rg Gaus, a methodological expert and Petra Mikkolainen, a specialist on Finnish Development Cooperation, support the team as Senior Evaluators. Dr. Maja Flaig, an expert in meta-analysis and Janis Wicke, an expert on qualitative data analysis, complement the team as Evaluators. The multi-disciplinary and gender-mixed team benefits from complementary competencies while fulfilling the standards set in the tender. Given the tight metaevaluation schedule, the size of the evaluation team is justified.

## **Dr. Susanne Johanna V  th**

As the Team Leader, Susanne Johanna V  th took the overall responsibility for the assignment and was involved in all stages of the analysis. She worked in close exchange with the MFA and the evaluation service management coordinator to develop precise ToRs during service order one and enable a well-suited evaluation team. During the inception phase, she was responsible for updating the general metaevaluation approach and developing the methodological design for the new component 3 on usage and usefulness assessment and for drafting the inception report. Further, she guided comprehensive pre-testing and inter-team calibration to ensure inter-rater reliability. In the implementation phase, she was involved in the quality and content assessments of the reports to be analysed and provided backstopping, supervised data collection, and contributed to cross-checks. To facilitate a joint analysis, Susanne Johanna V  th was in close exchange with all team members, provided individual guidance for synthesis and created space for collaborative exchange. Finally, she took the lead during results reporting and ensured proper presentation of metaevaluation results.





### **Dr. Stefan Silvestrini**

As Deputy Team Leader, Stefan Silvestrini worked in close cooperation with Susanne Johanna Väh. During the inception phase, he primarily took the role of a critical friend to strengthen methodological revisions, identify potential data gaps and mitigation strategies, and bring in additional insights based on his far-reaching experience in the field of metaevaluation. Further, he supervised programming, pre-testing and technical implementation of the survey. In the implementation phase, he was responsible for the random selection of evaluation reports and the implementation of independent cross-checks on quality and content assessment. Further, he conducted interviews with key informants to assess the usage and usefulness of the evaluation reports.

### **Dr. Hansjörg Gaus**

As a meta-evaluator with a strong methodological background, Hansjörg Gaus supported Susanne Johanna Väh in developing the metaevaluation design for the new component 3. Thus, he contributed to developing the corresponding data collection instruments for key informant interviews and the online surveys during the inception phase. He was substantially involved in the quality assessment of the reports to be analysed during the implementation phase and contributed to the content assessment. In joint analysis, Hansjörg Gaus took a considerable stake in synthesising findings from the quality and content analysis.

### **Petra Mikkolainen**

As a meta-evaluator with in-depth knowledge of Finnish development cooperation, Petra Mikkolainen was substantially involved in the context analysis during the inception phase. Further, she supported Susanne Johanna Väh in updating the methodology to ensure its appropriateness with respect to the specificities of Finnish development cooperation. During the implementation phase, she was strongly involved in content analysis. Further, she conducted interviews with key informants to assess the usage and usefulness of the evaluation reports. With regard to the joint analysis, Petra Mikkolainen took a considerable stake in synthesising results related to addressing the priority areas of Finnish development policy, cross-cutting objectives and general recommendations from the content assessment.

### **Dr. Maja Flaig**

As an expert in the meta-analysis, Maja Flaig was involved in the quality and content assessment of the reports to be analysed. Furthermore, she took a considerable stake regarding the implementation of the online surveys and performed quantitative data analysis. During the joint analysis, she supported synthesising results from the quality and content assessment, as well as from the usage and usefulness assessment. Furthermore, she worked on database maintenance and supported documentation throughout the implementation and reporting phases.

### **Janis Wicke, M.A.**

As an expert in qualitative data collection and analysis, Janis Wicke supported the evaluation team with the qualitative component of the content analysis. He was in charge of analysing and aggregating the lessons learnt and recommendations prevalent in the sampled evaluation reports. Drawing on qualitative content analysis, he further worked on a set of categories describing the most frequent reasons underlying the evaluators' positive or negative assessment of interventions



with respect to their performance against the OECD-DAC criteria. It was also his task to present the findings of this analysis in this report.

### **Work plan and communication and dissemination plan**

In general, all tasks completed were assigned to at least two team members ensuring continuous application of the four-eyes principle. Supervision and backstopping by the team leader were provided during all phases of the assignment. In addition, the deputy team leader looked particularly through the lens of an internal quality assurer on the assignment. Thus, several measures were taken to avoid individual biases, include multiple perspectives, foster close inter-team exchange, and ensure high quality of results.

Table 5 presents a work plan displaying single task, division of labour and responsibilities, individual working days per assigned tasks and month(s) of implementation. Table 6 below presents the outline of the evaluation's communication and dissemination plan.



**Table 5: Work plan with a division of assigned tasks and timeline**

<b>Legend:</b>	<b>SJV</b>	<b>StS</b>	<b>HG</b>	<b>PM</b>	<b>MF/ JW</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>PHASE 0 AND 1: PLANNING AND START-UP PHASE</b>	<b>8</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>										
Kick-off meeting SO1 (08.06.2021) and internal meetings															
Pre-proposal budget and components															
Draft ToR incl. team composition (25.6.2021)															
<b>PHASE 2: INCEPTION PHASE</b>	<b>19,5</b>	<b>10,5</b>	<b>5</b>	<b>10</b>	<b>5</b>										
Review of relevant documents and first screening of exemplary reports															
Identifying and filling potential data gaps (e.g additional data request, explorative interviews if necessary)															
Updating methodology incl. analysis grid; quality & ToR assessment tool (component 1), content assessment tool (component 2) incl. operationalisation according to OECD DAC update, minimal criteria for inclusion into summative meta-analysis, developing quantitative content analysis for Finnish priority areas and cross-cuttings															
Developing methodology incl. analysis grid for usage and usefulness (component 3) incl. operationalisation (i.e., interview guidelines with open and standardised questions for commissioners, survey for implementers)															
Drafting of inception report incl. finalisation of work plan and division of labour, deadline: 13.08.2021															
Inception meeting with reference group, administrative meeting (incl. preparation, documentation) 03.09.2021, virtual															
Finalisation of inception report, approval: 10.09.2021															
Pre-test of assessment tools incl. virtual inter-team coordination, calibration workshop with adjustments if necessary															
Programming and pre-test of online survey															
<b>PHASE 3: IMPLEMENTATION PHASE</b>	<b>37</b>	<b>15,5</b>	<b>37</b>	<b>32</b>	<b>42</b>										
Quality assessment of 80 evaluation reports															
Content assessment of 72 evaluation reports															
26 key informant interviews incl. scheduling															
Online survey															
Quantitative content analysis i.e key word analysis for CCO															
Joint analysis and synthesis incl. internal workshop via Skype (Component 1: 17.12.2021, Component 2: 21.12.2021, Component 3: 4.1.2022)															
External validation FCR meeting incl. preparation, 13.01.2022, virtual															
<b>PHASE 4: REPORTING &amp; DISSEMINATION PHASE</b>	<b>12,5</b>	<b>6,5</b>	<b>5</b>	<b>5</b>	<b>5</b>										
Drafting of the metaevaluation report, 07.02.2022															
Finalisation of the report based on comments (received by 17.02.2022): 28.02.2022															
Management meeting and public presentation of results incl. preparation 31.03. March 2022, in Helsinki															
Webinar of results incl. Preparation 31.03.2022 March 2022, virtual															
<b>Total</b>	<b>77</b>	<b>32,5</b>	<b>47</b>	<b>47</b>	<b>52</b>										

Note: SJV: Susanne J. Våth, Team Leader, StS: Stefan Silvestrini, Deputy Team Leader, HG: Hansjörg Gaus, Methodological Expert, PM: Petra Mikkolainen, Finn Evaluation Expert, MF: Maja Flaig, Meta-analysis expert, JW: Janis Wicke, Qualitative analysis expert (recruited in December 2021) WD: Working day.



**Table 6: Communication and dissemination plan**

<b>COMMUNICATION &amp; DISSEMINATION OF EVALUATION DELIVERABLES</b>	<b>TARGET AUDIENCE</b>	<b>TIMING</b>	<b>RESPONSIBLE ACTORS</b>
Distribution of final inception report	EVA-11, Reference Group	September 2021	EVA 11-Evaluation Manager
Introduction letter	MFA internal stakeholders	October 2021	EVA 11-Evaluation Manager
Findings-Conclusions-Recommendations workshop	EVA-11, Reference Group	January 2022	Evaluation Team
Public presentation event	MFA staff Evaluation and development cooperation community	March 2022	EVA 11-Evaluation Manager & Evaluation Team
Webinar	MFA staff Evaluation and development cooperation community in Finland and abroad		
Evaluation Summary (3-pager)	The general public, in addition to the stakeholders mentioned above	Beginning of implementation phase	Team Leader
Evaluation Summary (4-pager)	The general public, in addition to the stakeholders mentioned above	March 2022	Team Leader (draft), EVA 11-Evaluation Manager (finalisation)
A debriefing meeting with the management and EVA-11	MFA management, EVA-11	Before the final presentation	Team Leader
After the evaluation process has been concluded, if necessary, thematic workshops (peer learning) (currently not considered in the evaluation budget)	EVA-11 Manager, relevant MFA Units and Departments	After the final presentation	Team Leader
Disseminating the evaluation in social media	The general public, in addition to the stakeholders mentioned above	March 2022	EVA-11 Evaluation Manager NIRAS



# Annex 1: Terms of Reference

## Terms of Reference (ToR)

### Metaevaluation OF MFA'S PROJECT AND PROGRAMME EVALUATIONS IN 2018 – 2020

#### Rationale of the evaluation

The Ministry for Foreign Affairs of Finland (MFA) assesses Finnish development cooperation by carrying out two types of evaluations. One type is the comprehensive, policy level evaluations (centralized evaluations) managed by the Development Evaluation Unit (EVA-11). The second type is the project and programme evaluations (decentralized evaluations) commissioned by the units, departments or embassies responsible for the project or programme in question. Evaluations are conducted in order to learn from successes and setbacks, and the lessons learned are used for developing future cooperation. The development evaluations are based on the Evaluation norm of the Ministry for Foreign Affairs of Finland 1/2015.

EVA-11 commissions regularly metaevaluations to synthesize findings, explore issues and assess the reliability of decentralized evaluations to facilitate learning from MFA's evaluation activities and products. This provides a comprehensive overview of scattered findings and various lessons learned from a wide range of different development cooperation interventions funded by Finland. Further, it serves for accountability with respect to the evaluation function of MFA during a certain period of time by classifying decentralized evaluation reports by commissioner, evaluation type, sector etc. and by assessing the quality and synthesizing the content of the reports. This results in improved transparency towards partner countries, general public, parliamentarians, academia, media and development professionals outside the MFA.

Beyond purely desk-based analysis, the perspectives of key stakeholders from MFA, Finnish embassies and project implementers can add important insights regarding the use of decentralized evaluations and their usefulness, and thus, further foster learning, provide guidance for development cooperation and ensure that the evaluations will be used and useful by improving evaluation activities and products.

This is the Terms of Reference (ToR) for the metaevaluation of project and programme evaluations (decentralized evaluations) carried out between 2018 and 2020. The evaluation will be based on the assessment of the decentralized evaluation reports and corresponding Terms of References (ToR) documents as well as on primary data collection from key stakeholders.

#### Purpose and objectives of the evaluation

**The purpose** of the metaevaluation is to continue the series of metaevaluations in order to see the development since the previous analysis, to learn from successes and setbacks as well as to see how the recommendations of the previous metaevaluations have been used and whether they have led to better quality of evaluations and development results. In addition, the results of the metaevaluation will be used when drafting the Development Policy Results Report 2022 to the Parliament.



**The primary users** of the metaevaluation are the MFA's staff responsible for design and implementation of development projects and programmes, the Department for Development Policy as well as related evaluations, CSOs and experts conducting evaluations.

**Overall objective of the metaevaluation is** to provide concise conclusions and recommendations that help the MFA enhancing the quality of evaluations, the evaluation management practices and its overall evaluation capacity development improving its development cooperation based on aggregated insights from decentralized evaluations regarding quality and results of Finland's development cooperation, identified issues and lessons learned increasing the usage and usefulness of decentralized evaluations for results-based management, evidence informed learning and decision making in the future based on primary data collected from key stakeholders.

**The evaluation consists of three components with the following objectives:**

**1) The metaevaluation of the quality of the decentralized evaluations 2018 – 2020** to update the standardized quality assessment tool of methodological issues to assess the quality of the evaluation reports and Terms of References by using the standardized assessment tool to put the results of the metaevaluation into perspective to the earlier metaevaluation of decentralized evaluation reports 2015 – 2017

**2) The meta-analysis (content analysis) of overall quality and results of development cooperation based on decentralized evaluations 2018 – 2020** to update the standardized assessment tool for content analysis of the decentralized evaluations to extract reliable evaluation findings and issues rising from decentralized evaluation reports on Finland's development cooperation by using the standardized assessment tool to form a synthesis of reliable evaluation results (findings and conclusions) and recommendations and issues on development policy and cooperation emanating from decentralized evaluation reports to put the results of the meta-analysis into perspective to the earlier metaevaluation of decentralized evaluation reports 2015 – 2017 to provide information for results reporting of Finnish Development Cooperation based on the decentralized evaluation reports.

**3) The assessment of usage and usefulness of the decentralized evaluations 2018 – 2020** to develop a methodology for this newly added component that focuses on the use and usefulness of decentralized evaluations to assess the usage and usefulness of decentralized evaluations with respect to evaluation processes and reports for commissioners and implementers

## Scope of the evaluation

The scope of the metaevaluation consists of the decentralized evaluation reports on development policy and cooperation commissioned by the MFA between 2018 and 2020. The focus is on the MFA's mid-term and final evaluations of individual projects and programmes (bilateral, multilateral and multi-bi) conducted by various units and departments of the MFA, embassies and multilateral and other implementing partners. So called multi-bi projects/programmes are funded partly by MFA. The administration of these projects and their evaluations may have been done by a partner organization in which case MFA has participated in commenting ToRs and evaluation reports but has not been the commissioner of the evaluation.

The key documents to be analyzed include evaluation reports (mid-term evaluations, final evaluations, ex-post evaluations and impact evaluations, and their ToRs as key planning document.



Appraisal reports will be excluded from this metaevaluation altogether as they are considered to be planning document instead of evaluations. The subject to this metaevaluation is approximately 80 decentralized reports which all undergo quality assessment.

In the metaevaluation (component 1) the reliability of the reports and their ToRs will be assessed applying the OECD DAC evaluation principles and standards. Those reports complying with minimal methodological standard are eligible for the summative meta-analysis (component 2). This ensures that only reliable evaluation findings will be synthesized to demonstrate how the goals of Finnish development policy priority areas have been achieved. In the meta-analysis development results are analysed along the OECD DAC evaluation criteria.

The meta-analysis will also produce an overview of MFA's decentralized evaluation activities by classifying evaluation reports by key characteristics (e.g. regional spread, thematic focus, funding etc.) and by providing comparisons of relevant sub-groups (e.g. MFA-commissioned vs. other commissioner).

The methodology and standardized assessment tool for metaevaluations will be updated for this upcoming evaluation by the evaluation team. In addition, the usage and usefulness of decentralized evaluations will be explored by using surveys and interviews (component 3). The implementation of additional participatory methods like focus group discussions will depend on considerations regarding feasibility and expected value-addition. The informants are the users of the evaluations at MFA, embassies and among stakeholders as well as the implementers of evaluated projects and programmes.

## Evaluation questions (MFA, TL)

In the following sets of detailed evaluation questions are presented:

**Component 1: Metaevaluation** to assess the state of evaluation capacity:

How can MFA's decentralized **evaluation portfolio** be described (e.g. by commissioner, budgets, sector and regional distribution, evaluation type)?

How is the **quality and reliability of decentralized evaluation reports**? Is there a difference between the quality of MFA-commissioned evaluation reports and the quality of evaluation reports that are commissioned by MFA's partners? Are there other **differences** between relevant sub-groups (e.g. final vs. mid-term evaluations, individual/independent evaluator vs. consulting firm/institutes) if applicable?

How is the **quality of the ToRs** of decentralized evaluations? Are there **systematic patterns** regarding the **quality ToRs** and the **quality of the corresponding evaluation reports**?

What are key differences and possible trends compared to the **previous metaevaluation**? Are there **gaps regarding MFA's evaluation capacity**?

**Component 2: Meta-analysis** to serve the needs for results reporting:

What can be said about the **quality of Finnish development cooperation** according to the **OECD DAC criteria** based on the reliable decentralized evaluation reports (i.e. relevance, effectiveness, efficiency, impact, sustainability, coherence)?



What can be said about the **consideration of policy priority areas, HRBA and cross-cutting objectives** in Finnish development cooperation based on the reliable decentralized evaluation reports?

What are the **major recommendations and lessons learned to improve Finnish development cooperation** emerging from the reliable decentralized evaluation reports?

What can be said about the **overall quality of Finnish development cooperation** based on the reliable decentralized evaluation reports? What are the **major strengths and major challenges** emerging from the reliable decentralized evaluation reports? Are there **typical factors for success or challenges**?

What are **key differences and possible trends** compared to the overall results of the **previous metaevaluation**?

### **Component 3: Usage and Usefulness assessment:**

How are decentralized **evaluations reports used by key stakeholders** (i.e. representatives of MFA, embassies, implementers)?

How **useful** are decentralized evaluations (i.e. **processes and reports**) assessed by key stakeholders? To what extent are decentralized evaluations **timely**? To what extent provide decentralized evaluations **relevant and realistic recommendations**?

Are there **typical factors facilitating or hampering the usefulness** of decentralized evaluations (e.g. report quality)? What are the **major recommendations and lessons learned to improve the use and usefulness** of decentralized evaluations based on the **perspectives of key stakeholders**?

### **Overall recommendations from the assignment:**

What are **recommendations to improve the quality of MFA's decentralized evaluations reports** (and processes if applicable)?

What are **recommendations to improve the quality of Finnish development cooperation**?

What are **recommendations to improve the usage and the usefulness** of decentralised evaluations?

What are **recommendations to improve the methodology of metaevaluations** in the future?

The detailed evaluation questions will be reviewed during the inception phase and finally agreed upon in the approved Inception Report (IR).





## General approach and methodology

The main method used for component 1 and 2 of the metaevaluation will be document review and thus, secondary data analysis.

Assessment tools developed in the course of the last metaevaluation will be used as much as possible to ensure a high comparability of the periods 2015-2017 and 2018-2020. However, given modifications of the OECD DAC criteria and Finnish policy priority areas, some adjustments are necessary. The methodology for both metaevaluation and meta-analysis will be clearly described as well as the criteria based on which the reliability of evaluation reports is assessed.

The main sources of information will be the decentralized evaluation reports (mid-term evaluations, final evaluations, ex-post evaluations, impact evaluations) and their corresponding ToRs as well as Development Policy Programme documents, guidelines, the previous metaevaluation, the management response of the previous metaevaluation, Government Reports to the Parliament and administrative in-house norms.

As evaluation reports under consideration considerably vary with regard to thematic focuses, context conditions, implementing partner organizations, scope and scale of the evaluation as well as evaluation designs and data sources, a high degree of content-related and methodological heterogeneity has to be taken into consideration for the quality assessment.

A revised checklist with criteria and sub-criteria enabling a fair and adequate grading will be repeatedly applied. It grounded originally on insights from MFA's earlier metaevaluations and will be modified to above-mentioned conceptual changes and feedback provided by MFA during the inception phase. Criteria comprise but are not limited to credibility, completeness, adequacy of documentation and appropriateness of evaluation methods applied. Issue areas for modifications comprise e.g., the introduction of the coherence criterion.

The evaluation team is expected to maintain and develop where necessary a four-step grading system with unambiguous grades to facilitate objective rating. The revised assessment tool has to be adjusted in line with MFA's feedback and pre-tested after approval by MFA. Findings of the quality assessment will be aggregated and presented in summarizing results tables to identify general trends, display heterogeneity and prepare the ground for enhancing the quality of evaluations. They will be further contextualized with findings from the previous metaevaluation.

In a second-stage a content assessment provides insights on the joint contribution of MFA's development cooperation and will be conditional on minimal methodological standards in the context of the available material and comparable assignments. The evaluation team will also identify any emerging issues, both positive and negative, from the material.

The evaluation team is expected to cross-analyse approximately 10% of all reports using random selection in order to prove avoidance of significant subjective biases.

For component 3 of the metaevaluation a mix of qualitative and quantitative primary data will be collected and analysed to capture the perspectives of key stakeholders regarding the actual use and the usefulness of decentralized evaluation reports (and processes if applicable). The main source of information are representatives of the MFA, Finnish embassies and project and programme implementers.



Primary data will be collected by conducting semi-structured key informant interviews and by launching two standardized online surveys; one among commissioners and one among implementers of the interventions. As due to their different mandates commissioners and implementers have different information needs regarding decentralized evaluations, their assessment regarding the usage and the perceived usefulness of decentralized evaluations is expected to vary. Thus, two tailor-made online surveys are expected to be the most efficient way to pay attention to such specificities and to gain valuable insights from these stakeholders going beyond generic statements. Qualitative and quantitative data provide for comprehensive and ideally representative insights on how decentralized evaluations are used and how useful they are. The evaluation team will also give attention to recommendations from key stakeholders towards further improving decentralized evaluations. Data analysis complies with good evaluation practice and comprise qualitative content analysis and descriptive statistical analysis.

Finally, the evaluation team provides synthesized conclusions from the results of each of the three components of the metaevaluation and derives corresponding recommendations.

The consultant is encouraged to raise issues that are important to the evaluation but are not mentioned in this ToR. Similarly, in consultation with EVA-11, the consultant might exclude issues that are in the ToR but may not be feasible and those remarks will be presented by latest in the inception report and agreed upon with its approval.

The evaluation must be gender and culturally sensitive and respect the confidentiality, protection of source and dignity of those interviewed.

## **Management of the evaluation**

The evaluation is commissioned by the EVA-11. The Evaluation Manager of EVA-11 will be responsible for the overall management of the evaluation process. The Evaluation Manager will work closely with other units/departments of the MFA and other stakeholders in Finland and abroad. This evaluation is managed through the EMS and it will be conducted by an independent evaluation team recruited by the EMS service provider (Particip GmbH – NIRAS Finland Oy).

There will be one Management Team responsible for the overall coordination of the evaluation. This consists of the EVA-11 Evaluation Manager, the evaluation Team Leader and the EMS Service Coordinator. A reference group (RG) for the evaluation will be established and chaired by the EVA-11 Evaluation Manager.

The RG facilitates the participation of relevant stakeholders in the design and scoping of the evaluation, informing others about the progress of the evaluation, raising awareness of the different information needs, quality assurance throughout the process and in using and disseminating the evaluation results. The mandate of the reference group is to provide quality assurance, advisory support and inputs to the evaluation e.g., through participating in the planning of the evaluation and commenting deliverables. The use of an RG is key in guaranteeing the transparency, accountability and credibility as well as use of evaluation and in validating the results.



The members of the RG are:

Suvi Virkkunen (the deputy will be nominated later), KEO-10

Jaakko Jakkila and Arto Valjas, ALI

Minna Hares (the deputy is Antero Klemola), ASA

The RG may be amended during the inception phase.

The evaluation team will be managed by the Team Leader. This requires careful planning to ensure that a common, consistent approach is used, in order to achieve comparability of the data gathered and the approach used in analysis. The Team Leader will develop a set of clear protocols for the team to use, and will convene regular team meetings to discuss the approach. During the process particular attention should be paid to strong inter-team coordination and information sharing within the team. The evaluation team is responsible for identifying relevant stakeholders to be interviewed and organizing the interviews. The MFA and embassies will not organize these interviews or meetings on behalf of the evaluation team, but will assist in identification of people and organizations to be included in the evaluation.

## Evaluation process, timelines and deliverables

The evaluation will take place during 2021/2022. It began in May 2021 by launching the process for identifying Team Leader candidates. The evaluation follows the general phasing of the Evaluation Management Services (EMS) framework used by the Development Evaluation Unit (EVA-11). The timetables are tentative, except for the final report.

**Phase A: Planning phase: Preparation of the draft Terms of Reference:** Deadline for the draft ToR: 23 June 2021

**Phase B: Start-up phase: June 2021** Start-up meeting: 8 June 2021 Finalisation of the ToR and submission for approval, recruitment of the evaluation team members: by 5 July 2021

**Phase C: Inception phase: July - September 2021** Submission of Draft Inception Report by 13 August 2021 Inception Meeting on 3 September 2021 (tentative), followed by Administrative Meeting Final Inception Report by 10 September 2021

**Phase D: Implementation phase: September 2021- January 2022** Review of decentralized evaluation reports, implementation of expert interviews and online surveys

**Phase E: Reporting/Dissemination Phase: mid-January – end March 2022** Findings, conclusion and recommendations (FCR) workshop, mid-January 2022 Draft Final Report submission by 7 February 2022 (tentative) Final Report 28 February 2022 (tentative) Webinar: end March 2022 (tentative)

## Deliverables and reporting

The language of all reports and any other deliverables is English, while abstracts and summaries of final evaluation report will be in Finnish and Swedish. Deliverables are as follows:

**Phase A (planning phase):** Team Leader submit the draft ToR in liaison with the Evaluation Manager and the EMS Service Coordinator. After approval, a half-page summary of the ToR and evaluation process will be prepared for dissemination to wider audiences and publications.



**Phase B (start-up phase):** Presentation of the approach and methodology by the Team Leader. Selection of the team members (TL already approved) and preparation of the draft cost-calculator.

**Phase C (inception phase):** The inception phase includes in-depth desk analysis and preparation of detailed metaevaluation plan. The desk study includes a comprehensive context and document analysis based on the previous meta-evaluation and policy documents, developing and finalizing the assessment tools and preparation of a detailed evaluation plan. The inception report consists of the detailed metaevaluation plan and finalized assessment tools including introduction and context; finalization of the methodology and assessment tools; final work plan and division of work between team members; tentative table of contents of final report; data gaps and other limitations and their mitigation; detailed implementation plan for quality assessment, content analysis and for usefulness and utility assessment; budget; and communication and dissemination plan.

**Phase D (implementation):** The implementation phase includes a review of 80 decentralized evaluation reports, conducting expert interviews and implementing two online surveys. During the review particular attention should be paid to the methodological quality of the decentralized evaluations at hand (component 1) and to their content with respect to the OECD/DAC criteria, the priority areas of Finnish Development Cooperation as well as on lessons learnt and recommendations (component 2). In expert interviews and online surveys with representatives of different stakeholder groups attention has to be paid to the utility of decentralized evaluations and their perceived usefulness. The team is encouraged to use statistical evidence whenever possible and to triangulate primary and secondary data.

Direct quotes from interviewed stakeholders may be used in the reports, but only anonymously ensuring that the interviewee cannot be identified from the quote.

**Phase E (reporting/communication phase):** A Findings-Conclusions-Recommendations (FCR) workshop will be organised after reviewing all evaluation reports and after completion of the collection of primary data. In the meeting initial findings (and very preliminary conclusions or recommendations) will be presented. It is planned to be held in mid-January.

The draft final report must include abstract and summary in English (including the table on main findings, conclusions and recommendations). It must be of high and publishable quality. It must be ensured that the translations use commonly used terms in development cooperation. The consultant is responsible for the editing, proof-reading and quality control of the content and language.

Production of the first draft of the 4-pager for communication purposes will be the responsibility of the TL/Evaluation Team of the evaluation. The first draft will be provided simultaneously with the first draft of the final report.

The final report must include abstract and summary (including the table on main findings, conclusions and recommendations) in Finnish, Swedish and English. The final report will be delivered in Word-format (Microsoft Word 2010) with all the tables and pictures also separately in their original formats. The revised reports have to be accompanied by a table of received comments and responses to them. In addition, the MFA requires access to the evaluation team's tools, data sets or interim evidence documents, e.g. completed matrices, although it is not expected that these should be of publishable quality. The MFA treats these documents as confidential if needed.



Each deliverable is subjected to specific approval. The evaluation team is able to move to the next phase only after receiving a written statement of acceptance by the MFA. In addition to written deliverables, the Team Leader and the evaluation team are expected to participate in workshops and give oral presentations, often supported by PowerPoint slides (esp. during phases D and E). Should the Covid-19 situation allow, the publication event will be held in Helsinki, with the team leader and at least one evaluation team members present. In addition, the team leader and other team members will give short presentation of the findings in a public Webinar. This presentation can be delivered from distance. In the event of continued travel restrictions, these two publication events may be combined. The consultant is expected to provide agreed visual materials, such as minimum of 3 alternatives for cover picture, for EVA-11's acceptance. The MFA will prepare a management response to the recommendations.

## Expertise required

One Team Leader level expert will be identified as the Team Leader of the whole evaluation. The Team Leader will lead the work and will be ultimately responsible for the deliverables. The evaluation team will work under the leadership of the Team Leader who carries the final responsibility of completing the evaluation.

It is foreseen that the evaluation team consists of a Team Leader, Deputy Team Leader, two senior evaluators (one Finnish speaking) and one evaluator. For the exceptional case that a team member might have been involved in a decentralized evaluation, this team member will be completely excluded from assessing that report to avoid potential conflict of interest. The tight schedule requires a sufficiently large evaluation team to review and analyse 80 decentralized evaluation reports, their utility and usefulness. The skills and experience of the proposed experts have to correspond or exceed the minimum requirements of the evaluation team members. The EVA-11 will approve the experts.

The evaluation team should cover the following knowledge/expertise areas: strong expertise on metaevaluations, meta-analyses and systematic reviews proven experience of quantitative methods and statistical analysis of data as well as qualitative content analysis and data literacy expertise and experience in applying the qualitative data analysis software MaxQDA and the statistical data analysis package SPSS, STATA or R. expertise and experience in conceptualizing and implementing complex evaluations be familiar with the principles of Results Based Management (RBM) and measuring results readiness to use variety of data collection methods (e.g. survey, in-depth interviews, participatory methods) expertise and experience in comprehensive development policy and cooperation evaluations readiness to disseminate evaluation results and recommendations supporting decision making as well as programme design and learning knowledge of the main goals and priorities of the Finnish development policy and cooperation expertise in applying and evaluating Ministry's cross-cutting objectives and human rights-based approach in development policy and cooperation.

The competencies of the team members shall be complementary and jointly meet the expertise areas as mentioned above. However, given the nature of the metaevaluation proven in-depth knowledge on quantitative and qualitative methods are mandatory for all team members. All team members shall have fluency in English and at least one senior evaluator must have fluency in Finnish, because part of the documentation is available only in Finnish. MFA document material classified



as restricted use (classified as IV levels in the MFA, or confidential in other organizations) cannot be saved, processed or transmitted by any cloud services or unsecured emails and google translators or any other web-based translators cannot be used to translate these documents.

The Team Leader and the team have to be available until the reports have been approved by the EVA-11, even when the timetables change.

## Quality assurance

**Internal quality assurance** The consortium implementing this evaluation will put in place a three-layer system of quality assurance for all products/reports: at the level of the Team Leader of the individual evaluation, through the EMSC&D and through in-house senior QA advisors. Consultant is in charge of the impeccable quality of English, Swedish and Finnish texts of the reports and related proofreading. The EMSC will be responsible for Finnish translations of good quality. All deliverables shall be of publishable quality.

The evaluation team should make their best efforts not to exceed the total length of 80 pages for the main evaluation report and to prepare an executive summary that is publishable as a stand-alone document and that includes visualisations to ensure the summary is reader-friendly also for less closely involved stakeholders in the Vietnam transition process. A separate volume on annexes may be produced. This will be agreed during the inception phase, which of the final deliverables are to be published. The report should be kept clear, concise and consistent.

The report must follow the writing instructions and template provided by the MFA and it should contain, inter alia, the evaluation findings, conclusions and recommendations. The logic between those should be clear and based on evidence. The final draft report(s) will be sent for a round of comments by EVA-11. The purpose of the comments is only to correct any misunderstandings or factual errors. External quality assurance It should be noted that internationally recognized expert has been contracted by EVA-11 as a Critical Friend (external peer reviewer) for the whole evaluation process. The person interacts directly with EVA11 and provides expert opinions on the planning and implementation of the evaluations. EVA-11 may or may not integrate any such external advice as part of their overall feedback and management responses of the evaluation.

**External quality assurance** It should be noted that internationally recognized expert has been contracted by EVA-11 as a Critical Friend (external peer reviewer) for the whole evaluation process. The person interacts directly with EVA11 and provides expert opinions on the planning and implementation of the evaluations. EVA-11 may or may not integrate any such external advice as part of their overall feedback and management responses of the evaluation.



## **Budget**

The evaluation will not cost more than 286,000 Euros (VAT excluded).

## **Mandate**

The evaluation team is entitled and expected to discuss matters relevant to this evaluation with pertinent persons and organizations. However, it is not authorized to make any commitments on behalf of the Government of Finland or the Ministry. The evaluation team does not represent the Ministry for Foreign Affairs of Finland in any capacity.

All intellectual property rights to the result of the Service referred to in the Contract will be exclusive property of the Ministry, including the right to make modifications and hand over material to a third party. The Ministry may publish the end result under Creative Commons license in order to promote openness and public use of evaluation results.

## **Authorisation**

Helsinki, 6.7.2021

Anu Saxén

Director, Development Evaluation Unit

Ministry for Foreign Affairs of Finland

Please add here the original terms of reference of the evaluation assignment.



## Annex 2: People Interviewed

NAME	ROLE	CURRENT POSITION	ORGANISATION	DEPARTMENT	UNIT
Ville Lahelma	HQ	Desk Officer	MFA	Political Department	Unit for Arms Control
Suvi Sipilä	HQ	Programme Officer	MFA	Africa and the Middle East	Middle East/ North Africa
Anu Eskonheimo	HQ	Programme Officer, Ethiopia	MFA	Africa and the Middle East	Unit for the Horn of Africa and Eastern Africa
Mika Vehnämäki	HQ	Commercial Counsellor	MFA	International Trade	Sustainable Trade Unit
Henna Riiikka Pihlapuro	HQ	Desk Officer, Concessional credits	MFA	Development Policy	Development Finance and Private Sector Cooperation
Hannele Tikkanen	HQ	Desk Officer	MFA	Development Policy	Unit for Development Finance and Private Sector Cooperation
Sanna-Liisa Taivalmaa	HQ	Senior Adviser	MFA	Development Policy	Unit for Sectoral Policy
Marjo Ahvenainen	HQ	Desk Officer	MFA	Russia, Eastern Europe, Central Asia	Eastern Europe and Central Asia
Tuulikki Parviainen	HQ	Team Leader	MFA	Africa and the Middle East	Southern and Western Africa
Venla Voutilainen	HQ	Program Manager	MFA	Americas and Asia	Eastern Asia and Oceania
Minni Hyrkkänen	HQ	Desk Officer	MFA	Political Department	Management of Political Department
Juoko Eskelinen	HQ	Senior Advisor	MFA	Russia, Eastern Europe, Central Asia	Development Cooperation with Ukraine
Tony Paso	HQ	Team Leader	MFA	Africa and the Middle East	Eastern Africa and Horn of Africa (ALi20)
Milma Kettunen	HQ	Communication Specialist	MFA	Communications	Communications on Sustainable Development and Trade
Janne Oksanen	HQ	Team leader	MFA	Africa and the Middle East	Eastern Africa
Petri Wanner	HQ	Senior Desk Officer	MFA	Africa and the Middle East	Southern Western Africa





NAME	ROLE	CURRENT POSITION	ORGANISATION	DEPARTMENT	UNIT
<b>Annika Kaipola</b>	Embassy	Senior Adviser, Development Policy	MFA	Development Policy	General Development Policy
<b>Arto Valjas</b>	Embassy	Senior Adviser, Development Policy	MFA	Africa and the Middle East	Management of the Department for Africa and the Middle East
<b>Tina Byring-Ilboudo</b>	Embassy	Consultant, Contract	NIRAS	none	none
<b>Gezahegn Tadesse</b>	Embassy	Special Adviser	MFA	none	Embassy of Finland, Addis Abeba
<b>Sami Frestadius</b>	Embassy	Senior Specialist Development Policy, Natural Resources	MFA	none	Embassy of Finland, Addis Abeba
<b>Workaferahu Eshetu</b>	Embassy	Special Adviser	MFA	none	Embassy of Finland, Addis Abeba
<b>Hanna Laitinen</b>	Embassy	Counsellor, Education and Social Development in Somalia	MFA	none	Embassy of Finland, Nairobi
<b>Sari Uusi-Rauva</b>	Embassy	Deputy Head of Mission	MFA	none	Embassy of Finland, Dar es Salaam
<b>Jan Koivu</b>	Embassy	Senior Specialist	MFA	none	Embassy of Finland, Lusaka
<b>Marko Saarinen</b>	Implementer/ Embassy	Senior Specialist Development Policy	MFA	none	Embassy of Finland, Pretoria
<b>Jaakko Jakkila</b>	Implementer/ HQ/ Embassy	Senior Adviser, Development Policy	MFA	Africa and the Middle East	Management of the Department for Africa and the Middle East

N.B. Titles and positions reflect the situation that prevailed at the time of the interviews in 2021.



## Annex 3: Documents Consulted

- EU Commission. (2015). *ROM Handbook: Results Oriented Monitoring*. Brussels, Belgium.
- Independent Evaluation Group-World Bank. (2007). *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards*. Washington, D.C., USA.
- Development Evaluation Unit. (2018) *Evaluation Manual*. Ministry for Foreign Affairs of Finland. [https://um.fi/documents/384998/o/Evaluation+Manual\\_2021+%281%29.pdf/6fe6d6a9-86e2-2f5b-4d35-fa232582a703?t=1617079302186](https://um.fi/documents/384998/o/Evaluation+Manual_2021+%281%29.pdf/6fe6d6a9-86e2-2f5b-4d35-fa232582a703?t=1617079302186)
- MFA. (n.d.) *Theme 11 - Templates, tools and additional reading*. Ministry for Foreign Affairs of Finland. [https://www.eoppiva.fi/kurssit/evaluation-manual-2/#/lessons/gWmg0J46CkNwU4y\\_kptp\\_haJMZrvc4HO](https://www.eoppiva.fi/kurssit/evaluation-manual-2/#/lessons/gWmg0J46CkNwU4y_kptp_haJMZrvc4HO)
- MFA. (2021). *Meta-evaluoinnin 2015-2017 johdon päätöksen toimeenpanon seurantaraportti*. Ministry for Foreign Affairs of Finland.
- NORAD. (2017). *The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation (01)*. Oslo, Norway.
- OECD. (2010). *Quality Standards for Development Evaluation*. DAC Guidelines and Reference Series. Paris, France.
- United Nations Evaluation Group. (2016). *Norms and Standards for Evaluation*. New York, USA
- UN Women. (2017). *What can we learn from UN-Women Evaluations? A meta-analysis of evaluations managed by UN-Women in 2016*. (UNW/2017/CRP.10). New York, USA.



# Annex 4: Analysis Grid

EVALUATION QUESTION	DATA SOURCES USED	DATA ANALYSIS METHOD
<b>FOR THE METAEVALUATION:</b>		
1. How can MFA's decentralised evaluation portfolio be described (e.g., by commissioner, budgets, sector and regional distribution, evaluation type)?	Decentralised evaluation reports	Descriptive statistics
2. How is the quality and reliability of decentralised evaluation reports? Is there a difference between the quality of MFA-commissioned evaluation reports and the quality of evaluation reports that are commissioned by MFA's partners? Are there other differences between relevant sub-groups (e.g., final vs. mid-term evaluations, individual/independent evaluator(s) vs. consulting firm/institutes) if applicable?	Decentralised evaluation reports	Quality assessment tool, Descriptive statistics, OLS regressions
3. How is the quality of the ToRs of decentralised evaluations? Are there systematic patterns regarding the quality of ToRs and the quality of corresponding evaluation reports?	ToRs, Decentralised evaluation reports	Quality assessment tool, Descriptive statistics OLS regressions
4. What are key differences and possible trends compared to the previous metaevaluation? Are there gaps regarding MFA's evaluation capacity?	Decentralised evaluation reports, Previous metaevaluation, MFA evaluation manual, MFA staff	Descriptive statistics, Qualitative content analysis
<b>FOR THE SUMMATIVE META-ANALYSIS:</b>		
5. What can be said about the quality of Finnish development cooperation according to the OECD DAC criteria based on reliable decentralised evaluation reports (i.e., relevance, coherence, effectiveness, efficiency, impact, sustainability)?	Decentralised evaluation reports (Implementers)	Content assessment tool, Qualitative content analysis Descriptive statistics
6. What can be said about the consideration of cross-cutting objectives, the human rights-based approach (HRBA) and policy priority areas in Finnish development cooperation based on reliable decentralised evaluation reports?	Decentralised evaluation reports (Implementers)	Quantitative key word analysis Content assessment tool Qualitative content analysis Descriptive statistics
7. What are the major recommendations and lessons learnt to improve Finnish development cooperation emerging from reliable decentralised evaluation reports?	Decentralised evaluation reports (Implementers)	Content assessment tool Qualitative content analysis Descriptive statistics



EVALUATION QUESTION	DATA SOURCES USED	DATA ANALYSIS METHOD
8. What can be said about the overall quality of Finnish development cooperation based on reliable decentralised evaluation reports? What are major strengths and challenges emerging from reliable decentralised evaluation reports? Are there typical factors for success or challenges?	Decentralised evaluation reports (Implementers)	Content assessment tool Descriptive statistics Qualitative content analysis
9. What are key differences and possible trends compared to the overall results of the previous metaevaluation?	Decentralised evaluation reports (Implementers) Previous metaevaluation	Descriptive statistics, Qualitative content analysis
<b>FOR THE USE AND USEFULNESS ASSESSMENT</b>		
10. How are decentralised evaluations reports used by key stakeholders (i.e., MFA commissioners, embassies, implementers)?	MFA staff, Implementers	Descriptive statistics, Qualitative content analysis
11. How useful are decentralised evaluations (i.e., reports and processes) assessed by key stakeholders? To what extent are decentralised evaluations timely? To what extent provide decentralised evaluations relevant and realistic recommendations? To what extent have recommendations been implemented?	MFA staff, Implementers	Descriptive statistics, Qualitative content analysis
12. Are there typical factors facilitating or hampering the usefulness of decentralised evaluations (e.g., report quality, timing)? What are major recommendations and lessons learnt to improve the use and usefulness of decentralised evaluations based on the perspectives of key stakeholders?	MFA staff, Implementers	Descriptive statistics, Qualitative content analysis
<b>OVERALL RECOMMENDATIONS FROM THE ASSIGNMENT</b>		
13. What are recommendations to improve the quality of MFA's decentralised evaluations reports (and processes if applicable)?	Findings of the quality assessment	Expert judgement
14. What are recommendations to improve the quality of Finnish development cooperation?	Findings of the content assessment	Expert judgement
15. What are recommendations to improve the use and the usefulness of decentralised evaluations?	Findings of the use and usefulness assessment	Expert judgement
16. What are recommendations to improve the methodology of metaevaluations in the future?	Findings of the assignment	Expert judgement



# Annex 5: Methodological details

## Development, modifications, and application of the assessment tools

Assessment tools were **developed as part of the previous metaevaluation in 2017** to prepare the ground for replication. They were based on reviews of MFA documents, our experience with metaevaluations and systematic reviews, and similar assignments conducted by others. They aim at (i) establishing a robust toolkit for MFA to evaluate the quality of its decentralised evaluations, (ii) providing reliable insights on the interventions for accountability purposes, and (iii) drawing emerging issues of projects and programmes from the evaluation reports' point of view.

The **structure of the quality assessment tool follows** the chapters of the evaluation report according to the **MFA evaluation manual** (MFA of Finland, 2013), i.e., (i) introduction, (ii) methodology, (iii) context and intervention logic, (iv) findings, (v) conclusions, (vi) recommendations, and (vii) annexes. However, the metaevaluation team considered the information regardless of where the evaluators placed it. (viii) Cross-cutting objectives and (ix) general issues were assessed after that, and checking (x) the summary marked the end of the assessment process. This structure facilitated easy application for the metaevaluation team.

A review of the original quality assessment tool against the recent quality standards for evaluation by OECD DAC confirmed that the most important aspects are still covered. **Only additions** regarding the new **OECD DAC coherence criterion** were necessary, and few new aspects on the old OECD DAC criteria were added as extra features. A review against the latest version of the MFA evaluation manual (MFA, 2021b) further disclosed the **need to adjust** assessments regarding **cross-cutting objectives** and **add** checks on whether referral to the **policy priority areas** is provided (MFA, 2020c). Finally, reflecting increased evaluation quality standards leads to slightly stronger measures on the quality of evidence provided by the evaluators.

The **assessment of the underlying ToR** is inevitable to ensure a fair quality assessment because reports may not provide information on aspects as they were explicitly excluded or not requested by the corresponding ToR. Further ToR assessment allows reviewing compliance of ToR with MFA guidelines. Consequently, ToR assessment was already in the predecessor metaevaluation regarded as part of the quality assessment. A check against the recent MFA manual did **not require any modifications**. Thus, the application of the tool was fully replicated in this assignment. It follows the same logic of aspects, sub-sections and sections as presented above and covers the following sections: (i) intervention, (ii) purpose, objectives, and scope of the evaluation, (iii) evaluation questions, (iv) evaluation criteria, (v) methodology, (vi) feasibility, (vii) evaluation process and quality assurance, and (viii) overarching and cross-cutting objectives.

To recall the principal **functionality of the tool**, we cite from the last metaevaluation (Silvestrini et al., 2017) report: “The main sections consist of sub-sections with very specific statements, so-called aspects, which were checked in terms of true or false—for example, the first section *1. Introduction and background* contain the sub-section *1.1 Rationale and purpose*. Within this section, there are two statements that the meta-evaluator has assessed. For example, one of these statements is: *1.1a Report describes the purpose of evaluation*. The metaevaluation team checked if the original evaluator had described the purpose of the evaluation in the evaluation report and selected one of the answer options, “yes” or “no”. Most statements in the quality assessment tool could be answered with “yes” or “no” because many aspects refer to checking for the existence of



certain information in the report. Still, in several cases, there are more answer options (on a four-step scale) that refer to different grades of completeness or more specific assessments introduced by the evaluator. Table 7 provides an example to assess whether the sources of information are described.”

**Table 7: Example: four-step scale question**

ASPECT	ANSWER OPTIONS	GUIDANCE FOR CHOOSING THE ANSWER
<p><b>2.2a The sources of information are described.</b></p>	<p>(1) no, (2) short and incomplete, (3) short and complete, (4) detailed and complete</p>	<p>(1) no information (2) cryptic, incomplete, not naming types of documents or different groups to be interviewed etc., (3) short but naming all sources of information, (4) minimum one paragraph with three or more sentences with all sources of information</p>

Source: previous metaevaluation

The original **content assessment tool** was **built on the MFA Evaluation Manual** (MFA, 2013), the **MFA Manual for Bilateral Cooperation** (2012) (MFA, 2012a) and the **OECD DAC criteria** (OECD DAC, 1991). A review against the revised evaluation criteria and the recent publication on their thoughtful application (OECD, 2021) disclosed at large still an acceptable fit, despite the **necessity to integrate the new coherence criterion** and, in turn, to remove the section on aid effectiveness and triple C (coherence, complementarity, coordination). Although OECD (2021) suggests several adjustments and integrating new aspects into the old criteria (i.e., relevance, effectiveness, efficiency, impact, and sustainability), the metaevaluation team decided with EVA-11 to stick to major elements of the content assessment tool. The benefit of **allowing comparability with predecessor metaevaluation** results was thus expected to **outweigh perfect alignment** with recent developments **at international level**.

A review against MFA’s 2018 Manual for Bilateral Cooperation (MFA, 2018c), the Guideline for the Cross-cutting Objectives in Finnish Development Policy and Cooperation (MFA, 2020a), and the Theories of Change and Aggregate Indicators for Finland’s Development Policy (MFA, 2020b), however, revealed additional **modification needs regarding cross-cutting objectives** and the new **integration of the policy priority areas** of Finland.

The **content assessment tool** consists of two main elements: The first part focuses on transferring evaluators’ general assessments on the six OECD DAC criteria and cross-cutting objectives (if available) and then extracting their assessments on single sub-criteria/topics and single aspects. In this regard, the tool is conceptually similar to the quality assessment tool by transferring assessments to a four-step scale and applying aggregation of single aspects, sub-criteria/topics and criteria/topics.

Both assessment tools **ensure suitability** to a high degree of **content-related and methodological heterogeneity** among the sample of decentralised evaluation reports (i.e. (i) varying context conditions, e.g. poverty levels, degree of political stability, etc. in the countries under consideration, (ii) differences among implementing partner organisations, e.g. level of operations, financial resources etc., (iii) different thematic focuses, (iv) varying working approaches, e.g.



technical, human rights-based, (v) differing intervention budgets, (vi) different scope and scale of the evaluations, e.g. mid-term vs final, programme vs project evaluation etc., (vii) different evaluation designs with accordingly varying data sources and analysis methods used, e.g. contribution analyses, ex-post-facto designs, and (viii) varying evaluation budgets).

In the second part of the content assessment, decentralised evaluation reports were **coded with MAXQDA**. By doing so, **underlying positive and negative reasons for key assessments** provided were captured. Further **thematic coding** was applied to allocate **lessons learnt and recommendations** to specific fields<sup>4</sup> as already done by the original content assessment of the predecessor metaevaluation and providing a better understanding of success and failure. Axial coding was applied whenever a lesson learnt or a recommendation did not fit any category (i.e., the identification of new categories throughout later analysis steps was encouraged).

As in the previous metaevaluation, we only considered lessons that were either formulated according to the OECD DAC definition (OECD, 2010): “*Generalisations based on evaluation experiences with projects, programs, or policies that abstract from the specific circumstances to broader situations*” or lessons formulated in a way that generalisation with a reasonable level of expert judgement was possible. So-called lessons that required arbitrary interpretation to go beyond intervention-level recommendations were not considered.

The same analysis steps as in the previous metaevaluation were applied. Thus, we cite accordingly: “To provide a meaningful synthesis of rather heterogeneous lessons learnt and recommendations made by various evaluators in different reports, a three-step approach has been utilised. In a first step, the lessons or recommendations found in the evaluation reports were broadly assigned to categories corresponding to the main thematic interests of the metaevaluation. This allowed identifying first tendencies with regard to the frequency of certain topics. Subsequently, in a second step, the lessons or recommendations within each broader category were generalised and clustered to the extent possible. Finally, in a third step their overall frequency and their importance based on the expert judgement of the metaevaluation team were assessed. Recommendations appearing in more than 50% of the reports and lessons learnt [appearing in more than 10% of the reports] were synthesised and generalised further; they form the main part of the synthesis. Less frequent lessons or recommendations were treated anecdotally and added as illustrating examples when perceived as relevant.”

## Aggregation and further analysis

**Data processing** lays the foundation for metaevaluations and is analytically **not clearly distinguishable from data analysis**. Thus, the first step of the analysis process has already been presented above. This section briefly recalls the way data was aggregated and further analysed.

First, **grades of single aspects were calculated for sub-sections/criteria/topics**. **Weights** were introduced to balance single aspects according **to their importance**. The default weight was set as “1” and was increased for comparable important aspects. Taking weights into account, **arithmetic means** were calculated at **different levels**. **Weights** of the predecessor metaevaluation were kept constant. Further details are provided in annexes 6 and 7.

---

<sup>4</sup> I.e. financial aspects, personnel, time, capacity, equipment, management, communication, scope, participation, outreach, M&E, relevance, coherence, effectiveness, efficiency, impact, sustainability.



The **analysis centred around sub-section/criteria/topics and section/criteria/topics level**. Single aspects were only considered to elaborate on results. Further **overall scores** were used to provide a general picture of the decentralised evaluation reports' and ToRs' quality as well as on overall quality with respect to Finnish development cooperation. The predecessor metaevaluation excluded cross-cutting objectives, aid effectiveness and triple C due to limited data available from the overall scores. To allow **comparability**, we provided overall scores following that reasoning in **static analysis** and, by doing so, excluded the new coherence criterion and newly introduced additional aspects on the old criteria. In addition, we calculated overall scores for **dynamic analysis, including the coherence criterion**. As in the predecessor metaevaluation, scores of single OECD DAC criteria assessed by the evaluator were generally aggregated and divided through the sum of OECD DAC criteria to be assessed as requested by the ToR (i.e., evaluations which per ToR were not requested to assess all OECD DAC criteria were not rated on missing criteria).

**Overall scores** were taken to run ordinary least squares (**OLS**) regressions and **ordered logistic regressions** with robust standard errors to identify determinants of the quality of evaluation findings and Finnish development cooperation. We ran regression analyses with different sets of independent variables like MFA commissioned reports vs partner reports, individual/independent consultant(s) vs consulting firms/institutes, mid-term vs final evaluation, national vs regional/global interventions, different regions, different sectors, different evaluation budgets). **Spearman's correlation coefficients** were employed to analyse potential linkages between two variables, **Mann-Whitney test statistics** to check for significant differences between two groups, and **Kruskal-Wallis test statistics** to detect differences among several groups. Analyses were conducted with the statistical software package R.

Further, **thematic and axial codes** underwent **qualitative content analysis** identifying influencing factors and general trends to derive systematic lessons from and recommendations for Finnish development cooperation. **Replication** of the predecessor aggregation and analysis methods were thereby applied.

As an additional feature, MFA's cross-cutting objectives were assessed by **quantitative keyword analysis** with MAXQDA. At first, the documents were grouped into evaluations commissioned by MFA and those commissioned by other organisations than MFA. Corresponding codes were created, including only the findings, conclusions, and recommendations sections in the analysis. In this way, the bias that could be created by including introductory paragraphs of the reports and attachments like the ToRs was minimised.

A dictionary of keywords and close derivatives was created by collecting relevant terminology from two MFA documents; the Guideline for Cross-Cutting Objectives in the Finnish Development Policy and Cooperation (MFA, 2020a) and Theories of Change and Aggregate Indicators for Finland's Development Policy (MFA, 2020b) (see Table 9 in Annex 13).

**MAXQDA word frequency's function** was used for generating the results of the analyses. Further data processing took place in Excel. In subsequent steps, the **keyword in context function** was applied to understand the contexts of the terms used in the evaluation reports.

In addition, **thematic codes of main gender results were classified** into four categories: "general comments on results", "specific gender results", "lessons learnt and good practices", and "challenges", thus delving deeper into detail concerning the most prominent cross-cutting objective of Finnish development cooperation.





# Annex 6: Quality Assessment Tool for Reports and ToR

name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
keyinfo	Key information									
ProjectID	Evaluation ID									
ProjectTitle	Evaluation name									
ProjectYear	Year									
ToRavailable	ToR available									
ProjectType	Type of Evaluation									
EvaluatorID	Evaluator ID									
review	Selected for review									
ReviewerID	Reviewer ID									
13	Introduction and Context	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(1.1+1.2+1.3+1.4+1.5 + 3.1*2)/7$	2	22	45	11		80	
1	Introduction and background	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(1.1+1.2+1.3+1.4+1.5)/5$	2	12	36	30		80	
1.1	Rationale and purpose	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(1.1a*2+1.1b)/3$	11	1	28	40		80	
1.1a	Report describes purpose for evaluation.	no (1), yes (4)	general statement on rational/purpose	12			68		80	
1.1b	Report describes intended user(s) of evaluation.	no (1), yes (4)	Organizations/divisions/persons are described that will use the results of the evaluation.	39			41		80	
1.2	Objectives of the evaluation: Report describes objectives of evaluation.	no (1), yes (4)	statement on objectives	14			66		80	
1.3	Evaluation object	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(1.3a+1.3b+1.3c+1.3d+1.3e+1.3f+1.3g+1.3h +1.3i)/9$	2	19	33	26		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
1.3a	The description of the intervention includes time period.	no (1), yes (4)	Start AND end of intervention	10			70		80	
1.3b	The description of the intervention includes budget.	no (1), yes (4)		25			55		80	
1.3c	The description of the intervention includes intervention area.	no (1), yes (4)	Description where exactly the intervention takes places in the country/region.	31			49		80	
1.3d	The description of the intervention includes components of the intervention.	no (1), yes (4)	Different components of the intervention are described	15			65		80	
1.3e	The description of the intervention includes target groups.	no (1), yes (4)	Who is going to benefit from the intervention?	23			57		80	
1.3f	The description of the intervention includes objectives of the intervention.	no (1), yes (4)		11			69		80	
1.3g	The description of the intervention includes stakeholders.	no (1), yes (4)	(4) different stakeholder groups are mentioned e.g. (N)Go's, implementers, external experts, (secondary) beneficiaries	30			50		80	
1.3h	The description of the intervention includes implementation arrangements (incl. organisational set-up).	no (1), yes (4)	(4) Which partners are involved in the project/program? What is their labour division? With whom was the project negotiated?	43			37		80	
1.3i	The description of the intervention includes changes regarding implementation.	no (1), yes (4)		58			22		80	
1.4	Scope of evaluation	no (1), yes (4)	1.4a	29			51		80	
1.4a	The scope of the evaluation is described.	no (1), yes (4)	What is evaluated? Time, area, components	29			51		80	
1.4b	The scope is coherent with ToR, otherwise justification is given.	no w/o justification (1), no w/ justification or yes (4), no ToR available, n.T., n.a.	In case of large differences ask MFA for IR.	4			50	26	80	54
1.5	Evaluation questions are reported.	no eq reported (1), few eq are reported (2), more than half of eq or the main eq are reported (3), all eq are reported (4)	(2) only few eq are reported, the selection seems arbitrary, (3) given a different priorities, the main eq e.g. heading eqs are reported or at least half of the eq are reported, also in annex ok with reference.	21	1	3	55		80	
1.6	Results of previous evaluations are mentioned.	no (1), yes (4)		59			21		80	
2.	Methodology	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(2.1+2.2+2.3+2.4+2.5+2.6)/6		45	35			80	
2.1	Evaluation design	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(2.1a+2.1b)/2	36		34	10		80	



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
2.1a	The general evaluation approach is described.	no (1), yes (4)	participatory, theory-based, formative, exploratory, empowerment etc. mixed methods	37			43		80	
2.1b	The evaluation design is described.	no (1), yes (4)	A design is developed. I.e. is there a strategy on how to answer the evaluation questions e.g. pre-post design, comparison groups, contribution analysis, etc.	69			11		80	
2.2	Sources of evidence	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(2.2a*2+2.2b+2.2c+2.2d+2.2e+2.2f+2.2g+2.2h+2.2i*2+7.3a + 7.3b/13$		5	50	25		80	
2.2a	The sources of information are described.	no (1), short and incomplete (2), short and complete (3), detailed and complete (4)	(2) one sentence, cryptic, incomplete, not naming types of documents or different groups to be interviewed etc., (3) at least two sentences and naming all sources of information, (4) minimum one paragraph with three or more sentences with all sources of information	5	15	12	48		80	
2.2b	Project documents have been used in the evaluation.	no (1), yes (4)		1			79		80	
2.2c	M&E data has been used in the evaluation.	no (1), yes (4)		32			48		80	
2.2d	Additional literature has been used in the evaluation.	no (1), yes (4)		46			34		80	
2.2e	The implementing organisation(s) has/have been used as source of information for the evaluation.	no (1), yes (4)		3			77		80	
2.2f	The beneficiaries have been used as source of information of the evaluation.	no (1), yes (4)		9			71		80	
2.2g	The institutional environment e.g. external experts, (N) GOs have been used as source of information in the evaluation.	no (1), yes (4)		26			54		80	
2.2h	Other source(s) of information has/have been used	no (1), yes -specify- (4)		69			11		80	
2.2h.specify	specify:	free input						80	80	0
2.2i	The mix of sources of information is appropriate (data triangulation).	completely inappropriate (1), rather inappropriate (2), rather appropriate (3), completely appropriate (4)	(1) only secondary data or only one source, (2) two sources, (3) three sources, (4) three or more source with mixture of primary and secondary data.	1	2	11	66		80	
2.3	Data collection	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(2.3a*2+2.3b+2.3c+2.3d+2.3e+2.3f+2.3g*2+2.3h*2+2.3i*2+2.3j*2+7.3f)/16$		38	41	1		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
2.3a	Data collection techniques are described in the report.	no (1), short and incomplete (2), short and complete (3), detailed and complete (4)	(2) one sentence, cryptic, incomplete, not naming techniques etc., (3) at least two sentences and naming all techniques, (4) minimum one paragraph with three or more sentences with all techniques	3	14	32	31		80	
2.3b	Interviews have been conducted in the evaluation.	no (1), yes (4)	If not method section, indications from findings can be considered				80		80	
2.3c	Focus group discussions have been conducted.	no (1), yes (4)	If not method section, indications from findings can be considered	31			49		80	
2.3d	Participatory observation has been conducted.	no (1), yes (4)	If not method section, indications from findings can be considered	58			22		80	
2.3e	A survey(s) has been conducted.	no (1), yes (4)	If not method section, indications from findings can be considered	49			31		80	
2.3f	Other data collection method(s) has/have been used	no (1), yes -specify- (4)	e.g. video diaries, experiments	73			7		80	
2.3f.specify	specify:	free input					1	79	80	1
2.3g	A mix of data collection techniques is applied.	no (1), yes (4)	(1) only one, (4) two or more	9			71		80	
2.3h	Data collection techniques are applied without severe failures.	no (1), yes (4)	(1) e.g. extreme size of focus group discussions, survey population size smaller than 50	3			77		80	
2.3i	Validity of data is assessed by the evaluators.	no (1), yes (4)	There is a paragraph discussing the validity. Measure the instruments what they want to measure? Discussion of internal vs. external validity.	74			6		80	
2.3j	Reliability of data is assessed by the evaluators.	no (1), yes (4)	There is a paragraph discussing the reliability of data e. g. would a repetition of the study yield the same results?	79			1		80	
2.4	Sampling	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(2.4a*2+2.4b*2+2.4c)/5	36	18	11	15		80	
2.4a	The sample is described.	no (1), brief and incomplete (2), moderate but incomplete (3), complete (4)	(1) no information at all, (2) very incomplete information (e.g. total number of persons involved), (3) incomplete information (e.g. number of persons involved and affiliations but information not connected to the data collection instruments), (4) detailed information (number of persons and affiliation for each data collection technique are provided)	22	24	20	14		80	



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
2.4b	The sampling strategy is described.	no (1), yes (4)	Methods or criteria to select the persons from whom to collect data are described.	51			29		80	
2.4c	The evaluators justify the sampling strategy.	no (1), yes (4)	Reasons for the sampling strategy are described.	61			19		80	
2.4d	Data collection acknowledges all groups of key stakeholders.	no (1), yes (4), n.a.	Compare purpose and sampling strategy. Are groups involved who are key stakeholders given the purpose of the evaluation?	11			59	10	80	70
2.5	Data analysis methods	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(2.5a+2.5b+2.5c)/3	3	28	36	13		80	
2.5a	Data analysis methods are described.	no (1), brief and incomplete (2), moderate but incomplete (3), complete (4)	(1) no information at all, (2) very incomplete information (for few data the data analysis method is described), (3) incomplete information (for most data the data analysis method is described), (4) detailed information (for each data the data analysis method is described)	42	23	12	3		80	
2.5b	The mix of data analysis methods is appropriate (triangulation of methods).	no (1), yes (4)	Qualitative and quantitative analysis methods are used e.g. content analysis, grounded theory, summary statistics, correlations, cross tabulations. Focus on mixture of qualitative analysis and quantitative analysis (tables with figures). This does not mean primary quantitative data has to be collected, but at least secondary data like project documents have to be analysed quantitatively.	33			47		80	
2.5c	Data analysis methods are applied without severe failures.	no (1), yes (4)	e.g. ignoring basic statistics like mixing up pure numbers and causal effects, generalizing based on single interviews etc.	3			77		80	
2.6	Limitations and challenges	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(2.6a*2+2.6b+2.6c+2.6d*2+2.6e)/7	20	35	19	6		80	
2.6a	Limitations regarding data collection are described.	no (1), yes (4)		21			59		80	
2.6b	Limitations regarding the evaluation process are described.	no (1), yes (4)		38			42		80	
2.6c	Limitations regarding data analysis methods are described.	no (1), yes (4)		71			9		80	



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
2.6d	Possible influence of limitations on the evaluation is discussed.	no (1), yes (4)		67			13		80	
2.6e	Coping strategies for limitations are described.	no (1), yes (4)		60			20		80	
3.	Context and intervention logic		No aggregation: Context combined with introduction section, intervention logic integrated in findings							
3.1	Context	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(3.1b+3.1c+3.1d+3.1e+3.1f+3.1g+3.1h+3.1i+3.1j)*2/10$	1	12	42	5	20	80	
3.1a	A context analysis is provided in the report.	no (1), yes (4)		20			60		80	
3.1b	In the context analysis it is referred to (inter)national key actors in the sector.	no (1), yes (4), n.a.		11			49	20	80	
3.1c	In the context analysis it is referred to international policies or strategies.	no (1), yes (4), n.a.		29			31	20	80	
3.1d	In the context analysis it is referred to Finnish development policies or strategies.	no (1), yes (4), n.a.		39			21	20	80	
3.1e	In the context analysis it is referred to national/regional policies (e.g. sector strategies, poverty reduction policies).	no (1), yes (4), n.a.		17			43	20	80	
3.1f	In the context analysis it is referred to the country/ regional context (socio-economic, political, cultural factors if applicable).	no (1), yes (4), n.a.		15			45	20	80	
3.1g	In the context analysis it is referred to gender equality.	no (1), yes (4)	rights of women and girls	55			25		80	
3.1h	In the context analysis it is referred to non-discrimination.	no (1), yes (4)	equal opportunities to participate, rights of the most vulnerable, reduction of inequality	59			21		80	
3.1i	In the context analysis it is referred to climate sustainability.	no (1), yes (4)	climate change preparedness and mitigation, climate resilience/low emission development	63			17		80	
3.1j	Overall, the context description is in relation with intervention.	no (1), rather no (2), rather yes (3), yes (4), n.a.	(1) not at all in relation, (2) few parts in relation, (3) most parts in relation, (4) all parts in relation (direct reference important)	1	3	13	43	20	80	
3.2	Intervention logic	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(3.2a*2+3.2b+3.2c*2+3.2d)/6$	36	15	16	13		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
3.2a	The intervention logic (IL), logical framework (LF), programme theory (PT) or the theory of change (ToC) is described.	no (1), brief and incomplete (2), moderate (3), complete and comprehensive (4), n.a.	(1) not at all, (2) one-two sentences, rather cryptic, incomplete (3) paragraph or table, giving an idea but program does not become fully clear or table is not described in the text, (4) minimum one paragraph with three sentences and very comprehensive table with explanation or very detailed description without table, logic of the programme becomes clear, overall comprehensive and easy to understand, (n.a.) if evaluators mentions the lack of an (appropriate) framework	31	17	16	16		80	
3.2b	A results model (IOOI) is provided.	no (1), yes (4)	Input, expected output, outcome and impact are in the report.	57			23		80	
3.2c	The IL, LF, PT, ToC or the (IOOI) is assessed by the evaluator as appropriate, otherwise shortcomings are disclosed.	no (1), yes (4), n.a.		53			27		80	
3.2d	Underlying assumptions of the intervention logic are reviewed by evaluator.	no (1), yes (4), n.a.		59			21		80	
4.	Findings	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(4.1*2+4.2+3.2+4.3dac)/5$	3	56	20	1		80	
4.1	Findings	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(4.1a+4.1b+4.1c+4.1d)/4$	1	21	46	12		80	
4.1a	Findings are evidence-based.	no (1), yes (4)	General Assessment. More differentiated assessment below. The findings refer clearly to the data collected.	7			73		80	
4.1d	All findings are evidence-based.	no (1), rather no (2), rather yes (3), yes (4)	(1) not at all, (2) findings are rarely based on evidence or sources of evidence are rarely specified, (3) findings are based on evidence but the sources of evidence are only specified for longer passages in the text and not for each statement, (4) almost scientific level; all findings are evidence-based and for each statement the sources of evidence is specified.	1	38	37	4		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
4.1b	Results are put into perspective with referral to different data sources.	no (1), rather no (2), rather yes (3), yes (4)	the focus here is on different sources = triangulation. (1) not put at all into perspective, (2) very rarely put into perspective e.g. only two, three times within the report, (3) often parts put into perspective e.g. around half of the results, (4) vast majority put into perspective (e.g. interviews showed xx but the focus groups came to different results. Or in the survey respondents showed xx which was confirmed by the interviews.)	15	39	19	7		80	
4.1c	Only findings are presented. (No conclusions, no recommendations)	no (1), yes (4)		46			34		80	
4.2	Causal Inference	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(4.2a+4.2b)/2	61		16	3		80	
4.2a	Attribution of intervention to results is discussed.	no (1), yes (4)	Evaluators critically discuss the ability of the intervention to attribute to the results.	61			19		80	
4.2b	Confounding factors are discussed.	no (1), yes (4)		77			3		80	
4.3dac	DAC Criteria		(4.3+4.4+4.5+4.6+4.7+4.8)/6	6	48	25	1		80	
4.3	Relevance	Is the intervention doing the right things? The extent to which the intervention's objectives and design respond to beneficiaries', global, country and partner/ institution needs, policies and priorities, and continue to do so if circumstances change.	4.3b	14	21	36	9		80	
4.3a	Relevance is discussed.	no (1), yes (4)	RELEVANCE IS ALWAYS LINKED TO THE INTERVENTION	10			70		80	
4.3b	Relevance is appropriately captured.	no (1), rather no (2), rather yes (3), yes (4)	READ THE SECTION AND RATE 4.3c-k, AFTERWARDS ASSESS SECTION IN GENERAL CONSIDERING THESE ASSESSMENTS.	14	21	36	9		80	
4.3c	Does the report discuss, if the intervention meets the needs of the target group?	no (1), yes (4), n.a.	n.a. if there is no target group (i.e. only final beneficiaries), NOTE GOOD PRACTICE	20			48	12	80	68





name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
4.3d	Does the report discuss, if the intervention meets the needs of the final beneficiaries (population)?	no (1), yes (4)	NOTE GOOD PRACTICE	28			52		80	
4.3e	Does the report discuss, if the intervention is consistent and supportive of the partner government/regional policies?	no (1), yes (4)	NOTE GOOD PRACTICE	25			55		80	
4.3f	Does the report discuss, if the intervention is consistent with the MFA development cooperation policy?	no (1), yes (4)	NOTE GOOD PRACTICE	52			28		80	
4.3g	Does the report discuss, if the intervention is addressing international conventions, policies, strategies or goals?	no (1), yes (4)	NOTE GOOD PRACTICE	44			36		80	
4.3h	Does the report discuss, if the intervention is sensitive to the context conditions?	no (1), yes (4)	NOTE GOOD PRACTICE	41			39		80	
4.3i	Does the report discuss, if the intervention is sensitiveness to the capacity conditions?	no (1), yes (4)	NOTE GOOD PRACTICE	66			14		80	
4.3j	Does the report discuss, if the intervention's design is appropriate?	no (1), yes (4)	NOTE GOOD PRACTICE	43			37		80	
4.3k	Does the report discuss, if the intervention has been adapted to context changes to remain relevant?	no (1), yes (4), n.a.	NOTE GOOD PRACTICE	57			18	5	80	75
4.8	Coherence	How well does the intervention fit? The compatibility of the intervention with other interventions in a country, sector or institution. The extent to which other interventions (particularly policies) support or undermine the intervention and vice versa.	4.8b	44	21	13	2		80	
4.8a	Coherence is discussed.	no (1), yes (4)		38			42		80	
4.8b	Coherence is appropriately captured.	no (1), rather no (2), rather yes (3), yes (4)	READ THE SECTION AND RATE 4.8c-f, AFTERWARDS ASSESS SECTION IN GENERAL CONSIDERING THESE ASSESSMENTS.	44	21	13	2		80	
4.8c	Does the report discuss, if the intervention possess internal coherence (i.e. is coherent with other intervention by MFA)?	no (1), yes (4)	NOTE GOOD PRACTICE	59			21		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
4.8d	Does the report discuss, if the intervention possess external coherence?	no (1), yes (4)	CALCULATE MANUALLY (4 in both, e and f = 4; 1 in both, e and f = 1; 4 and 1 in e and f = 4. (4.8e+4.8f)/2)	52			28		80	
4.8e	Does the report discuss, if the intervention is complementary to other donor's activities (i.e. adds value)?	no (1), yes (4)	NOTE GOOD PRACTICE	57			23		80	
4.8f	Does the report discuss, if the intervention is coordinated with other donor's activities (i.e. avoiding duplication)	no (1), yes (4)	NOTE GOOD PRACTICE	56			24		80	
4.4	Effectiveness	Is the intervention achieving its objectives? The extent to which the intervention achieved, or is expected to achieve, its objectives and its results, including any differential results across groups.	4.4b	7	31	33	9		80	
4.4a	Effectiveness is discussed.			4			76		80	
4.4b	Effectiveness is appropriately captured.	no (1), rather no (2), rather yes (3), yes (4)	READ THE SECTION AND RATE 4.4c- i AFTERWARDS ASSESS SECTION IN GENERAL CONSIDERING THESE ASSESSMENTS.	7	31	33	9		80	
4.4c	Does the report discuss, if the outputs of the intervention have been achieved?	no (1), yes (4)	NOTE GOOD PRACTICE	14	1		65		80	
4.4d	Does the report discuss, if the outcomes of the intervention have been achieved?	no (1), yes (4)	NOTE GOOD PRACTICE	23	1		56		80	
4.4e	Does the report discuss, if the intervention has resulted in benefits for the target group?	no (1), yes (4), n.a.	n.a. if there is no target group (i.e. only final beneficiaries), NOTE GOOD PRACTICE	12			56	12	80	68
4.4f	Does the report discuss, if the intervention has resulted in benefits for the final beneficiaries?	no (1), yes (4)	NOTE GOOD PRACTICE	24			56		80	
4.4g	Does the report discuss, if the results are different for men and women? (differentiate between men and women?)	no (1), yes (4)	NOTE GOOD PRACTICE	60			20		80	
4.4h	Does the report discuss, if results are differentiated across groups (beyond gender, e.g. age groups, different target groups)	no (1), yes (4)	NOTE GOOD PRACTICE	63			17		80	
4.4i	Does the report discuss, if the importance of results is varying?	no (1), yes (4), n.a.	NOTE GOOD PRACTICE. Statements like the one or the other result is more important than another one.	80					80	



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
4.5	Efficiency	How well are resources being used? The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.	4.5b	14	37	28	1		80	
4.5a	Efficiency is discussed.	no (1), yes (4)		10			70		80	
4.5b	Efficiency is appropriately captured.	no (1), rather no (2), rather yes (3), yes (4)	READ THE SECTION AND RATE 4.5c-g, AFTERWARDS ASSESS SECTION IN GENERAL CONSIDERING THESE ASSESSMENTS.	14	37	28	1		80	
4.5c	Does the report discuss, if the implementation of the intervention is/was on time?	no (1), yes (4)	NOTE GOOD PRACTICE	40			40		80	
4.5d	Does the report discuss, if the inputs have been converted into high quality outputs?	no (1), yes (4)	NOTE GOOD PRACTICE	36			44		80	
4.5h	Does the report discuss, if the inputs have been converted into high quality outcomes?	no (1), yes (4)	NOTE GOOD PRACTICE	59			21		80	
4.5e	Does the report discuss, if the intervention is efficient regarding costs?	no (1), yes (4)	NOTE GOOD PRACTICE	31			49		80	
4.5f	Does the report discuss, if the intervention is efficient regarding personnel?	no (1), yes (4)	NOTE GOOD PRACTICE	54			26		80	
4.5g	Does the report discuss, if the implementation management is efficient?	no (1), yes (4)	NOTE GOOD PRACTICE	31			49		80	
4.6	Impact	What differences does the intervention make? The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.	(4.6b*2+4.6c)/3	22	22	32	4		80	
4.6a	Impact is discussed.	no (1), yes (4)		18			62		80	
4.6b	Impact is appropriately captured.	no (1), rather no (2), rather yes (3), yes (4)	READ THE SECTION AND RATE 4.6c-h, AFTERWARDS ASSESS SECTION IN GENERAL CONSIDERING THESE ASSESSMENTS.	28	26	22	4		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
4.6c	Does the report discuss, if the intervention contributed to its overall objective, reach its intended impact?	no (1), yes (4)	NOTE GOOD PRACTICE	38			42		80	
4.6d	Does the report discuss, if the intervention has any unintended impacts?	no (1), yes (4)		65			15		80	
4.6e	Does the report discuss, if the intervention contributes to enhance the quality of life of the final beneficiaries?	no (1), yes (4)	NOTE GOOD PRACTICE	52			28		80	
4.6f	Does the report discuss, if the intervention contributes to enhance institutional quality (i.e. institutions/services in the partner country/region have been improved)?	no (1), yes (4), n.a.	n.a. if the intervention does not address the institutional level, NOTE GOOD PRACTICE	41			33	6	80	74
4.6g	Does the report discuss, if the intervention contributed to changes in the partner country's/region's policies/ to sector reforms?	no (1), yes (4)	NOTE GOOD PRACTICE	59			21		80	
4.6h	Does the report discuss differential impacts across groups (e.g. by gender, age groups, different target groups like persons with disabilities)	no (1), yes (4)	NOTE GOOD PRACTICE	65			15		80	
4.7	Sustainability	Will the benefits last? The extent to which the net benefits of the intervention continue or are likely to continue.	4.7b	14	40	22	4		80	
4.7a	Sustainability is discussed.	no (1), yes (4)		8			72		80	
4.7b	Sustainability is appropriately captured.	no (1), rather no (2), rather yes (3), yes (4)	READ THE SECTION AND RATE 4.7c-n, AFTERWARDS ASSESS SECTION IN GENERAL REFLECTING THESE ASSESSMENTS.	14	40	22	4		80	
4.7c	Does the report discuss the economic sustainability of the intervention?	no (1), yes (4)		46			34		80	
4.7d	Does the report discuss the social sustainability of the intervention?	no (1), yes (4)	e.g. intervention is accepted by population, approach useful for population etc.	54			26		80	
4.7e	Does the report discuss the environmental sustainability of the intervention?	no (1), yes (4), n.a.		57			16	7	80	73
4.7f	Does the report discuss the sustainability as a multifaceted concept?	no (1), yes (4)		62			18		80	
4.7g	Does the report discuss if the benefits of the intervention are likely to continue after the completion of the intervention? (i.e. Do the final beneficiaries further benefit after the intervention ends?)	no (1), yes (4)	NOTE GOOD PRACTICE	31			49		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
4.7h	Does the report discuss, if the target group/beneficiaries has the capacity to make the intervention sustainable?	no (1), yes (4)		37			43		80	
4.7i	Does the report discuss, if the target group/beneficiaries has the financial means to make the intervention sustainable?	no (1), yes (4)		44			36		80	
4.7j	Does the report discuss, if the implementing partner organisations / intermediaries have the institutional capacity to make the intervention sustainable?	no (1), yes (4)	often the same as target group, but can be different e.g. International NGO, Consulting etc. NOTE GOOD PRACTICE	41			39		80	
4.7k	Does the report discuss, if the implementing partner organisations / intermediaries have the financial means to make the intervention sustainable?	no (1), yes (4)	often the same as target group, but can be different e.g. International NGO, Consulting etc. NOTE GOOD PRACTICE	48			32		80	
4.7l	Does the report discuss, if risks threatening the sustainability of the intervention?	no (1), yes (4)	NOTE GOOD PRACTICE	44			36		80	
4.7m	Does the report discuss, if trade-offs exist regarding the sustainability of the intervention?	no (1), yes (4)	NOTE GOOD PRACTICE	76			4		80	
4.7n	Does the report discuss, if the enabling environment allows sustainability?	no (1), yes (4)	NOTE GOOD PRACTICE	56			24		80	
56.	Conclusions and Recommendations		$(5.a+6.) / 2$	7	8	44	21		80	
5.	Conclusions	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(5.a*4+5.b+5.c+5.d+5.e+5.f) / 10$	13	5	25	37		80	
5.a	Conclusions are derived from findings.	no (1), yes (4)	not necessarily direct reference but perceived as consistent with findings.	15			65		80	
5.b	Relevance is discussed.	no (1), yes (4)		31			49		80	
5.c	Effectiveness is discussed.	no (1), yes (4)		24			56		80	
5.d	Efficiency is discussed.	no (1), yes (4)		37			43		80	
5.e	Impact is discussed.	no (1), yes (4)		40			40		80	
5.f	Sustainability is discussed.	no (1), yes (4)		31			49		80	
6.	Recommendations	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(6.a*2+6.b+6.c+6.d+6.e+6.f) / 7$	5	39	25	11		80	
6.a	Recommendations are derived from findings and conclusions.	no (1), rather no (2), rather yes (3), yes (4)	(1) no logical conjunction to conclusions, (2) very rarely logical conjunction to conclusions e.g. only two, three times, (3) often logical conjunction to conclusions e.g. around half, (4) for vast majority logical conjunction to conclusions or findings	3	14	17	46		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
6.b	Recommendations are directed to actors.	no (1), yes (4)	majority of recommendations is clearly directed to actors	25			55		80	
6.c	Recommendations are prioritised.	no (1), yes (4)		71			9		80	
6.d	Recommendations indicate an actor responsible for implementation.	no (1), yes (4)	More concrete indication than 'directed to actor'	47			33		80	
6.e	Recommendations are time-bound.	no (1), yes (4)		65			15		80	
6.f	Lessons learned are derived.	no (1), yes (4)		27			53		80	
7.	Annex									
7.1	7.1 Evaluation Team		7.1h		2	3	8	67	80	13
7.1a	Team members are presented.	no (1), yes (4)		56			24		80	
7.1b	Team composition is justified.	no (1), yes (4)		71			9		80	
7.1c	Team is gender-balanced, according to report.	no (1), yes (4); n.a.		8			20	52	80	28
7.1d	Team has thematic expertise, according to report.	no (1), yes (4); n.a.		1			17	62	80	18
7.1e	Team has evaluation expertise, according to report.	no (1), yes (4); n.a.		4			17	59	80	21
7.1f	Team has local expertise, according to report.	no (1), yes (4); n.a.		5			20	55	80	25
7.1g	There is incidence in the report for lack of independence.	no (1), yes (4)		77			3		80	
7.1h	Team composition is appropriate. (agg)	completely inappropriate (1), rather inappropriate (2), rather appropriate (3), completely appropriate (4), n.a.	summary indicator from above, (1) three or more of the following; incidence for lack of independence, no local expertise, no evaluation expertise, no thematic expertise, and no gender-balance, (2) if max. three of the former, (3) only gender-balance and one other item can be missing but not lack of independence, (4)only gender-balance can be missing		2	3	8	67	80	13
7.2	Report contains ToRs.	no (1), yes (4)		17			63		80	
7.3	Other annexes							80	80	0
7.3a	Report contains list of people interviewed.	no (1), yes (4)		21			59		80	
7.3b	Report contains documents consulted.	no (1), yes (4)		27			53		80	
7.3c	Report addresses internal quality assurance.	no (1), yes (4)		69			11		80	
7.3d	Report addresses external quality assurance.	no (1), yes (4)		71			9		80	
7.3e	Report contains a two-pager as communication tool.	no (1), yes (4)		67			13		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
7.3f	Data collection instruments are provided.	no (1), rather no (2), rather yes (3), yes (4)	(1) no data collection instruments are provided, (2) one data collection instrument, (3) most data collection instruments, (4) all data collection instruments	50	11	5	14		80	
8.	Cross-cutting objectives and HRBA		$(8.1+8.2+8.4+8.5)/4$	20	41	19			80	
8.1	Gender equality is integrated in the report.	no (1), rather no (2), rather yes (3), yes (4)	(1) not integrated at all, (2) integrated only sporadically in few (e.g. only two chapters) (3) reference to topics in findings, conclusions and recommendations but not comprehensively, (4) integrated in findings, conclusions and recommendation with separate sections or paragraphs PLEASE ALSO INCLUDE RIGHTS OF WOMEN AND GIRLS	13	31	19	17		80	
8.1a	The intervention is targeted to gender equality.	no (1), yes (4)		49			31		80	
8.2	Non-discrimination is integrated in report.	no (1), rather no (2), rather yes (3), yes (4)	(1) not integrated at all, (2) integrated only sporadically in few (e.g. only two chapters) (3) reference to topics in findings, conclusions and recommendations but not comprehensively, (4) integrated in findings, conclusions and recommendation with separate sections or paragraphs PLEASE ALSO INCLUDE REDUCTION OF INEQUALITY/EQUAL OPPORTUNITIES TO PARTICIPATE/RIGHTS OF THE MOST VULNERABLE	44	24	6	6		80	
8.2a	The intervention is targeted to non-discrimination.	no (1), yes (4)		58			22		80	
8.2b	The inclusion of persons with disabilities is discussed in the report.	no (1), rather no (2), rather yes (3), yes (4)	(1) not integrated at all, (2) integrated only sporadically in few (e.g. only two chapters) (3) reference to topics in findings, conclusions and recommendations but not comprehensively, (4) integrated in findings, conclusions and recommendation with separate sections or paragraphs	63	8	3	6		80	



name	Specification	Rating.1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
8.4	Climate sustainability is integrated in report.	no (1), rather no (2), rather yes (3), yes (4)	(1) not integrated at all, (2) integrated only sporadically in few (e.g. only two chapters) (3) reference to topics in findings, conclusions and recommendations but not comprehensively, (4) integrated in findings, conclusions and recommendation with separate sections or paragraphs PLEASE ALSO INCLUDE CLIMATE CHANGE PREPAREDNESS AND MITIGATION/ CLIMATE RESILIENCE/LOW EMISSION DEVELOPMENT	54	16	7	3		80	
8.4a	The intervention is targeted to climate sustainability.	no (1), yes (4)		66			14		80	
8.5	Human rights-based approach is integrated in report	no (1), rather no (2), rather yes (3), yes (4)	(1) not integrated at all, (2) integrated only sporadically in few (e.g. only two chapters) (3) reference to topics in findings, conclusions and recommendations but not comprehensively, (4) integrated in findings, conclusions and recommendation with separate sections or paragraphs	41	23	10	6		80	
11	Policy priority areas	no (1), yes (4)		74			6		80	
11.1	1 women's and girls' rights	no (1), yes (4)	(4) if evaluator direct referral to the PPA	79			1		80	
11.2	2 sustainable economies and decent work	no (1), yes (4)	(4) if evaluator direct referral to the PPA	78			2		80	
11.3	3 democratic societies and education	no (1), yes (4)	(4) if evaluator direct referral to the PPA	79			1		80	
11.4	4 natural resources and climate	no (1), yes (4)	(4) if evaluator direct referral to the PPA	76			4		80	
11.5	5 humanitarian assistance	no (1), yes (4)	(4) if evaluator direct referral to the PPA	79			1		80	
12	Covid-19 related issues									
12.1	Pandemic is acknowledged in the report.	no (1), yes (4), n.a.	n.a. if intervention and evaluation were completed before the pandemic.	2			11	67	80	13
12.2	Pandemic led to remote evaluation.	no (1), yes (4), n.a.	n.a. if intervention and evaluation were completed before the pandemic.	9			3	68	80	12
12.3	Pandemic led to semi-remote evaluation.	no (1), yes (4), n.a.	n.a. if intervention and evaluation were completed before the pandemic.	5			7	68	80	12
12.4	Pandemic negatively affected quality of the evaluation.	no (1), yes (4), n.a.	n.a. if intervention and evaluation were completed before the pandemic.	8			4	68	80	12
9.	General issues									
9.1	Documentation on evaluation process									





name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
9.1a	Deviations from planned implementation of evaluation are described.	no (1), yes (4)		57			23		80	
9.1b	Report mentions validation by stakeholders, i.e. validation workshop.	no (1), yes (4)	Project staff, representatives of beneficiaries, implementing organisation	57			23		80	
9.1c	Report mentions validation by MFA or other commissioners.	no (1), yes (4)		62			18		80	
9.2	Structure and style									
9.2a	Report is structured according to MFA template. (check annex)	no (1), yes (4)	Summary, Introduction, Methodology, Context Analysis, Findings, Conclusions, Recommendations, References, Evaluation Team, ToR, People Interviewed, Documents Consulted, if chapters missing, specify in comments	74			6		80	
9.2b	Report is properly edited.	no (1), yes (4)	Clear labelling of graphs and tables. Clear headlines and visual structure.	18			62		80	
9.2c	Report is written in clear language.	no (1), yes (4)		9			71		80	
9.3	Evaluation questions									
9.3a	The evaluation report answers evaluation questions defined in the ToR.	no (1), yes (4)	comment and be rather generous. n.a. of ToR or evaluation questions missing.	19			61		80	
10.	Summary	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(10.2*4+10.3+10.4*2)/7$	11	41	28			80	
10.1	Report contains executive summary.	no (1), yes (4)		4			76		80	
10.2	Completeness of summary	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	$(10.2a+10.2b+10.2c+10.2d+10.2e+10.2f+10.2g+10.2h+10.2i+10.2j*4+10.k)/14$	6	31	38	5		80	
10.2a	Summary describes rationale/purpose of evaluation.	no (1), yes (4)		39			41		80	
10.2b	Summary describes objectives of evaluation.	no (1), yes (4)	Look for elaborations.	50			30		80	
10.2c	Summary describes the intervention.	no (1), yes (4)		30			50		80	
10.2d	Summary describes the scope of the evaluation.	no (1), yes (4)	time, area, components	64			16		80	
10.2e	Summary describes the evaluation design.	no (1), yes (4)		73			7		80	
10.2f	Summary describes the methods.	no (1), yes (4)		42			38		80	
10.2g	Summary describes the findings.	no (1), yes (4)		8			72		80	
10.2h	Summary describes the conclusions.	no (1), yes (4)		30			50		80	
10.2i	Summary describes recommendations.	no (1), yes (4)	Also within summarising table ok	9			71		80	



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
10.2j	Summary contains a summarising table (incl. findings, conclusions and recommendations).	no (1), very incomplete (2), partly incomplete (3), complete (4)	(1) no table at all, (2) incomplete table with only findings, conclusions OR recommendations, (3) incomplete table with only two of this three elements, (4) complete table.	48	3	6	23		80	
10.2k	Summary describes lessons learned.	no (1), yes (4)		57			23		80	
10.3	Style			9			71		80	
10.3a	Summary is written in clear language.	no (1), yes (4)		9			71		80	
10.4	Consistency			28			52		80	
10.4a	Summary is consistent with report.	no (1), yes (4)		28			52		80	
OverallRating	OVERALL RATING of the Evaluation Report	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	(13+2+4+56+10+8)/6	3	40	37			80	
TOR	ToR available			10			70		80	
21.	Intervention	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate 21.1 + 21.3	2	19	39	10	10	80	70
21.1	Context of the development intervention	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate 21.1a.,21.1h, 21.1i, 21.1j. 21.1.k	9	29	26	6	10	80	70
21.1a	Finnish policy context	no (1), yes (4)		41			29	10	80	70
21.1b	Policy priority areas	no (1), yes (4)		25			45	10	80	70
21.1c	1 women's and girls' rights	no (1), yes (4)	no if policy priority area does not fit to the intervention, yes if intervention fits to the policy priority area. If reference is provided, leave a note in the comment	50			18	12	80	68
21.1d	2 sustainable economies and decent work	no (1), yes (4)	no if policy priority area does not fit to the intervention, yes if intervention fits to the policy priority area. If reference is provided, leave a note in the comment.	43			21	16	80	64
21.1e	3 democratic societies and education	no (1), yes (4)	no if policy priority area does not fit to the intervention, yes if intervention fits to the policy priority area. If reference is provided, leave a note in the comment.	48			19	13	80	67
21.1f	4 natural resources and climate	no (1), yes (4)	no if policy priority area does not fit to the intervention, yes if intervention fits to the policy priority area. If reference is provided, leave a note in the comment.	41			24	15	80	65



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
21.1g	5 humanitarian assistance	no (1), yes (4)	no if policy priority area does not fit to the intervention, yes if intervention fits to the policy priority area. If reference is provided, leave a note in the comment.	58			4	18	80	62
21.1h	international policy context	no (1), yes (4)		50			20	10	80	70
21.1i	target area's policy context	no (1), yes (4)		26			44	10	80	70
21.1j	development context	no (1), yes (4)	Is there reference on whether the intervention tackles a developmental issue in this country, region, or area?	21			49	10	80	70
21.1k	context with respect to cross-cutting issues	no (1), yes (4)		42			28	10	80	70
21.2	reference to relevant issues of previous evaluations	no (1), yes (4)	Please code	40			30	10	80	70
21.3	Objectives, strategies and implementation of the Intervention	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate		7	33	30	10	80	70
21.3a	description of intervention objectives	no (1), yes (4)		1			69	10	80	70
21.3b	description of implementation strategies of the intervention	no (1), yes (4)		19			51	10	80	70
21.3c	description of resources for implementation of the intervention	no (1), yes (4)		30			40	10	80	70
21.3d	reference to cross-cutting issues relevant for intervention	no (1), yes (4)		43			27	10	80	70
21.3e	description of stakeholders and their role	no (1), yes (4)		23			47	10	80	70
21.3f	description of period of the intervention	no (1), yes (4)		11			59	10	80	70
21.4	Purpose, objectives, and scope of the evaluation	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate		10	39	21	10	80	70
21.4.1	Rationale and purpose	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	2	2	31	35	10	80	70
21.4.1a	rationale for evaluation	no (1), yes (4)		3			67	10	80	70
21.4.1b	rationale for point of time of evaluation	no (1), yes (4)		20			50	10	80	70
21.4.1c	intended users of evaluation	no (1), yes (4)		19			51	10	80	70
21.4.1d	intended use of evaluation	no (1), yes (4)		9			61	10	80	70
21.4.2	Objectives	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate (21.4.2a*2+21.4.2b)/3	4		49	17	10	80	70
21.4.2a	objectives of the evaluation	no (1), yes (4)		4			66	10	80	70



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
21.4.2b	prioritization of objectives	no (1), yes (4)		53			17	10	80	70
21.4.3	Scope	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	2	14	34	20	10	80	70
21.4.3a	intervention dimensions to be evaluated	no (1), yes (4)	yes, if dimensions are specified. also yes, when it is clear that all dimensions of the intervention shall be evaluated.	18			52	10	80	70
21.4.3b	stakeholder groups involved	no (1), yes (4)		30			40	10	80	70
21.4.3c	geographical area	no (1), yes (4)		29			41	10	80	70
21.4.3d	time span	no (1), yes (4)		13			57	10	80	70
21.4.3e	connection of evaluation to other supporting sectors or themes	no (1), yes (4)		63			7	10	80	70
21.5	Evaluation questions	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	14		41	15	10	80	70
21.5a	evaluation questions adapted to the specific information needs	no (1), yes (4), n.a.	n.a. if no questions	15			55	10	80	70
21.5b	maximum of 12 evaluation questions	no (1), yes (4), n.a.	n.a. if no questions	54			16	10	80	70
21.6	Evaluation criteria (relevant criteria for the evaluation, OECD/DAC and others)	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate of 21.6a-21.6h	1	6	16	47	10	80	70
21.6a	relevance	no (1), yes (4)		9			61	10	80	70
21.6b	coherence	no (1), yes (4)	formerly triple C and aid effectiveness separately	36			34	10	80	70
21.6c	effectiveness	no (1), yes (4)		4			66	10	80	70
21.6d	efficiency	no (1), yes (4)		9			61	10	80	70
21.6e	impact	no (1), yes (4)		18			52	10	80	70
21.6f	sustainability	no (1), yes (4)		9			61	10	80	70
21.6g	other criteria added by the commissioner	no (1), yes (4)		38			32	10	80	70
21.6h	specification of other criteria	free input		4			3	73	80	7
21.7	Methodology	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	11	30	26	3	10	80	70
21.7a	request for mix of qualitative and quantitative methods	no (1), yes (4)		30			40	10	80	70
21.7b	request for triangulation	no (1), yes (4)		41			29	10	80	70



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
21.7c	request for disaggregated analysis	no (1), yes (4)		51			19	10	80	70
21.7d	specification of available materials	no (1), yes (4)		40			30	10	80	70
21.7e	specification of envisaged data collection techniques	no (1), yes (4)		36			34	10	80	70
21.7f	specification of envisaged data analysis techniques	no (1), yes (4)		67			3	10	80	70
21.8	Feasibility (Scope of work and given timeframe and resources are feasible.)	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	19	25	12	14	10	80	70
21.8a	evaluation budget in ToR	no (1), yes (4)		40			30	10	80	70
21.8b	feasible scope of evaluation given budget	no (1), yes (4)	No (1) if there is no budget in ToR	44			26	10	80	70
21.8c	feasible scope of evaluation given time resources	no (1), yes (4)		35			35	10	80	70
21.9	Evaluation Process and QA (The evaluation process is clearly explained in the ToR.)	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	1	26	34	9	10	80	70
21.9.1	Evaluation process	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	3	23	13	34	10	80	70
21.9.1a	outline of phases of evaluation process	no (1), yes (4)		22			48	10	80	70
21.9.1b	outline of sequencing of activities	no (1), yes (4)		16			54	10	80	70
21.9.1c	outline of approximate duration of activities	no (1), yes (4)		26			44	10	80	70
21.9.1d	place of work for activities	no (1), yes (4)		28			42	10	80	70
21.9.1e	specification of roles and responsibilities of commissioner and evaluator(s)	no (1), yes (4)		44			26	10	80	70
21.9.2	Deliverables	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate	3	28		39	10	80	70
21.9.2a	specification of deliverables	no (1), yes (4)		4			66	10	80	70
21.9.2b	specification of milestones with timeline	no (1), yes (4)		30			40	10	80	70
21.9.3	Quality assurance (reference to what kind of quality assurance is desired)	no (1), yes (4)	Please code.	51			19	10	80	70
21.10	Overarching and cross-cutting criteria	inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate 1.10b+21.10d+21.h+21.10i+21.10j	24	19	18	9	10	80	70
21.10a	pointing to gender equality as cross-cutting objective	no (1), yes (4)	rights of women and girls	31			39	10	80	70
21.10b	gender equality requested to be analysed by evaluator	no (1), yes (4)		32			38	10	80	70



name	Specification	Rating,1-4	Guidance	1	2	3	4	missing/ n.a./ no ToRs	Total	Total valid
21.10c	pointing to non-discrimination as cross-cutting objective	no (1), yes (4)	reduction of inequality, equal opportunities to participate, rights of the most vulnerable,	46			24	10	80	70
21.10d	non-discrimination requested to be analysed by evaluator	no (1), yes (4)		50			20	10	80	70
21.10e	special emphasis given to persons with disabilities	no (1), yes (4)		67			3	10	80	70
21.10f	inclusion of persons with disabilities requested to be analysed by evaluator	no (1), yes (4)		67			3	10	80	70
21.10g	pointing to climate sustainability as cross-cutting objective	no (1), yes (4)	climate change preparedness and mitigation, climate resilience, low emission development	49			21	10	80	70
21.10h	climate sustainability requested to be analysed by evaluator	no (1), yes (4)		50			20	10	80	70
21.10i	pointing to HRBA	no (1), yes (4)		41			29	10	80	70
21.10j	HRBA requested to be analysed by evaluator	no (1), yes (4)		49			21	10	80	70
Ethics	request for ethical considerations	no (1), yes (4)		58			12	10	80	70
Overall		inadequate (1), need for improvement (2), satisfactory (3), good or very good (4)	aggregate (21+21.4+21.5+21.6+21.7+21.9+21.10)/7		24	45	1	10	80	70
exclusion	Evaluation should be taken out of the sample because of severe quality issues.	no (1), yes (4)	Reasons can be that the report is completely unclear written, unclear structured, does not flow logically from data and findings to conclusions and recommendations or that data collection and analysis methods do not meet evaluation standards.	72			8		80	
	colour reflects newly inserted aspects									
	colour indicates aggregates affected by new aspects (applied for dynamic comparison)									



# Annex 7: Content Assessment Tool

Specification	Rating.1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
Key information									
Evaluation ID									
Evaluation name									
Year									
ToR available									
Type of Evaluation									
Evaluator ID									
Selected for review									
Reviewer ID									
Specification	Rating 1-4	Guidance							
1. Relevance		Is the intervention doing the right things? The extent to which the intervention's objectives and design respond to beneficiaries', global, country and partner/institution needs, policies and priorities, and continue to do so if circumstances change.							
1.1a Relevance is discussed.	no (1), yes (4)	Transferred from quality tool	6			66		72	72
1.1b Relevance is methodologically appropriately captured.	n.a., no (1), rather no (2), rather yes (3), yes (4)	Transferred from quality tool, if 'no' or 'n.a.' no analysis of this subsection possible, all 'n.a.'	1	20	36	9	6	72	66
1.2 How do the evaluators assess the relevance of the intervention in the evaluation report?	n.a., not relevant at all (1), somewhat relevant (2), moderately relevant (3), highly relevant (4)	n.a. report is not analysing this aspect, (1) all aspects analysed in relevance are assessed negatively, (2) most aspects analysed in relevance are assessed negatively, (3) most aspects analysed in relevance are assessed positively, (4) all aspects analysed in relevance are assessed positively PLEASE ONLY REFER TO THE REPORT NOT TO THE EXECUTIVE SUMMARY, PLEASE ONLY TAKE THE OVERALL ASSESSMENTS PROVIDED IF ANY.		2	23	36	11	72	61



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
1.3a According to the evaluators, does the intervention meet the needs of the target group (i.e. those for whom the intervention has been designed)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) intervention does not meet the needs of the target group, (2) intervention does mostly not meet the needs of the target group, (3) intervention does somehow meet the need of the target group, (4) intervention does mostly meet the needs of the target group PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE MADE, E.G. LINK TO POLICIES, THEN RATE THIS SUBQUESTION WITH N.A. IF THE TARGET GROUP IS THE POOR/LOCAL POPULATION PLEASE COPY YOUR RATING TO 1.4a (ONLY IN THE EXCEL)		2	16	36	18	72	54
1.3b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL. IF THE TARGET GROUP IS THE POOR/LOCAL POPULATION PLEASE COPY YOUR RATING TO 1.4b (ONLY IN THE EXCEL)			1		71	72	1
1.3c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL. IF THE TARGET GROUP IS THE POOR/LOCAL POPULATION PLEASE COPY YOUR RATING TO 1.4c (ONLY IN THE EXCEL)			1		71	72	1
1.4a According to the evaluators, does the intervention meet the needs of the final beneficiaries (i.e. the local/poor people)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) intervention does not meet the needs of the final beneficiaries, (2) intervention does mostly not meet the needs of the final beneficiaries, (3) intervention does somehow meet the need of the final beneficiaries, (4) intervention does mostly meet the needs of the final beneficiaries PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY E.G. ONLY DISCUSSION ON THE LOCAL GOVERNMENT AS TARGET GROUP, THEN RATE THIS SUBQUESTION WITH N.A. THUS, THIS IS NOT ABOUT GUESSING YOURSELF HOW IMPROVED LOCAL GOVERNMENT IS RELEVANT FOR POOR PEOPLE.	2	1	22	19	28	72	44
1.4b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
1.4c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, list all negative explanatory factors provided by the evaluators, if ambiguous please specify in key words. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR A SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT A LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND THEM FROM THIS EXCEL.							





Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
1.5 According to the evaluators, is the intervention consistent and supportive of the partner government/regional policies?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is inconsistent with partner government policies, (2) the intervention is mostly not consistent and supportive of partner government policies, (3) the intervention is mostly consistent and supportive of partner government policies, (4) the intervention is fully consistent and supportive partner government policies PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED. FOR NATIONAL INTERVENTIONS LOOK AT PARTNER GOVERNMENT FOR REGIONAL INTERVENTIONS, AT REGIONAL POLICIES E.G. EU POLICIES. MINIMUM REQUIREMENT FOR ASSESSMENT (i.e. 1, 2, 3 or 4) IS NAMING.	1		4	51	16	72	56
1.6 According to the evaluators, is the intervention consistent with the MFA development cooperation policy?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is inconsistent with the MFA development cooperation policy, (2) the intervention is mostly not consistent with the MFA development cooperation policy, (3) the intervention is mostly consistent with the MFA development cooperation policy, (4) the intervention is fully consistent with the MFA development cooperation policy PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED. MINIMUM REQUIREMENT FOR ASSESSMENT (i.e. 1, 2, 3 or 4) IS NAMING.		2	6	24	40	72	32
1.7 According to the evaluators, is the intervention addressing international conventions, policies, strategies or goals?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is not addressing international conventions, policies, strategies or goals, (2) the intervention is mostly not addressing international conventions, policies, strategies or goals, (3) the intervention is mostly addressing international conventions, policies, strategies or goals, (4) the intervention is strongly addressing international conventions, policies, strategies or goals PLEASE LOOK AT EXPLICIT ANSWER, IF THE ASPECT IS NOT DISCUSSED RATE N.A. THIS IS NOT THE PLACE TO MENTION THAT CONSISTENCY WITH MFA POLICY DOES INDIRECTLY MEAN CONSISTENCY WITH INTERNATIONAL GOALS. PLEASE KEEP IN MIND IF REGIONAL INTERVENTIONS ONLY REFER TO REGIONAL POLICIES THIS WOULD BE N.A.		1	1	28	42	72	30
1.9 According to the evaluators, is the intervention sensitive to the context conditions in which it takes place?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is not sensitive to the context conditions, (2) the intervention is mostly sensitive to the context conditions, (3) the intervention is mostly sensitive to the context conditions, (4) the intervention is strongly sensitive to the context conditions PLEASE LOOK AT EXPLICIT ANSWER, IF THE ASPECT IS NOT DISCUSSED RATE N.A.	1	7	10	22	32	72	40
1.10 According to the evaluators, has the intervention's design been appropriate?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention's design is not appropriate, (2) the intervention's design is mostly not appropriate, (3) the intervention's design is mostly appropriate, (4) the intervention's design is strongly appropriate PLEASE LOOK AT EXPLICIT ANSWER, IF THE ASPECT IS NOT DISCUSSED RATE N.A. PLEASE ANTICIPATE BALANCING DIFFERENCES AND TRADE-OFFS BETWEEN DIFFERENT PRIORITIES AND NEEDS	1	17	17	2	35	72	37
1.11 According to the evaluators, has the intervention been adapted to changes in the context to remain relevant?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention has not been adapted to changes in the context, (2) the intervention has mostly not been adapted to changes in the context, (3) the intervention has mostly been adapted to changes in the context, (4) the intervention has been strongly adapted to changes in the context PLEASE LOOK AT EXPLICIT ANSWER, IF THE ASPECT IS NOT DISCUSSED RATE N.A.	2	3	9	13	45	72	27



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
1.12 According to the evaluators, has the intervention been adapted to Covid-19 pandemic?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention has not been adapted to Covid-19 pandemic, (2) the intervention has mostly not been adapted to Covid-19 pandemic, (3) the intervention has mostly been adapted to Covid-19 pandemic, (4) the intervention has been strongly adapted to Covid-19 pandemic PLEASE LOOK AT EXPLICIT ANSWER, IF INTERVENTION HAS BEEN COMPLETE BEFORE COVID-19 PANDEMIC RATE N.A.	1		2	1	68	72	4
1.8 Is this section on the intervention a success story?	no (1), yes (4)	Do you have the impression that this section is a very good example for a very successful project? Then select yes. PLEASE FOCUS ON EXTRAORDINARY WORK.	48			17	7	72	65
9. Coherence		How well does the intervention fit? The compatibility of the intervention with other interventions in a country, sector or institution. The extent to which other interventions (particularly policies) support or undermine the intervention and vice versa.							
9.1a Coherence is discussed.	no (1), yes (4)	Transferred from quality tool	32			40		72	72
9.1b Coherence is appropriately captured.	n.a., no (1), rather no (2), rather yes (3), yes (4)	Transferred from quality tool, if 'no' or 'n.a.' no analysis of 7.2. PLEASE ASSESS SINGLE ASPECTS IF PROVIDED BY EVALUATORS (e.g. old triple C and aid effectiveness discussions).	5	20	13	2	32	72	40
9.2 How do the evaluators assess the coherence of the intervention in the evaluation report?	n.a., not coherent at all (1), somewhat coherent (2), moderately coherent (3), highly coherent (4)	n.a. report is not analysing this aspect, (1) all aspects analysed in coherence are assessed negatively, (2) most aspects analysed in coherence are assessed negatively, (3) most aspects analysed in coherence are assessed positively, (4) all aspects analysed in coherence are assessed positively PLEASE ONLY REFER TO THE REPORT NOT TO THE EXECUTIVE SUMMARY, PLEASE ONLY CODE THE OVERALL ASSESSMENTS PROVIDED IF ANY. THIS IS NOT THE PLACE TO CODE ALL DETAILS; THEY ARE CAPTURED BELOW. ONLY APPLICABLE IF SUMMARY ASSESSMENT AVAILABLE (COHERENCE SECTION)	5	8	9	6	44	72	28
9.3 According to the evaluators, is the intervention synergistic to and interlinked with other interventions carried out by MFA (i.e. internal coherence)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) all aspects with regard to internal coherence are assessed negatively, (2) most aspects analysed with regard to internal coherence are assessed negatively, (3) most aspects with regard to internal coherence are assessed positively, (4) all aspects analysed with regard to internal coherence are assessed positively PLEASE TAKE ALL ASPECTS MENTIONED BY THE EVALUATOR INTO CONSIDERATION.	4	6	5	5	52	72	20
9.4 According to the evaluators, is the intervention consistent with other actors' interventions in the same context (i.e. complementarity, co-ordination with others, adding value while avoiding duplication of effort) (i.e. external coherence)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) all aspects with regard to external coherence are assessed negatively, (2) most aspects analysed with regard to external coherence are assessed negatively, (3) most aspects with regard to external coherence are assessed positively, (4) all aspects analysed with regard to external coherence are assessed positively PLEASE RATE ONLY 4 IF COMPLEMENTARITY AND COORDINATION ARE BOTH ASSESSED POSITIVELY.	2	5	10	6	49	72	23
9.5 Is this section on the intervention a success story?	no (1), yes (4)	Do you have the impression that this section is a very good example for a very successful project? Then select yes. PLEASE FOCUS ON EXTRAORDINARY WORK.	29			6	37	72	35



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
2. Effectiveness		Is the intervention achieving its objectives? The extent to which the intervention achieved, or is expected to achieve, its objectives and its results, including any differential results across groups.							
2.1a Effectiveness is discussed.	no (1), yes (4)	Transferred from quality tool	3			69		72	72
2.1b Effectiveness is appropriately captured.	n.a., no (1), rather no (2), rather yes (3), yes (4)	Transferred from quality tool, if 'no' or 'n.a.' no analysis of this subsection possible, all 'n.a.'	1	27	32	9	3	72	69
2.2 How do the evaluators assess the effectiveness of the intervention in the evaluation report?	n.a., not effective at all (1), somewhat effective (2), moderately effective (3), highly effective (4)	n.a. report is not analysing this aspect, (1) all aspects analysed in effectiveness are assessed negatively, (2) most aspects analysed in effectiveness are assessed negatively, (3) most aspects analysed in effectiveness are assessed positively, (4) all aspects analysed in effectiveness are assessed positively PLEASE ONLY REFER TO THE REPORT NOT TO THE EXECUTIVE SUMMARY, PLEASE ONLY CODE THE OVERALL ASSESSMENTS PROVIDED IF ANY. THIS IS NOT THE PLACE TO CODE ALL DETAILS; THEY ARE CAPTURED BELOW.	2	18	35	11	6	72	66
2.4a According to the evaluators, have the outcomes of the intervention been achieved?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. outcomes are not analysed in report, (1) no outcomes have been achieved, (2) most outcomes have not been achieved, (3) most outcomes have been achieved, (4) all outcomes have been achieved PLEASE RATE THIS IF THE ASSESSMENT IS RATHER AT THE LEVEL OF LONG-TERM OUTCOMES OF THE INTERVENTION. IF SO-CALLED OUTCOMES ARE OBVIOUS OUTPUTS/ACTIVITIES PLEASE RATE N.A.	4	11	34	14	9	72	63
2.4b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL. PLEASE LIST REASONS FOR OUTCOME ACHIEVEMENT NOT OUTCOMES.							
2.4c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL. PLEASE LIST REASONS FOR OUTCOME ACHIEVEMENT NOT OUTCOMES.							
2.9a According to the evaluators, has the outcome achievement been affected by Covid-19 pandemic?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. the intervention was implemented before Covid-19 pandemic, (1) the outcome achievement has not been affected by Covid-19 pandemic, (2) the outcome achievement has been only slightly affected by Covid-19 pandemic, (3) the outcome achievement has been moderately affected by Covid-19 pandemic, (4) the outcome achievement has been heavily affected by Covid-19 pandemic.		3	1	1	67	72	5
2.9b If yes, please specify.									
2.9c What reasons are provided?									



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
2.5a According to the evaluators, has the intervention resulted in benefits for the target group (i.e. those for whom the intervention was designed)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. results for the target group are not analysed in report, (1) no benefits for the target group have been achieved, (2) very few benefits for the target group have been achieved, (3) moderate benefits for the target group have been achieved, (4) many benefits for the target group have been achieved PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A. IF THE TARGET GROUP IS THE POOR/LOCAL POPULATION PLEASE COPY YOUR RATING TO 2.6a (ONLY IN THE EXCEL)	3	11	28	13	17	72	55
2.5b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL. IF THE TARGET GROUP IS THE POOR/LOCAL POPULATION PLEASE COPY YOUR RATING TO 2.6b (ONLY IN THE EXCEL)							
2.5c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL. IF THE TARGET GROUP IS THE POOR/LOCAL POPULATION PLEASE COPY YOUR RATING TO 2.6c (ONLY IN THE EXCEL)							
2.6a According to the evaluators, has the intervention resulted in benefits for the final beneficiaries (i.e. the local/poor people)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. results for the final beneficiaries are not analysed in report, (1) no benefits for the final beneficiaries have been achieved, (2) very few benefits for the final beneficiaries have been achieved, (3) moderate benefits for the final beneficiaries have been achieved, (4) many benefits for the final beneficiaries have been achieved PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	2	13	18	16	23	72	49
2.6b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
2.6c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
2.7 What are the main results of the gender-analysis provided by the evaluator?	no (1), yes (4)	no, if no gender-analysis. PLEASE CODE MAIN GENDER RESULTS.	14			31	27	72	45
2.8 Is this section on the intervention a success story?	no (1), yes (4)	Do you have the impression that this section is a very good example for a very successful project? Then select yes. PLEASE FOCUS ON EXTRAORDINARY WORK.	55			13	4	72	68



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
3. Efficiency		How well are resources being used? The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.							
3.1a Efficiency is discussed.	no (1), yes (4)	Transferred from quality tool	7			65		72	72
3.1b Efficiency is appropriately captured.	n.a., no (1), rather no (2), rather yes (3), yes (4)	Transferred from quality tool, if 'no' or 'n.a.' no analysis of this subsection possible, all 'n.a.'	3	33	28	1	7	72	65
3.2 How do the evaluators assess the efficiency of the intervention in the evaluation report?	n.a., not efficient at all (1), somewhat efficient (2), moderately efficient (3), highly efficient (4)	n.a. report is not analysing this aspect, (1) all aspects analysed in efficiency are assessed negatively, (2) most aspects analysed in efficiency are assessed negatively, (3) most aspects analysed in efficiency are assessed positively, (4) all aspects analysed in efficiency are assessed positively PLEASE ONLY REFER TO THE REPORT NOT TO THE EXECUTIVE SUMMARY, PLEASE ONLY CODE THE OVERALL ASSESSMENTS PROVIDED IF ANY. THIS IS NOT THE PLACE TO CODE ALL DETAILS; THEY ARE CAPTURED BELOW.	4	13	29	8	18	72	54
3.3a According to the evaluators, is/was the implementation of the intervention on time?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is/was not at all on time, (2) the intervention is/was mostly not on time, (3) the intervention is/was mostly on time, (4) the intervention is/was on time or ahead of schedule.	5	18	16	5	28	72	44
3.3b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.3c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.4a According to the evaluators, have the inputs been converted into high quality outputs?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) inputs have not been converted into high quality outputs, (2) the inputs have mostly not been converted into high quality outputs, (3) the inputs have mostly been converted into high quality outputs, (4) all inputs have been converted into high quality outputs PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A. CHECK FOR EXPLICIT ASSESSMENTS ON THE QUALITY, THIS IS NOT THE SAME AS WHETHER SOMETHING HAS BEEN REACHED. HOWEVER, THIS IS EXPECTED TO BE OFTEN N.A.	2	2	19	12	37	72	35
3.4b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.4c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							



Specification	Rating.1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
3.9a According to the evaluators, have the inputs been converted into high quality outcomes?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) inputs have not been converted into high quality outcomes, (2) the inputs have mostly not been converted into high quality outcomes, (3) the inputs have mostly been converted into high quality outcomes, (4) all inputs have been converted into high quality outcomes PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A. CHECK FOR EXPLICIT ASSESSMENTS ON THE QUALITY, THIS IS NOT THE SAME AS WHETHER SOMETHING HAS BEEN REACHED. HOWEVER, THIS IS EXPECTED TO BE OFTEN N.A.	2	1	9	6	54	72	18
3.9b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.9c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.5a According to the evaluators, is the intervention efficient regarding costs?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is not at all cost-efficient regarding costs, (2) the intervention is mostly not cost-efficient, (3) the intervention is mostly cost-efficient, (4) the intervention is fully cost-efficient	5	8	23	11	25	72	47
3.5b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.5c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.6a According to the evaluators, is the intervention efficient regarding personnel?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is not efficient regarding personnel, (2) the intervention is mostly not efficient regarding personnel, (3) the intervention is mostly efficient regarding personnel, (4) the intervention is fully efficient regarding personnel PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	7	16	8	3	38	72	34
3.6b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.6c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							



Specification	Rating.1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
3.7a According to the evaluators, is the implementation management efficient?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention is not efficient regarding implementation management, (2) the intervention is mostly not efficient regarding implementation management, (3) the intervention is mostly efficient regarding implementation management, (4) the intervention is fully efficient regarding implementation management	7	14	19	7	25	72	47
3.7b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.7c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
3.8 Is this section on the intervention a success story?	no (1), yes (4)	Do you have the impression that this section is a very good example for a very successful project? Then select yes. PLEASE FOCUS ON EXTRAORDINARY WORK.	59			3	10	72	62
4. Impact		What differences does the intervention make? The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.							
4.1a Impact is discussed.	no (1), yes (4)	Transferred from quality tool	14			58		72	72
4.1b Impact is appropriately captured.	n.a., no (1), rather no (2), rather yes (3), yes (4)	Transferred from quality tool, if 'no' or 'n.a.' no analysis of this subsection possible, all 'n.a.'	6	26	22	4	14	72	58
4.2 How do the evaluators assess the impact of the intervention in the evaluation report?	n.a., no impact at all (1), some impact (2), moderate impact (3), high impact (4)	n.a. report is not analysing this aspect, (1) the intervention has no impacts at all, (2) the intervention has mostly no impact, (3) the intervention has some impact, (4) the intervention has a high impact PLEASE ONLY REFER TO THE REPORT NOT TO THE EXECUTIVE SUMMARY, PLEASE ONLY CODE THE OVERALL ASSESSMENTS PROVIDED IF ANY. THIS IS NOT THE PLACE TO CODE ALL DETAILS; THEY ARE CAPTURED BELOW.	4	13	14	8	33	72	39
4.3 According to the evaluators, did the intervention contribute to its overall objective/reach its intended impact?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention did not contribute, (2) the intervention did contribute very little, (3) the intervention did contribute moderately, (4) the intervention did contribute highly PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A. THIS IS THE PLACE TO LOOK AT HIGHER LEVEL IMPACTS, THERE MIGHT BE OVERLAPS TO LONG-TERM OUTCOMES, THIS IS OKAY, HOWEVER DO NOT RATE ANY OUTPUT LEVEL ASSESSMENTS HERE. HERE WE DO NOT ASK FOR REASONS AS THEY ARE CAPTURED IN THE THEMATIC SUB-SECTIONS BELOW.	4	10	14	8	36	72	36



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
4.4a According to the evaluators, does the intervention have any unintended positive impacts?	n.a., no (1), yes (4)	n.a. report is not analysing this aspect, (1) the intervention did not have positive unintended impacts, (4) the intervention did have positive unintended impacts PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	2			12	58	72	14
4.4b If any, please specify									
4.4c What reasons are provided?		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
4.5a According to the evaluators, does the intervention have any unintended negative impacts?	n.a., no (1), yes (4)	n.a. report is not analysing this aspect, (1) the intervention did not have negative unintended impacts, (4) the intervention did have negative unintended impacts PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	6			4	62	72	10
4.5b If any, please specify									
4.5c What reasons are provided?		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
4.6a According to the evaluators, does the intervention contribute to enhance the quality of life of the final beneficiaries/ reach final beneficiaries?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention did not contribute, (2) the intervention did contribute very little, (3) the intervention did contribute moderately, (4) the intervention did contribute highly PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A., PLEASE DO NOT JUDGE WHETHER THE ASSESSMENT OF THE EVALUATOR IS VALID FROM YOUR PERSPECTIVE, RATHER CAPTURE THE ANSWER. IF SOMETHING SEEMS SUSPICIOUS, USE THE COMMENT FIELD.	2	8	14	7	41	72	31
4.6b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
4.6c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							





Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
4.7a According to the evaluators, does the intervention contribute to enhance institutional quality (i.e. institutions/services in the partner country/region have been improved)?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention did not contribute, (2) the intervention did contribute very little, (3) the intervention did contribute moderately, (4) the intervention did contribute highly PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.		8	18	9	37	72	35
4.7b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
4.7c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) negative?)		If clear, provide synthesis of positive explanatory factors provided by the evaluators. PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND FROM THIS EXCEL.							
4.8a According to the evaluators, has the intervention contributed to changes in the partner country's/ region's policies/ to sector reforms?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the intervention did not contribute, (2) the intervention did contribute very little, (3) the intervention did contribute moderately, (4) the intervention did contribute highly PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	6	5	13	8	40	72	32
4.8b If any, please specify									
4.8c What reasons are provided?		If applicable, list all explanatory factors provided by the evaluators, if ambiguous please specify in key words							
4.9 Is this section on the intervention a success story?	no (1), yes (4)	Do you have the impression that this section is a very good example for a very successful project? Then select yes. PLEASE FOCUS ON EXTRAORDINARY WORK.	46			6	20	72	52
<b>5. Sustainability</b>									
5.1a Sustainability is discussed.	no (1), yes (4)	Transferred from quality tool	3			69		72	72
5.1b Sustainability is appropriately captured.	n.a., no (1), rather no (2), rather yes (3), yes (4)	Transferred from quality tool, if 'no' or 'n.a.' no analysis of this subsection possible, all 'n.a.'	4	39	22	4	3	72	69
5.2 How do the evaluators assess the sustainability of the intervention in the evaluation report?	n.a., not sustainable at all (1), somewhat sustainable (2), moderately sustainable (3), highly sustainable (4)	n.a. report is not analysing this aspect, (1) all aspects analysed in sustainability are assessed negatively, (2) most aspects analysed in sustainability are assessed negatively, (3) most aspects analysed in sustainability are assessed positively, (4) all aspects analysed in sustainability are assessed positively PLEASE ONLY REFER TO THE REPORT NOT TO THE EXECUTIVE SUMMARY, PLEASE ONLY CODE THE OVERALL ASSESSMENTS PROVIDED IF ANY. THIS IS NOT THE PLACE TO CODE ALL DETAILS; THEY ARE CAPTURED BELOW.	6	32	20	4	10	72	62



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
5.3a According to the evaluators, are benefits of the intervention likely to continue after the completion of the intervention? (i.e. Do the final beneficiaries further benefit after the intervention ends?)	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) benefits are not at all likely to continue, (2) benefits are rather not likely to continue, (3) benefits are rather likely to continue, (4) benefits are likely to continue PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	8	19	18	3	24	72	48
5.3b What reasons are provided for the positive assessment? (i.e. Why did the evaluators assess it (rather) positive?)		If applicable, list all positive explanatory factors provided by the evaluators, if ambiguous please specify in key words PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR A SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT A LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND THEM FROM THIS EXCEL.							
5.3c What reasons are provided for the negative assessment? (i.e. Why did the evaluators assess it (rather) not relevant?)		If applicable, list all negative explanatory factors provided by the evaluators, if ambiguous please specify in key words PLEASE KEEP IN MIND THAT IT HAS TO BE ON A GENERAL LEVEL AS THIS IS FOR A SUMMATIVE ANALYSIS OF FINNISH DEVCO, AT A LATER STAGE WE HAVE TO BE ABLE TO UNDERSTAND THEM FROM THIS EXCEL.							
5.4 According to the evaluators, does the target group have the capacity to make the intervention sustainable? (i.e. knowledge, know-how)	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) target group does not have the capacity at all, (2) target group does rather not have the capacity, (3) target group rather has the capacity, (4) target group has the capacity, Consider capacity as comprehensive concept, not only human but also institutional capacity PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A. IN CASE THE TARGET GROUP IS AT THE SAME TIME THE IMPLEMENTING ORGANISATION PLEASE COPY YOUR RATINGS TO 5.6.	8	15	20	5	24	72	48
5.5 According to the evaluators, does the target group have the financial means to make the intervention sustainable?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) beneficiaries do not have the financial means, (2) beneficiaries do rather not have the financial means, (3) beneficiaries rather have the financial means, (4) beneficiaries have the financial means PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A. IN CASE THE TARGET GROUP IS AT THE SAME TIME THE IMPLEMENTING ORGANISATION PLEASE COPY YOUR RATINGS TO 5.7.	18	13	7	2	32	72	40
5.6 According to the evaluators, do the implementing partner organisations have the institutional capacity to make the intervention sustainable?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) partners do not have the institutional capacity, (2) partners do rather not have the institutional capacity, (3) partners rather have the institutional capacity, (4) partners have the institutional capacity PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	3	10	8	3	48	72	24



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
5.7 According to the evaluators, do the implementing partner organisations have the financial means to make the intervention sustainable?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) partners do not have the financial means, (2) partners do rather not have the financial means, (3) partners rather have the financial means, (4) partners have the financial means PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	7	7	4	2	52	72	20
5.10 According to the evaluators, does the enabling environment allow sustainability?	n.a., no (1), rather no (2), rather yes (3), yes (4)	n.a. report is not analysing this aspect, (1) the enabling environment does not allow sustainability, (2) the enabling environment does mostly not allow sustainability, (3) the enabling environment does mostly allow sustainability, (4) the enabling environment does allow sustainability PLEASE LOOK AT EXPLICIT ANSWERS PROVIDED, IF THERE IS ONLY INDIRECT REFERENCE, THEN RATE THIS SUBQUESTION WITH N.A.	3	8	7	4	50	72	22
5.8 According to the evaluators, does the intervention have an exit strategy?	n.a., no (1), yes (4)	n.a. report is not analysing this aspect, (1) the intervention does not have an exit strategy, (4) the intervention has an exit strategy.	10			9	53	72	19
5.9 Is this section on the intervention a success story?	no (1), yes (4)	Do you have the impression that this section is a very good example for a very successful project? Then select yes. PLEASE FOCUS ON EXTRAORDINARY WORK.	61			4	7	72	65
<b>10. Policy Priority Areas</b>									
10.1a Does the intervention belong to PPA 1- Rights of women and girls?	no (1), yes (4)	The rights and status of all women and girls have been enhanced. (SDG 5 AND 1, 3, 4, 10) NOT MUCH EXPECTED IN OUR SAMPLE; ONLY ABOUT SHRH & GBV	66			6		72	72
10.1b Does the intervention belong to PPA1 - Outcome Area 1?	no (1), yes (4)	Outcome 1: The right of women and girls of all abilities to access high-quality non-discriminatory sexual and reproductive health services is protected (SDG3, T7; SDG5, T6) ALL RELATED TO SRHR	69			3		72	72
10.1c Does the intervention belong to PPA1 - Outcome Area 2?	no (1), yes (4)	Outcome 2: Women and girls of all abilities enjoy the right to live a life free of violence and abuse, and to make decisions concerning their bodies in compliance with CEDAW (SDG5, T2) ALL RELATED TO GBV AT INDIVIDUAL LEVEL	69			3		72	72
10.1d Does the intervention belong to PPA1 - Outcome Area 3?	no (1), yes (4)	Outcome 3: The rights of persons with disabilities, including their right to enjoy life free of violence, stigma and discrimination are protected. ALL RELATED TO GBV AT INSTITUTIONAL LEVEL	69			3		72	72
10.2a Does the intervention belong to PPA 2 - Sustainable economies and decent work?	no (1), yes (4)	Developing countries' own economies have generated more jobs, livelihood opportunities and well-being (SDGS 5, 8, 9, 12 AND 17) PRIVATE SECTOR; SUSTAINABLE ECONOMIES; DECENT WORK; SUSTAINABLE BUSINESS	50			22		72	72
10.2b Does the intervention belong to PPA2 - Outcome Area 1?	no (1), yes (4)	Outcome 1: Increased number of people, especially women, youth and those in vulnerable situations, have their right to decent work, livelihoods and income fulfilled (SDG1, T4) ALL RELATED TO DECENT WORK AND LIVELIHOODS	53			19		72	72



Specification	Rating 1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
10.2c Does the intervention belong to PPA2 - Outcome Area 2?	no (1), yes (4)	Outcome 2: The private sector grows, is responsible and supports sustainable development (SDG 8, T2) ALL RELATED TO PRIVATE SECTOR AND RESPONSIBLE GROWTH	64			8		72	72
10.2d Does the intervention belong to PPA2 - Outcome Area 3?	no (1), yes (4)	Outcome 3: Developing country governments promote responsible business conduct and support a solid business enabling environment that enhances innovation (SDG 8, T3 and T8; SDG 9 T5) ALL RELATED TO RESPONSIBLE BUSINESS AT GOVERNMENTAL LEVEL.	59			13		72	72
10.2e Does the intervention belong to PPA2 - Outcome Area 4?	no (1), yes (4)	Outcome 4: The international community promotes responsible business conduct and innovations in a manner that benefits poor people, especially youth and women (SDG 17, T16) ALL RELATED TO RESPONSIBLE BUSINESS AT INTERNATIONAL ACTORS' LEVEL	69			3		72	72
10.3a Does the intervention belong to PPA 3 - Education and peaceful democratic societies?	no (1), yes (4)	People enjoy equitable quality education and accountable governance in inclusive, peaceful and democratic societies. (SDG 1, 4, 16, 17) EDUCATION AND PEACEFUL DEMOCRATIC SOCIETIES INCLUSIVE OF GIRLS EDUCATION	55			17		72	72
10.3b Does the intervention belong to PPA3 - Outcome Area 1?	no (1), yes (4)	Outcome 1: Access to quality primary and secondary education has improved, especially for girls and for those in most vulnerable positions (SDG4, T1 and 5) ALL RELATED TO EDUCATION	65			7		72	72
10.3c Does the intervention belong to PPA3 - Outcome Area 2?	no (1), yes (4)	Outcome 2: More effective and accountable public sector with transparent and inequality reducing tax system (SDG16, T3 and T6; SDG17, T1) ALL RELATED TO PUBLIC SECTOR GOVERNANCE; FINANCIAL MANAGEMENT, TAXES	71			1		72	72
10.3d Does the intervention belong to PPA3 - Outcome Area 3?	no (1), yes (4)	Outcome 3: More peaceful, stable and just societies with strengthened political and judicial institutions and inclusive state-building processes (SDG 16) ALL RELATED TO PEACE, MEDIATION, STATE BUILDING	61			11		72	72
10.3e Does the intervention belong to PPA3 - Outcome Area 4?	no (1), yes (4)	Outcome 4: The enabling environment for and capacity of the civil society and persons in vulnerable positions to influence and participate in decision-making has improved (SDG5, T5; SDG16, T7) ALL RELATED TO CITIZEN'S PARTICIPATION IN PUBLIC AND POLITICAL DECISION MAKING	66			6		72	72
10.3f Does the intervention belong to PPA3 - Policy Influencing Area?	no (1), yes (4)	Policy Influencing: Increased commitment to the principles of democracy, rule of law and human rights, and the global commitments to inclusive education and taxation, at national and international level EXCLUSIVELY FOCUSED ON POLICY INFLUENCING, EXPECTED TO HAVE LITTLE IN OUR SAMPLE	68			4		72	72
10.4a Does the intervention belong to PPA 4 - Climate and natural resources?	no (1), yes (4)	Climate resilience and low greenhouse gas emissions development are promoted by sustainable use of natural resources. (SDGS 1, 2, 6, 7, 11, 12, 13 AND 15) INCLUSIVE OF WASH	56			16		72	72
10.4b Does the intervention belong to PPA4 - Outcome Area 1 - Forests and biodiversity?	no (1), yes (4)	Outcome 1: Forests and biodiversity - All people benefit increasingly from clean environment and healthy ecosystems, conservation, sustainable management and use of renewable natural resources, such as forests and water bodies (SDG 12.2, 15.1, 15.2, 15.3, 15.5, supports also SDG 6.5, 13.1, 13.3,15.9) ALL RELATED TO FORESTS AND BIODIVERSITY	62			10		72	72



Specification	Rating.1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
10.4c Does the intervention belong to PPA4 - Outcome Area 2 - Energy?	no (1), yes (4)	Outcome 2: Energy - All people have improved and equitable access to affordable and clean, sustainably produced renewable energy (SDG7, supports also SDG 13.1 and 13.3) ALL RELATED TO ENERGY	70			2		72	72
10.4d Does the intervention belong to PPA4 - Outcome Area 3 - Meteorology and disaster risk reduction?	no (1), yes (4)	Outcome 3: Meteorology and disaster risk reduction - The vulnerability of all people to extreme weather events and natural disasters has decreased and their resilience to them has increased (SDG 1.5, 11.5, 13.1., 13.2., 13.3.) ALL RELATED TO METEOROLOGY AND DISASTER RISK REDUCTION	69			3		72	72
10.4e Does the intervention belong to PPA4 - Outcome Area 4 - Food and nutrition security?	no (1), yes (4)	Outcome 4: Food and Nutrition Security - All people have improved possibilities to produce and access safe, nutritious, and adequate food (SDG2.1; supports also SDG 13.1 and 13.3) ALL RELATED TO FOOD AND NUTRITION SECURITY	71			1		72	72
10.4f Does the intervention belong to PPA4 - Outcome Area 5 - Water?	no (1), yes (4)	Outcome 5: Water -All people have improved and equitable access to basic and sustainable drinking water, adequate sanitation services, and improved hygiene practices (SDG 6.1-.6.2; supports also SDG 13.1 and 13.3) ALL RELATED TO WATER	69			3		72	72
10.5a Does the intervention belong to PPA 5 - Humanitarian Assistance?	no (1), yes (4)	Lives are saved, human suffering is alleviated and dignity maintained during crises (SDGS 2, 3, 4, 5, 11, 12, 14, 15 AND 16, 17) HUMANITARIAN ASSISTANCE	71			1		72	72
10.5b Does the intervention belong to PPA5 - Outcome Area 1?	no (1), yes (4)	Outcome 1: Civilian population has access to basic commodities, services and facilities (SDG 2, SDG 3, SDG 4, SDG 5, SDG 6,)	72					72	72
10.5c Does the intervention belong to PPA5 - Outcome Area 2?	no (1), yes (4)	Outcome 2: The protection of all people affected or threatened by a humanitarian crisis is assured (SDG 5, SDG 16)	70			2		72	72
10.5d Does the intervention belong to PPA5 - Outcome Area 3?	no (1), yes (4)	Outcome 3: Humanitarian coordination and coherence is improved (SDG 11, SDG 12, SDG 14, SDG 15, SDG 17)	72					72	72
7. Lessons learnt		PLEASE CODE LESSONS WITH CORRESPONDING TOPIC'S CODE OR UNDER OTHER IN MAXQDA. OECD-DAC defines a lesson as follows: "Generalisations based on evaluation experiences with projects, programs, or policies that abstract from the specific circumstances to broader situations. Frequently, lessons highlight strengths or weaknesses in preparation, design, and implementation that affect performance, outcome, and impact." NO NEED TO SEARCH FOR LESSONS IN RECOMMENDATIONS. HOWEVER, IF EXTRACTING A LESSON WITH REASONABLE LEVEL OF EXPERT JUDGEMENT, THIS IS OKAY. IF IT IS NOT POSSIBLE TO CONCLUDE WHAT THE LESSON WOULD BE OR IF ARBITRARY JUDGEMENT WOULD BE REQUIRED, IT IS NOT CONSIDERED.							
7_Planning									
7_Scope									
7_Time									
7_Financial									
7_Personnel									



Specification	Rating.1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
7_Capacity									
7_Equipment									
7_Participation									
7_Management									
7_Communication									
7_Exit strategy									
7_M&E									
7_Relevance									
7_Effectiveness									
7_Efficiency									
7_Impact									
7_Sustainability									
7_Aid effectiveness									
7_Complementarity									
7_Coherence									
7_Coordination									
7_Gender									
7_Others (not captured above)	specify in comment								
8. Recommendations		PLEASE CODE ALL RECOMMENDATION WITH CORRESPONDING TOPIC'S CODE OR UNDER OTHER.							
8_Planning									
8_Scope									
8_Time									
8_Financial									
8_Personnel									
8_Capacity									
8_Equipment									
8_Participation									
8_Management									
8_Communication									
8_Exit strategy									



Specification	Rating.1-4	Guidance	1	2	3	4	missing/not applicable/ no ToRs	Total	Total valid
8_M&E									
8_Relevance									
8_Effectiveness									
8_Efficiency									
8_Impact									
8_Sustainability									
8_Aid effectiveness									
8_Complementarity									
8_Coherence									
8_Coordination									
8_Gender									
8_Others (not captured above)	specify in comment								
sucessstory	no, yes	Do you have the impression that this project is a very good example for a very successful project? Then select yes.	8				64	72	8
	colour indicates additional aspects inserted								



# Annex 8: Interview Guideline

Name of Interviewer: *[Your name]*

Interview number: *[Please assign a case number here]*

Name of Interviewee: *[Interviewee's name]*

Position: *[e.g., Senior Advisor, Desk Officer, Ambassador, etc.]*

Department: *[e.g., Department for the Americas and Asia, Embassy of Finland in Myanmar]*

Unit: *[e.g., Unit for South Asia]*

Date: *[2021-M-D]*

## Instructions

- Prepare the **Basic Information Table** (p. 4) before conducting the interview
- Conduct the **interview**
- Invite to the **survey**

## INTRODUCTION

*[Recall purpose of the interview, thanks for finding time, explaining mixture of interview and survey questions].*

## CHECK AND INQUIRE

Let us first go through some basic information and check with you whether this is correct and if there is something missing. *[Go through the projects in the “Basic Information Table” (see p.4) one by one and note all information in the respective column.]*

**Check 0.0** Can you tell me your position, department, unit? *[See information requested in the header of this guideline.]*

**Check 0.1** In our database, you are listed as a contact for the following evaluations: *[Read from Table]*. Is this information correct? If not, please specify.

**Inquiry 0.1** For these, projects which would be the embassy AND implementer counterparts?





Check 0.2 [*Multi-projects-only, case numbers #8,#17,#21,#30, #35*] As you are involved in an evaluation that is a multiple projects evaluation: Do you see this as **one** evaluation project or as **several separate** projects? [*Mark in the table.*]

Thank you very much, then let's begin with the interview.

### GENERAL USE

The **first section** of this interview centres around your **general experiences of the use** of decentralised evaluations.

From your experience in general...

**Q1.3** How do you rate the **timing and timeliness** of decentralised evaluations, and do you have any **suggestions to improve** them?

**Q1.6** From your experience, in general, how do you rate the relevance and applicability of the recommendations of decentralised evaluations, and do you have any **suggestions to improve the implementation** of relevant and realistic recommendations?

**Q1.8** From your experience, in general, do you have any **suggestions to improve learning** from decentralised evaluations in general?

**Q1.10** From your experience, in general, which **factors typically facilitate useful decentralised evaluations**?

**Q1.12** From your experience, in general, which **factors typically hamper useful decentralised evaluations**?

### OVERVIEW: EXAMPLE PROJECT

The **second section** of this interview centres around **the actual use of one specific project**. Which project on the list would you like to talk about? [*Let the interviewee decide which project should serve as an example here. It should be a project in which they really made use of the evaluation report. If they don't have a preference, choose the most recent project from the list.*]

Okay, so let's look at the example project **IPP** specifically:

**Q1.1.1** How did it go? What is your general impression? How do you feel about the evaluation?

**Q1.1.2** For which purpose did you use the evaluation report?

### RECOMMENDATIONS AND LEARNINGS: EXAMPLE PROJECT

Let us now look at the **recommendations by the evaluators, their implementation, and the learnings** for the example project [*NAME*].

**Q1.5** Can you please elaborate **on the recommendations** of this evaluation:



**Q 1.5.1** In which **intervention fields** have recommendations been implemented?

**Q1.5.2** Which influence did **the MFA implementation decision** (management response) and their **follow-up** have on the implementation of the recommendations?

**Q1.7.1** Can you please elaborate **on the learnings** of this evaluation: **What** did you learn, and **how** did the learning take place?

**Q1.7.2** **Who else** learned, and **how** did **dissemination** of learnings occur?

**Q1.7.3** And is there any learning with respect to the **COVID-19 pandemic**? *[ask only if applicable]*

### **WRAP UP**

Thank you very much for sharing these in-depth insights. There are a few more standardized questions to answer for this project *[NAME]*. Therefore, I will send you the link to a survey in a minute.

**Qo.3** Is there **anything you want to add (in general or for a specific project)**?

### **INVITATION TO THE SURVEY**

Thank you again for participating in this interview. We highly appreciate the time you spent supporting the metaevaluation.

I will now send you the link to our follow-up survey in the chat. The survey contains some standardized questions concerning the evaluation projects (yes/no, multiple choice, and few rating questions). [survey link: <https://ofb.ceval.de/surveymetaevaluation/>]

Please, click on the link and fill out the survey for each project that is on your list. In your case, this is *[number of projects]*. *[Copy project titles and codes to the chat if needed.]*

*[Multi projects: Remind interviewee when he or she wanted to treat the multi-project as “several” to also fill out the survey several time].*

In the survey you will be asked to enter the name of the project and the intervention code. Please, enter this information as we will need it to match your answers with our database. Do you have any further questions? *[If interviewee seems unsure, offer to stay in the chat to answer any questions while the interviewee is initiating the survey on ofb. Make sure that they have the **intervention codes** with the associated **project titles!**]*



## BASIC INFORMATION TABLE

INFORMATION FROM DATABASE					CHECK 0.1		INQUIRY 0.1		CHECK 0.2 [ONLY MULTI PROJECTS]
Case_Id	Project	Intervention Code	Year	Role	Correct y/n	No, Comment	Embassy contact counterpart	Project implementers	one/ several



# Annex 9: Survey Questions and Responses

QUESTION	ANSWER OPTIONS				
<b>Q1. What was your role in this project?</b>		No	yes		
	MFA Headquarters		49		
	MFA Embassy		29		
	Implementer		41		
	Other		0		
	Total		119		
<b>Q2: For which purpose did you use the evaluation?</b> <b>(* = not included in the implementers survey)</b>		no	yes	missing	Total
	I did not use this evaluation report	11	96	12	119
	Adaptive management of ongoing intervention (project or programme level)		39		
	Decision making on potential follow-up intervention (project or programme level)		46		
	Planning of follow-up intervention (project or programme level)		41		
	Designing new projects of Fin DevCo (no direct follow-up)*		8		
	Designing new country strategy of Fin DevCo (strategy understood as plan of action)*		7		
	Designing new sector strategy of Fin DevCo (strategy understood as plan of action)*		3		
	Informing policy development (policy understood as a principle of action)*		12		
	Individual learning after rotation		10		
	Learning in the project team		35		
	Learning in the country team		26		
	Learning in the sector team		9		
	Facilitating learning among donor community		12		
	Facilitating learning among partner countries		8		
	Facilitating learning among target groups		10		
Informing country analyses (e.g. as part of systematic review)		9			
Informing sector analyses (e.g. as part of systematic review)		9			
Raising awareness of Fin DevCo*		1			



QUESTION	ANSWER OPTIONS				
	Justifying funding decisions		23		
	Other		11		
<b>Q3: How do you assess...</b>		1= inadequate	2= need for improvement	3= satisfactory	4= (very) good
	...the overall usefulness of this evaluation?	2	2	44	32
<b>(4-Point Likert-Scale, from 1=inadequate to 4=very good, selection option="I don't know")</b>	...the overall quality of this evaluation report?	2	2	46	30
	...the overall quality of the corresponding executive summary?	2	1	45	27
	...the overall quality of this evaluation process?	2	8	43	21
	...the appropriateness of the evaluation team composition?	2	4	37	27
	...the overall quality of the corresponding ToR?	1	2	35	27
	...the appropriateness of the ToR with respect to (MFA HQ/ embassy/ implementer's) needs?	2	3	32	30
	...the overall quality of the intervention?	1	5	28	19
<b>Q3: How do you assess...</b>		1= inadequate	2= need for improvement	3= satisfactory	4= very good
	...the timing for this evaluation?	1	11	28	40
	...the timeliness of informal results delivery (FCR workshop, validation meeting) against (MFA HQ/ embassy/ implementer's) needs?	2	6	41	25
	...the timeliness of written, formal results delivery (approved report) against (MFA HQ/ embassy/ implementer's) needs?	1	5	36	31
<b>Q5: How do you assess...</b>		1= inadequate	2= need for improvement	3= satisfactory	4= (very) good
	...the relevance of the recommendations of this evaluation for (MFA HQ/ embassy/ implementer's) staff?	2	9	35	34
	...the recommendations being realistic?	3	6	47	23
	...the implementation of relevant and realistic recommendations?	2	8	43	19
	...the learnings from this evaluation?	2	6	39	29
<b>Q6: Did any of the following factors particularly facilitate the usefulness of this evaluation?</b>		no	yes	missing	Total
	I do not know	13	85	21	119
	Being a mid-term evaluation?		34		
	Being a final evaluation?		39		
	Inclusion of all DAC criteria?		12		
	Focus on specific DAC criteria?		7		
	Implementation as a rapid assessment?		6		



QUESTION	ANSWER OPTIONS				
	Implementation as an in-depth assessment?		9		
	Providing sufficient calendar time (anticipating of delays)?		9		
	Methodological competence of the evaluation team?		13		
	Technical competence of the evaluation team?		24		
	Appropriateness of ToR?		27		
	Appropriateness of evaluation budget (good cost-benefit ratio)?		3		
	Integrating diverse stakeholder groups (commissioners, implementers, target groups) in the inception phase?		10		
	Integrating diverse stakeholder groups (commissioner, implementers, target groups) in results validation?		10		
	High-quality management response?		17		
	Follow-up of the management response?		25		
	Others, to be added?		7		
		Total	252		
<b>Q7: Did any of the following factors particularly hamper the usefulness of this evaluation?</b>	I do not know	30	64	25	119
	Evaluation capacity gaps at implementers' level?		6		
	Evaluation capacity gaps at evaluators' level?		6		
	Evaluation capacity gaps at MFA level?*		4		
	Insufficient financial resources (evaluation budget)?		4		
	Insufficient time resources (deadlines inappropriate)?		13		
	Lack of high-quality M&E systems at intervention level (lack of baseline, end-line data)?		10		
	Covid-19 pandemic related travel restriction?		20		
	Lack of high-quality ToR?		3		
	Lack of coordination between MFA HQ, embassies, and implementers?		8		
	Lack of ownership?		4		
	Lack of leadership?		1		
	Lack of prioritisation?		3		
	Lack of management response?		5		
	Lack of follow-up of management response?		6		
	Lack of follow-up of management response?		6		
	Others, to be added?		18		
		Total	111		



## Annex 10: List of Evaluation Reports Received and Used

N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMISSIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALUATION
										yes	no			
1	2018	final	ILO decent work for women Phase II	ALI-10	MENA	Government and civil society	3.122.023	3.122.023		yes	no	yes		yes
2	2020	final	Somaliland Health Sector Support (MIDA Finnsom IV)	ALI-20	Africa	Reproductive Healthcare	4.300.000	4.300.000		no	no	yes		yes
3	2020	mid-term	General Education Quality Improvement Programme in Ethiopia	ALI-20	Africa	Education	850.000	850.000		no	yes	yes		yes
4	2020	final	Support Afghanistan Livelihoods and Mobility (UNDP/SALAM)	ASA-40	MENA	Other social services	4.500.000	4.632.527		no	no	yes		yes
5	2020	mid-term	Responsible and Innovative Land Administration II	ALI-20	Eastern Europe and Central Asia	Agriculture	7.100.000	8.800.000	59.000	no	yes	yes		yes
6	2019	mid-term	Support to Nepal's School Sector Development Plan (SSDP)	ASA-40	Latin America	Education	20.000.000	6.824.000.000		no	no	yes		yes
7	2018	mid-term	International Commission of Jurists	POL-40	Global	Conflict prevention, resolution, peace and security	1.370.000	1.370.000	20.000	no	no	no		yes
8a	2020	final	Integrated Land Use Assessment phase 2 (ILUA II)	ALI-31	Africa	Forestry	4.581.200	4.581.200	19.667	no	yes	yes		yes
8b	2020	final	Civil Society Environment Fund phase 2 (CSEF2)	ALI-32	Africa	Environment/Climate	4.600.000	4.600.000	19.667	no	yes	yes		yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMISSIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALUATION
										yes	no			
8c	2020	final	Decentralised Forest and other Natural Resources Programme (DFONRMP)	ALI-33	Africa	Environment/Climate	4.221.314	4.700.000	19.667	no	yes	yes		yes
9	2019	mid-term	Agro Business Induced Growth in Amhara Region (AgroBIG II)	ALI-20	Africa	Agriculture	9.400.000	10.340.000		yes	yes	yes		yes
10	2019	mid-term	Scaling-up Research and Capacity Building for Improved Development Policy in Mozambique	ALI-30	Africa	Business support services	2.498.212	3.727.827	80.000	no	yes	yes		yes
11	2019	final	Strengthening democratic institutions in the governance of natural resources	ALI-30	Africa	Government and civil society	1.500.000	1.523.483	80.000	no	yes	yes		yes
12	2019	mid-term	WCO-ESA Project "To Progress the Trade Facilitation Agenda, within the Framework of the WCO Mercator Programme"	TUO-10 & ALI-30	Africa	Business support services	3.000.000	3.000.000	59.900	yes	yes	yes		yes
13	2020	final	Rural Electrification Project	KEO-50	Latin America	Energy		33.791.520		no	yes	yes		yes
14			UNESCO: Strengthening Teacher Education in Myanmar									No		no
15	2018	final	IDEA Myanmar Constitution Centre (MyConstitution) Programme	ASA-40	Asia	Government and civil society				yes	no	no		yes
16	2019	mid-term	Rural Village Water Resources Management Program phase III	ASA-40	Asia	Water and sanitation	15.000.000		120.000	no	yes	yes		yes





N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMISSIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALUATION
										no	yes			
17a	2019	final	Participatory Sustainable Forest Management Project (SUFORD-SF, 74502102)	ASA-10	Asia	Forestry	7.974.315	35.422.014	50.750	no	yes	yes		yes
17b	2019	final	Strengthening National Geographic Services in Lao PDR (SNGS)	ASA-10	Asia	Forestry	6.000.000	6.500.000	50.750	no	yes	yes		yes
18	2019	final	Water and Sanitation Programme for Small Towns (WSPST)	ASA-10	Asia	Water and sanitation	32.707.000	38.918.000	130.000	no	yes	yes		yes
19	2019	final	Innovation Partnership Programme II	ASA-10	Asia	Innovation	9.900.000	11.000.000	90.000	no	yes	yes		yes
20	2019	final	Support to the Mekong River Commission + IUCN	ASA-10	Asia	Water and sanitation	211.443.258	114.853.050		no	yes	yes		For quality assessment only
21a	2019	final	Development of Management Information System for the Forestry Sector (FORMIS – Phase II)	ASA-11	Asia	Forestry	9.700.000	10.137.540	31.667	no	yes	yes		yes
21b	2019	final	People Participation in Improvement of Forest Governance and Poverty Alleviation in Vietnam (PFG)	ASA-12	Asia	Government and civil society	1.049.652	1.095.755	31.667	no	yes	yes		yes
22			Education Sector Project											no
23	2018	final	Coastal Surveillance System	KEO-50	Africa	Conflict prevention, resolution, peace and security	17.700.000	17.700.000	31.667	no	yes	yes		yes
24	2018	final	Adapting to climate change in Oceania, FINPAC	ASA-10	Global	Environment/Climate	500.000	3.700.000	100.000	no	yes	yes		yes
25	2018	mid-term	Afghanistan Independent Human Rights Commission	ASA-40	Eastern Europe and Central Asia	Government and civil society				no	no	yes		yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMISSIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALUATION
										yes	no			
26	2018	mid-term	LOTFA: the Law and Order Trust Fund for Afghanistan Support to Payroll Management Project	ASA-40	Eastern Europe and Central Asia	Conflict prevention, resolution, peace and security		1.146.084.819		yes	no	yes		yes
27	2018	final	UN Women Afghanistan	ASA-40	Eastern Europe and Central Asia	Unallocated/unspecified	600.000	38.195.943	124.878	no	no	yes		yes
28	2019	final	UN Women Supporting Women in Host Communities and in Jordan's Refugee Camps	ALI-10	MENA	Conflict prevention, resolution, peace and security		2.534.861		no	no	yes		yes
29	2019	final	UN Women Women, peace and security in the Arab States Programme Phase I	ALI-10	MENA	Conflict prevention, resolution, peace and security	3.204.925	3.204.925		no	no	yes		yes
30a	2018	final	Promoting Modernisation of Hydro-meteorological Services in Vietnam (PROMOSERV I and II)	ASA-10	Asia	Unallocated/unspecified	1.901.997	1.901.997	26,666,67	no	yes	yes		yes
30b	2018	final	Capacity Building for the Development of Selective Breeding Programs in Vietnam	ASA-10	Asia	Agriculture	1.103.834	1.103.834	26,666,67	no	yes	yes		yes
30c	2018	final	Developing and Implementing Climate Change Adaptation Measures at Local Level in Vietnam	ASA-10	Asia	Environment/Climate	1.973.000	1.973.000	26,666,67	no	yes	yes		yes
31	2019	final	UNDP/ Evaluation of Aid for Trade project phase III	ASA-40	Eastern Europe and Central Asia	Trade policy and regulation		9.032.000		no	no	yes		yes
32	2018	mid-term	MIDA Finnsom Education and Health South Central Somalia	ASA-40	Africa	Education	4.200.000			no	no	no		yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMIS- SIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALU- ATION
										yes	no			
33	2020	mid-term	Finland Ukraine Trust Fund at NEFCO	ASA-40	Eastern Europe and Central Asia	Energy	6.000.000	6.000.000		yes	no	yes		yes
34	2020	final	Syria Initiative Project (FELM)	ALI-10	MENA	Conflict prevention, resolution, peace and security				no	no	no		For quality assessment only
35a	2017	final	The Integrated Environmental and Forest Management Co-operation Project in Central America (Finfor-II)	ASA-30	Latin America	Forestry		3.800.000	86.667	no	yes	yes		yes
35b	2017	final	The Sustainable Forest Management (MFS) Programme	ASA-30	Latin America	Forestry	6.250.000	2.000.000	86.667	no	yes	yes		yes
35c	2017	final	Livelihood Improvement through Generation and Ownership of Forest Information by Local People	ASA-10	Asia	Forestry	2.000.000	8.080.000	86.667	no	yes	yes		yes
36	2018	final	UNAOC's Youth Solidarity Fund Projects	POL-50	Global	Government and civil society	215.053	215.053		yes	no	yes		For quality assessment only
37	2020	mid-term	UNDP / Strengthening Rule of Law and Human Rights to Empower People in Tajikistan, Phase II	ITÄ-20	MENA	Conflict prevention, resolution, peace and security				yes	no	yes		yes
38	2018	mid-term	IGAD Institutional Strengthening Action Programme (ISAP III)	ALI-20	Africa	Government and civil society	1.500.000	13.858.939		no	no	yes		For quality assessment only
39	2020	final	Independent Evaluation of IDA's Disability Catalyst Programme	POL-40	Global	Reproductive Healthcare	1.500.000		95.092	no	no	yes		yes
40	2018	mid-term	Sustainable Energy and Climate Change Multi-donor Trust Fund (MSC), SECCI	KEO-50	Latin America	Environment/ Climate		131.710.785		no	no	no		yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMIS- SIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALU- ATION
41	2019	mid-term	World Bank Multi-Donor Trust Fund for the Middle East and North Africa Region	ALI-10	MENA	Conflict prevention, resolution, peace and security		10.607.580		no	no	yes		yes
42	2020	final	MSI - Reproductive Health Programme	ASA-40	Eastern Europe and Central Asia	Reproductive Healthcare			10.000	no	no	yes		yes
43	2019	final	UNESCO/ Enhancement of Literacy in Afghanistan (ELA3)	ASA-40	Eastern Europe and Central Asia	Education	3.000.000	27.006.323		yes	no	no		yes
44	2018	final	Women and Girls First-programme	ASA-40	Asia	Reproductive Healthcare	7.300.000	9.138.224		yes	no	yes		yes
45	2019	mid-term	Training of Chemical Weapons Verification project (VERIFIN)	POL-20	Global	Conflict prevention, resolution, peace and security		12.657.782	35.000	no	yes	yes		For quality assessment only
46	2019	final	Technical Assistance for Competence based Soft Skills Development in School Education	ASA-40	Asia	Education	1.700.000	1.700.000		no	yes	yes		yes
47	2018	mid-term	ILO/ Decent Work Phase II	ITÄ-20	Eastern Europe and Central Asia	Conflict prevention, resolution, peace and security	4.000.000	4.000.000	21.000	yes	no	yes		yes
48	2020	mid-term	Learning Together Project Ukraine	ITÄ-20	Eastern Europe and Central Asia	Education	6.000.000	8.000.000	50.000	yes	yes	yes		yes
49	2018	mid-term	Developmental Performance Evaluation of the Network For Religious and Traditional Peacemakers	POL-60	Global	Conflict prevention, resolution, peace and security	2.700.000			yes	yes	yes		yes
50	2017	mid-term	Strengthening Health Security and Biosecurity in Tanzania by Biodetection Capacity Building	POL-20	Africa	Conflict prevention, resolution, peace and security				no	no	yes		yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMISSIONED	INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALUATION
51			External Review of Support to the Mozambique Revenue Authority										no
52	2018	final	Action Plan for Ukraine	ITÄ-20	Eastern Europe and Central Asia	Government and civil society	1.000.000	26.600.000	24.000	no	no	yes	yes
53	2017	mid-term	Euro-Burma Office in Myanmar (EBO/FELM)	ASA-40	Asia	Conflict prevention, resolution, peace and security		3.398.441		no	no	yes	yes
54	2018	final	Evaluation of Finnish Support to IESE (2016-2018)	ALI-30	Africa	Government and civil society	1.500.000	7.000.000		yes	yes	yes	yes
55	2018	final	School to Change the World Continuing Education Project	VIE-30	Global	Education			4.000	no	no	yes	For quality assessment only
56	2017	final	Development Policy Induction Programme (the "Influencer Programme")	VIE-30	Global	Unallocated/unspecified				yes	yes	yes	For quality assessment only
57	2017	mid-term	Support to Community-Led Accelerated WASH in Ethiopia (COWASH III)	ALI-20	Africa	Water and sanitation	14.100.000	41.400.000	59.000	yes	yes	yes	yes
58	2018	final	Cooperation in the Development of Information and Communication Technologies in Africa – Phase II	ALI-20	Africa	Communications/ICT	771.000	3.745.947	15.000	no	no	yes	yes
59	2019	final	Energy and Environment Partnership Programme Mekong	ASA-10	Asia	Energy	8.695.000	14.765.754		no	yes	yes	yes
60	2020	final	Support to National and Local Level Reconciliation in Somalia	ALI-20	Africa	Conflict prevention, resolution, peace and security	1.255.595	1.255.595		no	no	yes	yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMISSIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALUATION
										yes	no			
61	2017	mid-term	The Finnish - Southern African Partnership Programme to Strengthen NEPAD/ SANBio Network (BioFISA II)	ALI-30	Africa	Unallocated/ unspecified	6.000.000	7.820.000	30.000	yes	yes	yes		yes
62	2018	mid-term	Water Sector Trust Fund (WSTF)	ALI-20	Africa	Water and sanitation		17.300.000	50.000	yes	yes	yes		yes
63	2019	final	Social Protection/ United Nations Joint Protection Program	ALI-30	Africa	Conflict prevention, resolution, peace and security	9.770.000	13.600.000		yes	no	yes		yes
64	2017	final	ECDC: Early Intervention for Children with Disabilities (PYM)	ASA-40	Asia	Education	120.000	120.000		yes	no	no		yes
65	2017	final	UNDP: Strengthening the Rule of Law and Human Rights Protection System in Nepal Programme	ASA-40	Asia	Conflict prevention, resolution, peace and security	5.400.000	17.655.820		yes	no	yes		yes
66	2017	mid-term	UN Women: Women Economic Empowerment	ASA-40	Asia	Government and civil society	4.000.000	4.000.000		no	no	yes		yes
67	2017	mid-term	Digital Defenders Partnership (HIVOS)	POL-40	Global	Conflict prevention, resolution, peace and security		16.000.000	50.000	no	no	yes		yes
68	2017	mid-term	UN Environment-ILO-UNDP-UNIDO-UNITAR project 'Partnership for Action on Green Economy (PAGE)	KEO-90	Global	Environment/ Climate	1.645.725	23.906.039		yes	no	yes		yes
69	2018	mid-term	UNODC: Alternative Development & Community Forest Project in Southern Shan State	ASA-40	Asia	Conflict prevention, resolution, peace and security	3.000.000	5.715.742	46.484	no	no	yes		yes



N°	YEAR OF REPORT	EVALUATION TYPE	REPORT TITLE	MFA UNIT	REGION	SECTOR	PROJECT BUDGET (EUR) BY FINLAND	PROJECT BUDGET (EUR) OVERALL	EVALUATION BUDGET (EUR), VAT EXCLUDED	MFA COMMIS-SIONED		INDEPENDENT CONSULTANT	TOR AVAILABLE	USED FOR META-EVALU-ATION
70	2020	final	Phase II of the Strengthening Pre-Service Teacher Education in Myanmar (STEM) Project	ASA-40	Asia	Education	4.320.000	3.428.991		no	no	yes		yes
71			WB-Decentralized Funding to Schools											no
72	2017	final	UNICEF Afghanistan: Wash in Schools (WinS) Programme	ASA-40	Eastern Europe and Central Asia	Education				no	no	yes		yes
73			Regional Programme for Afghanistan and Neighbouring Countries											no
74	2017	mid-term	Afghanistan Reconstruction Trust Fund (World Bank)	ASA-40	Eastern Europe and Central Asia	Business support services		9.266.208.000		no	no	yes		For quality assessment only
75	2018	mid-term	WTO Standards and Trade Development Facility (STDF)	TUO-10	Global	Trade policy and regulation	2.800.000	18.781.150		no	no	yes		yes
76	2018	final	Zambia Green Jobs Programme	ALI-30	Africa	Business support services	10.000.000	10.000.000		yes	no	yes		yes
77	2019	mid-term	IOM: Fostering Health and Protection to Vulnerable Migrants Transiting Through Morocco, Tunisia, Egypt, Libya and Yemen	ALI-10	MENA	Other social services	2.726.954	2.726.954	17.000	yes	no	yes		yes

Note: 5 reports were not used due to following reasons: duplication (2 reports), aide memoire, no Finnish DAC, sector evaluation.



# Annex 11: Quality Assessment on single sections of the reports

## Quality of introductions and context analyses

### Highlights of the section:

- A bit less than half of the introductions is rated as satisfactory, a bit more than a third (very) good.
- Two cases are inadequate, and the remainder with a need for improvement.
- In 80% of the cases, context analysis is satisfactory or better; in 20%, there is a need for improvement, and one report is inadequate.
- However, one-fifth of the cases do not contain a context analysis at all.
- There are no significant changes to the previous metaevaluation.
- As in the previous meta, the lack of acknowledgement of CCOs is the most serious issue.
- Lack of contextualising Finnish development policies not only in non-MFA reports missing.

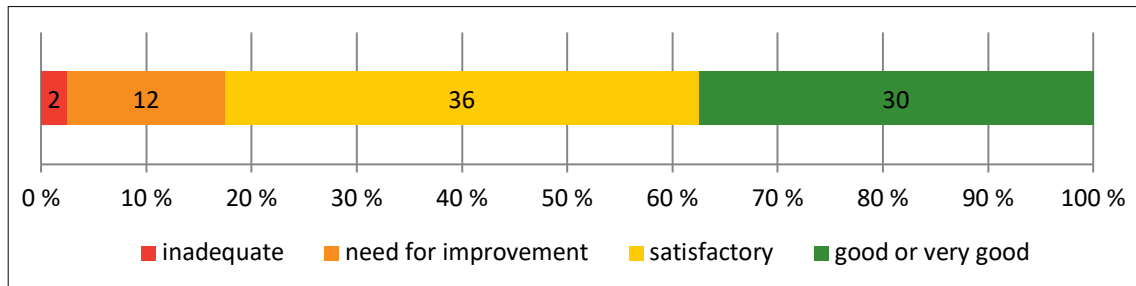
Different aspects of the introduction were analysed, which includes (i) **rationale and purpose**, (ii) **objectives of the evaluation**, (iii) **evaluation object**, (iv) **scope of the evaluation**, (v) **evaluation questions** and if any, (vi) **results of previous evaluations**. Each of these aspects is further broken down into sub-aspects and are listed in detail in annex 6.

Regarding the overall rating of introductions, shown in Figure 33, more than a third of the reports were (very) good, while almost half were rated as satisfactory. Two reports were considered inadequate, and the rest had introductions needing improvement. Not much difference was found compared to the previous metaevaluation, as both findings show that more than 80% of the reports rated satisfactory and (very) good in providing general information on the intervention and the evaluation conducted.





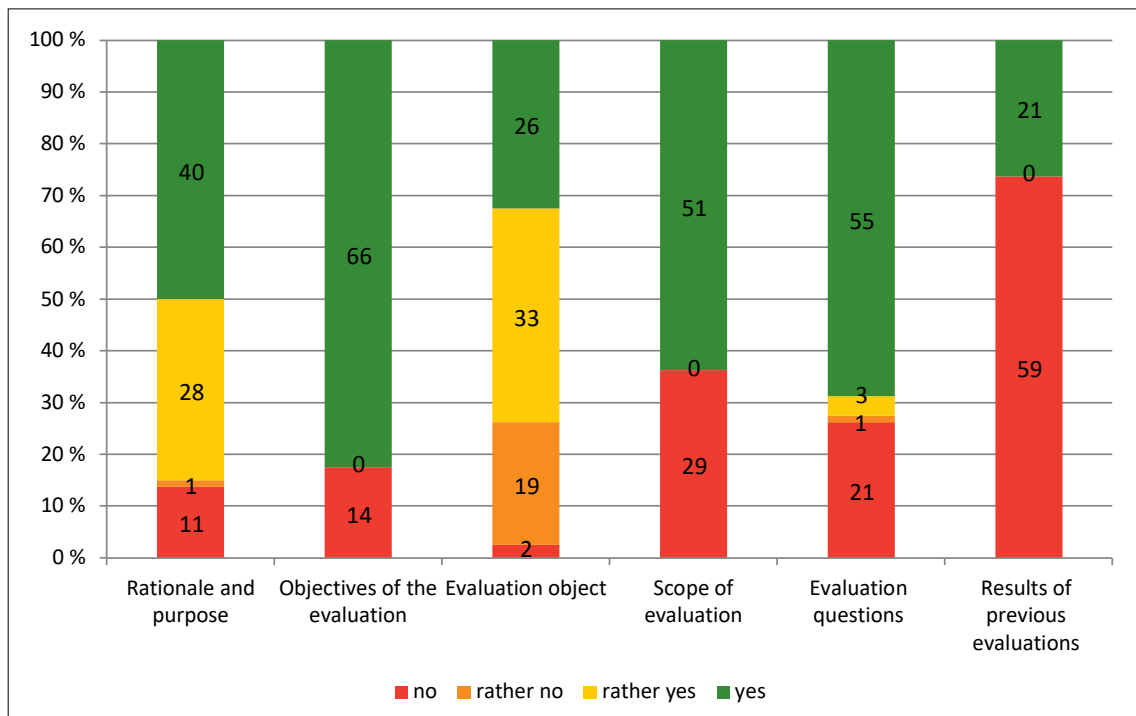
**Figure 33: Overall rating of introductions (n=80)**



Source: own statistics based on analysis of reports

Figure 34 demonstrates how well different aspects of an introduction were captured. Most reports (68,80%) stated the rationale and purpose satisfactorily or better, while more than 80% (66) described the objectives of the evaluation in the introduction. Also, the scope of evaluation and evaluation questions were addressed in most introductions satisfactorily (59,74% and 51,64%, respectively). A complete description of the evaluation object was in three-quarters of the cases (56,70%) satisfactorily or better, but only roughly a quarter (21,26%) of the introductions acknowledged results of previous evaluations.

**Figure 34: Contents of introduction (n=80)**

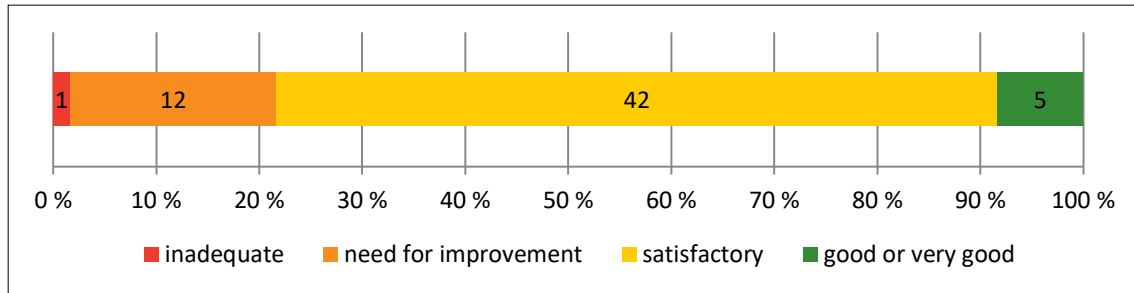


Source: own statistics based on analysis of reports

The overall rating of context analysis revealed that 20 out of 80 reports (25%) had not provided any context analysis. However, 80% of the cases with context analysis were rated as satisfactory or better, with 42 (53%) satisfactory and 5 (7%) (very) good, as displayed in Figure 35. Meanwhile, 20% (12) need improvement, and one report was rated inadequate.



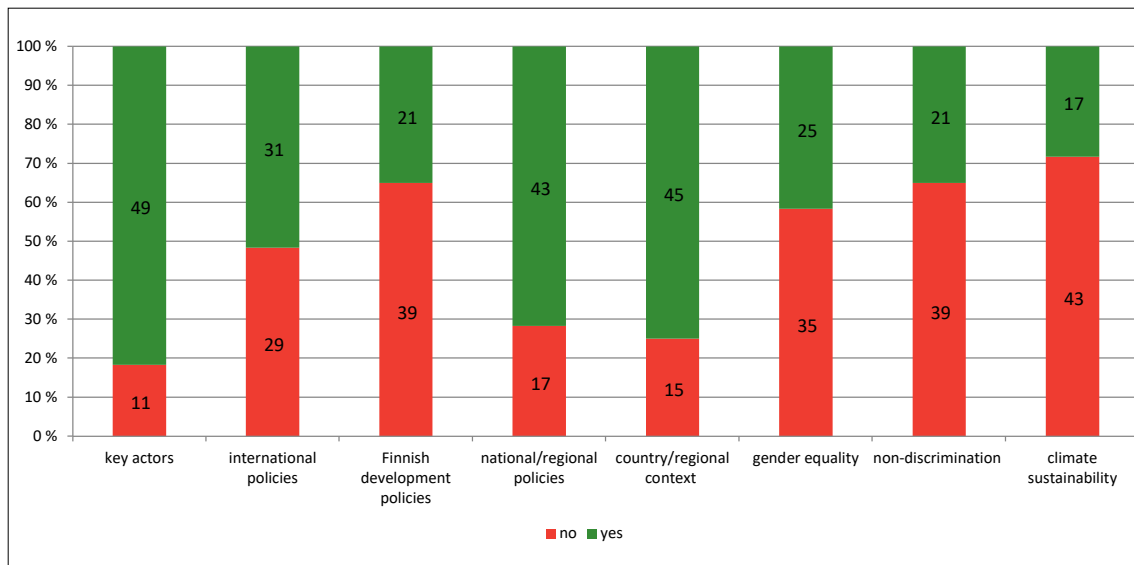
**Figure 35: Overall rating of context analysis (n=60)**



Source: own statistics based on analysis of reports

Figure 36 below illustrates the eight aspects rated to assess the overall quality of context analyses. Discussion of (i) **key actors** was most frequently discussed with 49 out of 60 cases (82%), followed by (ii) **country/regional context** with 45 out of 60 cases (75%), (iii) **national/regional policies** with 43 out of 60 cases (72%), and (iv) **international policies** in 31 out of 60 cases (52%). Meanwhile, less than half of the reports discussed (v) **gender equality** in the content analysis (25 out of 60, 42%), only a little more than third addressed (vi) **non-discrimination** and (vii) **Finnish development policies** (both 21 out of 60, 35%) and lastly, **climate sustainability** acknowledged only in 17 out of 60 cases (28%). The biggest share of reports captured all aspects (43 out of 60, 72%) or most parts (13 out of 60, 22%) of those aspects (see Annex 6).

**Figure 36: Contents of context analysis (n=60)**



Source: own statistics based on analysis of reports

A brief comparison with the previous metaevaluation did not reveal major changes. Similarly, as earlier, acknowledgement of CCOs is the most serious issue, and contextualisation of Finnish development policies is an issue.



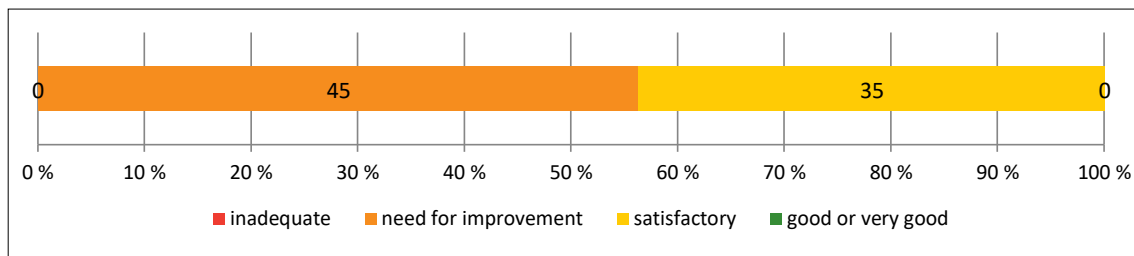
## Quality of evaluation methodologies

### Highlights of the section:

- More than half of the reports need improvements, while the remainder is satisfactory.
- Major room for improvements exists regarding informative sample description, reflections on limitations and presentation of evaluation designs.
- There are no significant changes to the previous metaevaluation.

Looking at the overall quality of evaluation methodologies as presented in Figure 37, less than half (35, 44%) are rated satisfactory and more than half (45, 56%) are assessed as needing improvement.

Figure 37: Overall rating on methodology (n=80)



Source: own statistics based on analysis of report

Figure 38 compares the assessment results of the six aspects considered in the analysis of the evaluation methodologies. According to the analysis, only a little more than half of the reports (44, 55%) **presented (i) the evaluation design** satisfactorily or better. A closer look at the sub-aspects (see Annex 6) reveals that only 11 (14%) presented an evaluation design, while only 43 (54%) describe the general evaluation approach. The rest of the reports did not describe the evaluation design or approach at all.

The (ii) **sources of evidence** fared better in that none of the reports were rated as inadequate, and about 94% (75) were rated satisfactorily or better. The same percentage was also rated as having at least a short description of sources of information, while 5 reports (7%) do not have this information. A closer look at the sub-aspects (see Annex 6) reveals that almost all cases use project documents (79, 99%), implementing organisation (77, 96%) and beneficiaries (71, 89%) as sources of information for the evaluation. M&E data and institutional environment are used in more than 60% of the cases. Consequently, the mix of sources of information was assessed as mostly completely appropriate (66, 83%) and rather appropriate (11, 14%).

Only one case was rated (very) good when it comes to describing (iii) **data collection** in the report, half of the report (41, 51%) is rated as satisfactory, and the other half (38, 48%) needs improvement. A closer look at the data collection methods used reveals that all cases have conducted interviews, and most of the evaluators (71, 89%) applied two or more data collection techniques. 61% of the sample (49) have used focused group discussion, and only 39% (31) have conducted surveys.

Regarding (iv) **sampling**, almost half of the reports (36, 45%) were rated inadequate in providing information on the sample. 22 out of 80 (28%) had no information at all on the sampling, only



14 (18%) were rated as having provided complete information, and the rest were either brief and incomplete (24,30%) or moderate but incomplete (20,25%). Furthermore, most of the reports (61,76%) do not justify the chosen sampling strategy and therefore fail to provide any information on why their selection of information sources was appropriate. Nevertheless, almost three-quarters of the reports were at least able to acknowledge all groups of stakeholders in their data collection.

(v) **Data analysis methods** are another aspect considered in analysing the quality of evaluation methodologies. About 61% (49) of the reports were rated satisfactory or better, and more than a third (28, 35%) were found to need improvement. However, a little more than half (42, 53%) have not described in their report which methods were used for the data analysis, while only 3 reports (4%) provided complete transparency in this regard.

Lastly, the description of (vi) **limitations and challenges** were rated inadequate in a quarter of the reports (20,25%), 35 reports (44%) need improvement and another quarter (19,24%) were assessed as satisfactory. Only 6 reports (8%) emerged as (very) good in discussing different sub-aspects of limitations and challenges in conducting the evaluation. As listed in annex 6, most of the limitations described are regarding data collection (59, 74%) and evaluation process (42,53%), while only 9 reports (11%) discussed limitations relating to data analysis methods. Furthermore, only 13 reports (16%) discussed possible influences of the limitations on the evaluation.

To sum up, aspects relating to evaluation design, information on the sampling strategy and reflections on limitations and challenges were assessed as weak.

**Figure 38: Description and appropriation of methods (n=80)**



Source: own statistics based on analysis of reports

A brief comparison with the previous metaevaluation did not reveal major changes.



## Quality of evaluation findings

### Highlights of the sections:

- Only half to reports mostly link findings to evidence.
- Confounding factors and attribution are rarely discussed.
- Quality of finding remains a major constructions site.
- Overall, there is no significant change to the previous metaevaluation. However, triangulation of findings and the evidence-based findings were rarer.

Three angles were considered in analysing the quality of evaluation findings, namely an assessment on (i) how the findings were obtained, (ii) whether the intervention logic was discussed/reviewed, and (iii) whether the different OECD DAC criteria were appropriately captured. Taking these aspects together, we find a great need for improvement in the overall quality of evaluation findings. The analysis of findings (see Annex 6) disclosed need improvement for the majority of the cases (51,70%), and 3 reports (4%) were assessed as inadequate. Only 1 report received a (very) good rating, and the remaining quarter (20,25%) was satisfactory.

To assess the quality of how the findings were obtained, the following aspects were analysed (i) **findings are evidence-based**, (ii) **findings are mostly linked to evidence** (iii) **triangulation of data** (iv) **only findings are found in section** (i.e. no mix with conclusions and recommendations, (v) **discussion of attribution**, and (vi) **discussion of confounding factors** (as displayed in Figure 39).

On a more general assessment that determines if findings refer clearly to the data collected, 73 reports (91%) were considered **somewhat evidence-based**. However, on a more differentiated assessment on whether **most findings are linked to evidence**, only a little more than half (41, 51%) were found to specify the source of information to a satisfactory level (37, 46%) or a (very) good level (4, 5%) when reporting findings.

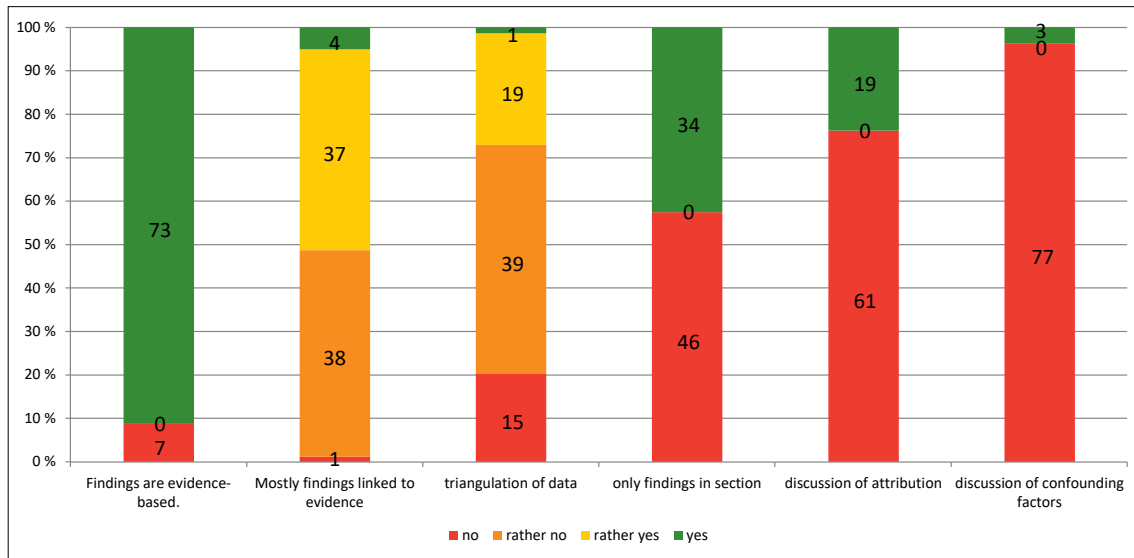
**Triangulation of data** was rather seldomly seen in almost half of the cases (39, 49%), while findings were not evidence-based in 15 reports (19%). Only one report was found to put the vast majority of the results into perspective, while 19 reports (24%) were found to do so to a satisfactory extent.

Less than half of the reports (34, 43%) presented **only findings in the section**. More than half (46, 57%) do not separate findings from conclusions and recommendations.

Finally, attribution of the intervention to observed results and confounding factors were rarely discussed in the findings. Only a quarter of the reports (19, 24%) discussed whether the intervention results could be attributed to the intervention. Meanwhile, only 3 reports (2%) were found to discuss further whether results are examined in consideration of confounding factors.



**Figure 39: Quality of findings (n=80)**



Source: own statistics based on analysis of reports

The intervention logic (i.e., the programme theory, logical framework, results model etc.) is important to structure the effectiveness and impact analyses of the evaluated interventions; thus, it was considered in assessing the quality regarding evaluation findings. The sub-aspects comprised of (i) whether the logical framework was described, (ii) whether the results model was provided, (iii) whether the intervention logic was assessed as appropriate and shortcomings were disclosed, and (iv) whether the evaluator reviews underlying assumptions of the intervention logic.

Overall, only 13 reports (16%) are rated (very) good, 16 (20%) are found to be satisfactory, another (15, 19%) in need of improvement, and the remaining 36 (45%) were assessed as inadequate in presenting, discussing, and reviewing the intervention logics of the interventions.

A closer look at the sub-aspects shows only one fifth (16,20%) of the reports provided a complete and comprehensive description of the intervention logic, almost double that amount (31,39%) have not done so, and the rest were either incomplete or moderately described in the report. Also, only 23 reports (29%) included the input, expected output, outcome, and impact, i.e., the results model, in the report. Only one-third of the evaluators (27, 34%) assessed the intervention logic and presented the shortcomings, while only a quarter (21, 26%) of the evaluators reviewed the underlying assumptions of the intervention logic.



## Appropriate capturing of OECD DAC criteria

### Highlights of the section addressing EQ 2 and 4:

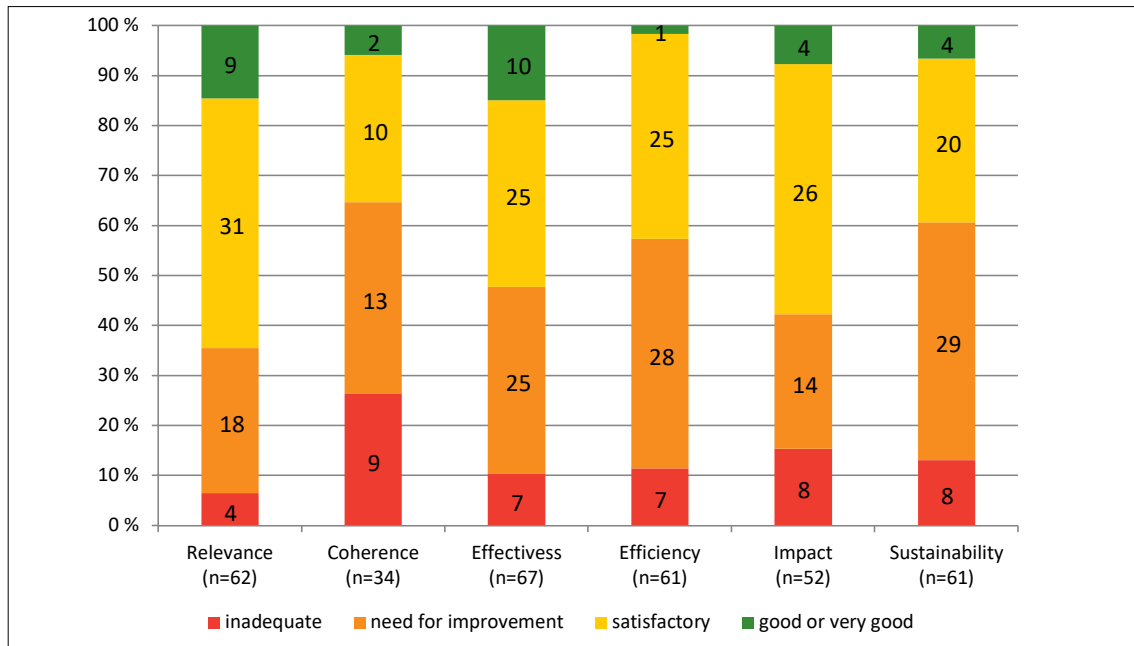
- Relevance was rated satisfactorily appropriately captured in 73% (45 out of 62) of the reports.
- Coherence is the least discussed with less than half (34 out of 80, 43%), and out of these 34 reports, 65% (22) were not satisfactory.
- Effectiveness is mostly appropriately captured in 67 out of 80 reports (84%) and received the greatest number of (very) good ratings (10).
- Efficiency is appropriately captured in three-quarters of the reports (61 out of 80, 76%), but more than half (35,57%) were graded less than satisfactory.
- Impact is the second least appropriately captured with 65% (52 out of 80) more than half of the reports (30, 58%) are rated satisfactory or better.
- Sustainability is appropriately captured in 61 reports, but only 39% (24) are satisfactory or better.
- OECD DAC scoring in the previous metaevaluation seemed to be partly better.
- Increased evaluation standards may cause deterioration (e.g., new coherence criterion)

Another aspect considered in assessing the quality of findings investigates how well evaluators captured the OECD DAC criteria, an important standard in the evaluation of development cooperation. It is also included in the MFA manual to evaluate Finnish development cooperation, which specifies what should be covered under each criterion. When assessing the appropriate capturing of DAC criteria, it was first assessed whether the criterion is discussed in general, and then it was analysed to which extent and at what quality level sub-aspects were addressed. Annex 6 provides details for relevance, coherence, effectiveness, efficiency, impact, and sustainability.

Looking at the aggregated assessment (see annex 6), only one report received (very) good scores in appropriate capturing of all DAC criteria, 25 reports (31%) reports were satisfactory, and most of the reports (48, 60%) were assessed in need of improvement.



**Figure 40: Appropriate capturing of DAC criteria**



Source: own statistics based on analysis of reports



The above Figure 40 summarises the assessment result with each bar representing a criterion and graded according to quality from inadequate, need for improvement, satisfactory and good or very good. Relevance is discussed in 62 reports and received the best rating regarding the quality of appropriately capturing the OECD DAC criteria, with roughly three quarters (45,73%) rated satisfactory or better. Less than a third (18, 29%) are considered needing improvement, and only 4 reports are inadequate (7 %).



As expected, being new, coherence is least discussed, with only 34 reports found to capture the criterion appropriately. Only 2 received a rating of (very) good (6%), and less than a third (10,29%) were found to be satisfactory, while more than half were either needing improvement (13, 38%) or inadequate (9 out of 34).



Meanwhile, with 67 reports, effectiveness is the most discussed and appropriately captured among the six OECD DAC criteria. In terms of quality of appropriate capturing of effectiveness, more than a third received a satisfactory rating (25,37%); at the same time, the exact same percentage of reports were assessed as needing improvement. Meanwhile, (very) good ratings were given to 10 reports (15%), and the remaining 10% (7 reports) were graded inadequate.



Efficiency is appropriately captured in 61 reports, and only 1 report received a (very) good rating. While less than half (25, 41%) are assessed as satisfactory, nearly another half (28, 46%) was determined to need improvement, and 7 reports (11%) were rated inadequate.





Concerning impact, 30 of the 52 reports addressing the criterion (58%) were rated satisfactory or better. Of the less than half that received less than satisfactory ratings, 14 (30%) needed improvement, and 8 (15%) were graded inadequate.



Lastly, sustainability was assessed as appropriately captured in 61 reports, and of these, 60% are found to either need improvement (29,48%) or inadequate (8,13%). A third of the sample (20, 33%) was assessed as satisfactory, and only 4 reports were considered (very) good (7%).

Comparing the previous metaevaluation unveils that OECD DAC criteria are partly worse captured nowadays. This deterioration may be caused by increased evaluation standards, as discussed in chapter 4.3.

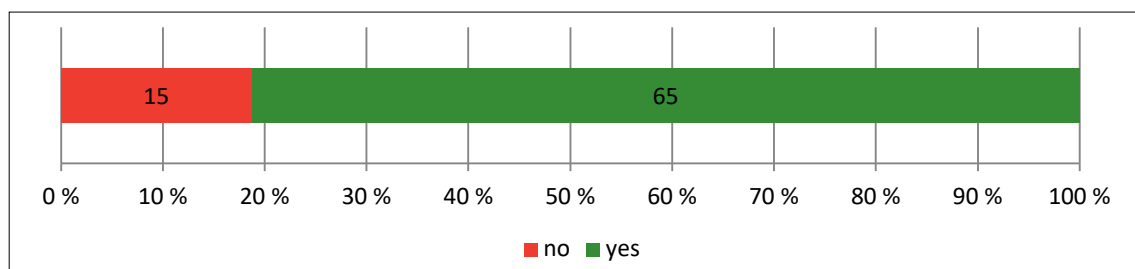
## Quality of conclusions and recommendations

### Highlights of the section:

- In 80%, conclusions are derived from findings; for 20%, this is not the case.
- The same holds true for deriving recommendations from findings and conclusions.
- No significant changes to the previous metaevaluations were observed.
- However, in this metaevaluation, recommendations are more often addressed to actors.

The report's conclusions were assessed as to whether they derived from findings. Figure 41 shows that about 80% (65) have conclusions consistent with the findings. However, in about 20% of the reports (15), linkages between findings and conclusions were missing. A most obvious inconsistency in this regard is that new information or data not yet presented in the findings are revealed in the conclusions.

**Figure 41: Conclusions are derived from findings (n=80)**

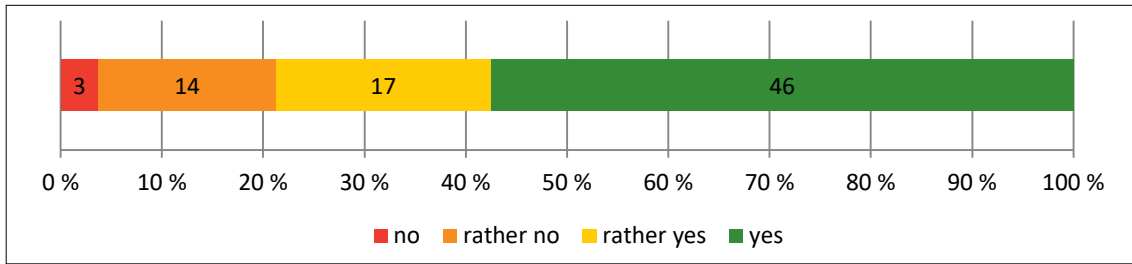


Source: own statistics based on analysis of reports

Figure 42 shows the result of assessing whether recommendations are derived from findings and conclusions. Similar to the quality of conclusions, almost 80% (63) had logical conjunction to conclusions or findings, either in a satisfactory manner (17,21%) or in (very) good terms (46,58%). A little over 20% (17) are rated to have recommendations that rarely (14,18%) or not at all (3,2%) derived from findings and conclusions.



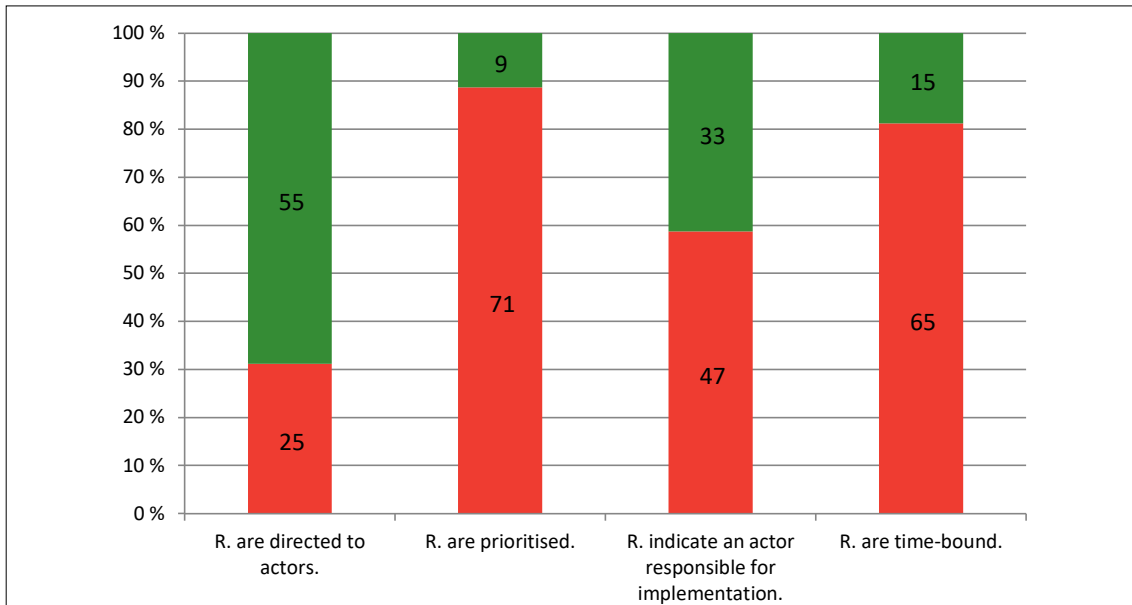
**Figure 42: Recommendations are derived from findings and conclusions (n=80)**



Source: own statistics based on analysis of reports

Furthermore, regarding recommendations, additional aspects were analysed whether recommendations are (i) directed to actors in general, (ii) prioritised, (iii) referred to an actor responsible for implementation, and (iv) time-bound. These four aspects are helpful to make recommendations more practical and actionable. Figure 43 reveals that recommendations are directed to actors in 69% (55) of the cases, but only less than half (33,41%) indicated actors responsible for implementation in their recommendations. On the contrary, only 9 reports (11%) prioritised the recommendations, and only 15 reports (19%) contained time-bound recommendations. Prioritisation and time-span are two aspects that need improvement to increase the usability of recommendations.

**Figure 43: Quality of recommendations (n=80)**



Source: own statistics based on analysis of reports

Although no major changes to the previous metaevaluations can be observed, recommendations were more often addressed to actors in this metaevaluation.



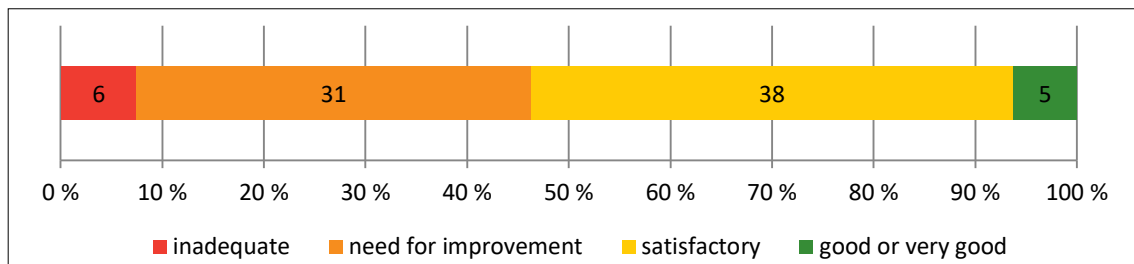
## Quality of executive summaries

### Highlights of the section addressing EQs 2 and 4:

- Almost half of the executive summaries were rated satisfactory; only five executive summaries were of high quality and comprehensive.
- About 40% of the summaries disclose the need for improvement, 6 are inadequate.
- Findings and conclusions are mostly provided.
- Evaluation design and scope of evaluation are least covered.
- No major changes to the previous metaevaluation can be observed.

The executive summary of a report is important because it is probably the most read portion of the report and provides a general hint to the quality of the report itself. Therefore, the quality of executive summaries concerning (i) their completeness, (ii) style and (iii) consistency with the report were analysed. As illustrated in Figure 44, almost half were rated satisfactory (38, 48%), and 5 executive summaries (6%) received a score of (very) good. More than a third (31, 39%) of the executive summaries are found to be incomplete and in need of improvement. 6 reports were rated inadequate (8%), and 4 of those did not provide any executive summaries at all.

**Figure 44: Overall quality of executive summaries (n=80)**



Source: own statistics based on analysis of reports

To be considered complete, the executive summary should resemble the evaluation report and include a description of the intervention and evaluation (objectives, scope, design, and methods), findings, conclusion, recommendations, and lessons learnt. On a more positive note, almost 90% of the executive summaries assessed provide findings and recommendations, more than 60% contain a conclusion and describe the objective of the evaluation, and roughly 50% describes the evaluation rationale/purpose and methods in the executive summary. On a less positive result, only 30 (38%) reports describe the intervention in the executive summary, only about a third (23,29%) provide lessons learnt, and the scope of the evaluation is only provided in 20% of the executive summaries (16). The least described component is the evaluation design, with only less than 10% (7) addressing it. Furthermore, the summarising table with findings, conclusions and recommendations requested by MFA was created only in 40% of the reports (32), with 23 summarising tables considered complete.



Finally, the executive summaries fared very well in writing style, with almost 90% (71) written in clear language. Although most executive summaries (52, 65%) were consistent with the report, inconsistencies were found in more than a third (28, 35%) of the summaries.

Again, a comparison with the previous metaevaluation did not show significant changes.

## Further aspects

### Highlights of the section:

- Only 6 reports were structured according to the MFA template.
- Three-quarters of the reports fail to discuss validation of findings, and a little more than 85% do not address quality assurance.
- The composition of the evaluation team was presented in only 30% of the reports.

In the quality assessment, attention was also paid to further aspects like (i) **the structure, style and annexes of the report**, (ii) **validation and quality assurance**, and (iii) **the composition of the evaluation team**.

Regarding **structure**, the MFA's manual provides a guideline on how the report should be structured, and only 6 reports adhere to this template. It is expected that reports not commissioned by MFA do not follow this structure, but more than 85% of the 44 MFA-commissioned reports did not comply with MFA's request.

The annexes fared well. Normally, regardless of the commissioning party, **ToRs** are requested to be annexed to the report, and 79% (63) of the reports complied with this. Almost three-quarters of the reports attach **lists of people interviewed** (59) and **documents consulted** (53). However, only 16% of the reports (13) annexed the **two-pager communication tool**.

Finally, in terms of style most of the reports, roughly 85%, were found to be comprehensible. 62 reports were considered adequately edited, and 71 were written in clear language.

When it comes to **validation of findings**, the reports were checked for mentioning validation by stakeholders and the commissioner, and roughly three-quarters of the reports failed to address this topic. Validation with stakeholders was discussed in about 29% of the reports (23) and only 23% (18) mentioned validation by MFA or other commissioners.

The reports were also reviewed if they address internal and/or external **quality assurance**. Only 11 (14%) contained a description of internal quality assurance, and only 9 reports (11%) addressed external quality assurance. Therefore, for a staggering majority of reports, it is unclear how or whether the quality of the evaluation is ensured.

To determine the appropriateness of the **composition of the evaluation team**, the following were analysed: (i) gender quality, (ii) thematic expertise, (iii) evaluation capacity, (iv) local expertise, and (v) lack of independence. The majority of the reports do not provide details of their evaluation team, and this information only existed in 24 reports (30%). Out of this information, it was assessed that only 11 of the evaluations were performed by an appropriate team composition.



The list of names on the cover page was gathered, and four more reports were added to assess gender balance, revealing that 20 out of 28 evaluations were performed by a gender-balanced team. Lastly, 3 reports were found to have a lack of independence due to numerous occurrences of subjective assessments.



# Annex 12: Interventions' Quality on Single OECD DAC Criteria

## Relevance

### Highlights of the section:

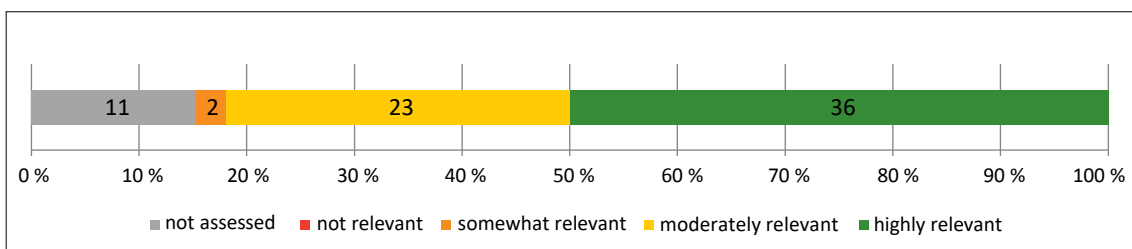
- Relevance was assessed in 61 reports, which accounts for 85% of all reports.
- 36 (50% of all) interventions were rated as highly relevant and 23 (32% of all) as moderately relevant, respectively.
- Only 2 interventions were assessed as somewhat relevant, while in no case an intervention was rated as not relevant.



The OECD DAC criterion of relevance describes the extent to which interventions are aligned with the policies and priorities of donors and the partner country and to what extent the intervention is tailored to the needs and requirements of its target group(s) and final beneficiaries.

Figure 45 shows that relevance was assessed in 61 reports (85% of all reports). 36 (50%) interventions were rated as highly relevant and 23 (32%) moderately relevant, respectively. Only 2 interventions were assessed as somewhat relevant, while no intervention was rated as not relevant.

**Figure 45: Projects' relevance according to the evaluation reports (n=72)**



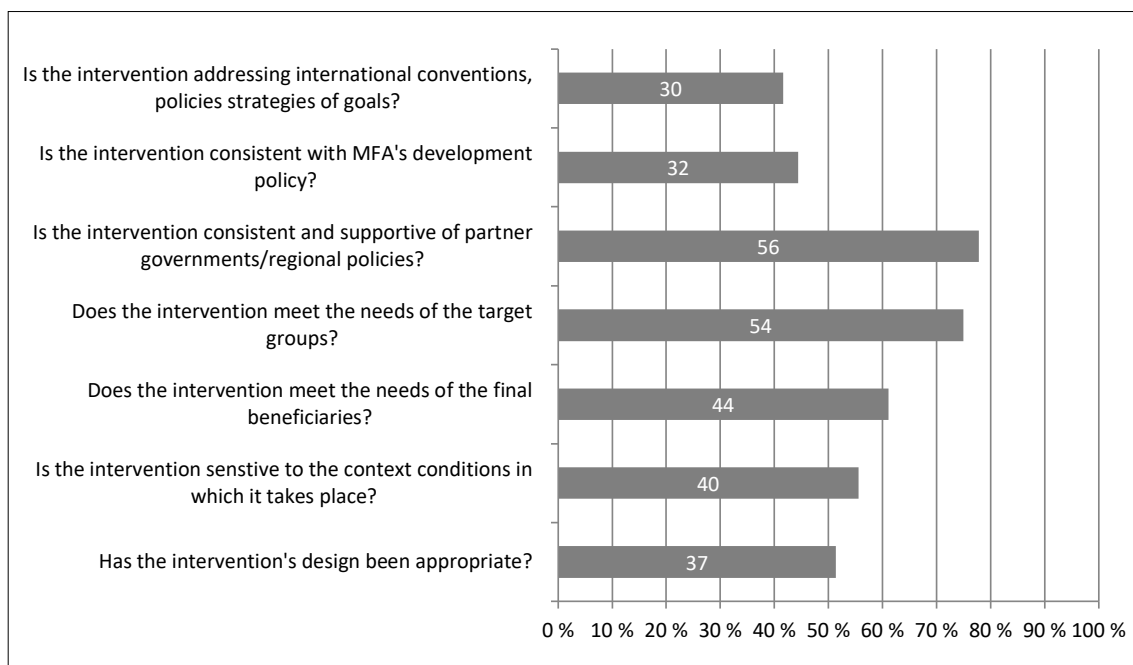
Source: own analysis of evaluation reports

Looking at different aspects of relevance, Figure 46 displays that the majority of evaluators assessed the alignment of interventions with the policies of partner countries (78%) and the extent to which interventions meet the needs of target groups (75%). The extent of meeting the needs of final beneficiaries, the context-sensitivity, and the quality of the design of interventions are also assessed in more than half of the reports (61%, 56% and 51%, respectively) while alignment with international policies and the consistency with MFA's development policy are addressed in less than half of the reports (42% and 44% respectively). Source: own analysis of evaluation reports



Figure 47 displays how the evaluators assessed the interventions' performance regarding the different aspects of relevance. 51 out of 56 interventions were assessed as consistent and 4 as rather consistent with partner government's policies, while 28 out of 30 interventions were assessed as consistent and one as rather consistent with international conventions and policies, respectively. The alignment with MFA's development policies was rated a bit lower, with 24 out of 32 reports rated as consistent and 6 as rather consistent. 36 out of 54 reports were assessed as meeting and 16 as rather meeting the needs of the target groups. Alignment with the needs of the final beneficiaries falls a bit behind, with 19 out of 44 reports assessed it with yes and 22 with rather yes. 22 out of 40 interventions were assessed as sensitive and 10 as rather sensitive to the context. What clearly stands out is the assessment of the interventions' design: for only 2 out of 37 interventions, the design was assessed as appropriate and for 17 as rather appropriate, while the designs of another 17 interventions were assessed as rather not appropriate.

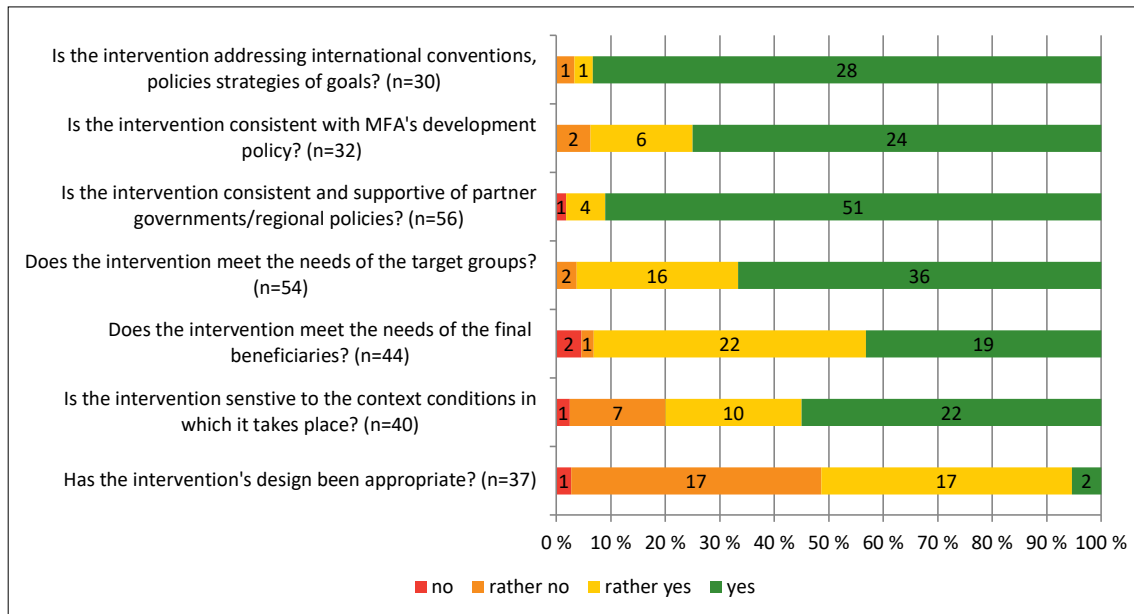
**Figure 46: Number of reports assessing different aspects of relevance (n=72)**



Source: own analysis of evaluation reports



**Figure 47: Assessment of different aspects of relevance**



Source: own analysis of evaluation reports

The most common positive reasons for assessing the extent of which projects meet the needs of the target groups were:

- Alignment with policies/strategy papers of partner institutions (13 cases)
- Responsiveness to demands/needs formulated by the target groups (17 cases)
- Direct reach out to stakeholders/partners, the conduct of a proper needs assessment or the direct involvement of stakeholders/partners in the planning of interventions (12 cases)
- Tailored intervention designs to specific local/country contexts or conducting of proper context assessments (10 cases)

By contrast, interventions were assessed as rather not or not meeting the needs of the target groups for the following reasons:

- Stakeholders were not involved in planning, or no needs assessment was conducted (4 cases)
- Inappropriate design: Limited scope or duration of interventions or the design was not adapted to the local context (6 cases)
- Wrong partners were selected for implementation, pointing to a mismatch of the target group and intervention design (3 cases)
- Target groups show a lack of ownership (3 cases)





*Both the survey results and Consultant's interview notes indicate that many beneficiaries, most importantly senior Government counterparts, indicated that the project established and maintained effective channels of communication with beneficiaries. Most notably, most beneficiaries indicated that the perspectives of their own agencies (Government agencies, industry/business associations, and other stakeholders) were adequately considered during the project design and implementation process (see Annex 10). (Report No. 31, Pos. 184)*

*Based on documentary evidence and information received through interviews, the capacity building by VIETADAPT [Developing and Implementing Climate Change Adaptation Measures at Local Level in Vietnam] has been highly relevant to the Vietnamese partners – responding to the identified needs in their key responsibility areas, such as water resources protection, assessment of climate change impacts on water resources, exchange and provision of water resources data and information and application of new technologies and relevant research on climate studies. In particular, VIETADAPT focused on capacity building in identifying and introducing feasible adaptation measures to national hazards and impacts of climate change, and building trust between scientists and local decision makers and stakeholders. (Report No. 30c, p. 32)*

*89% of all local staff surveyed believe that the training they received from diaspora was relevant for the host institution and professional their development (Report No. 32, p. 21)*

*UA and CoE stakeholders confirmed the AP interventions' clear alignment with partner needs and priorities. Interviewees provided a wealth of examples to underline relevance, many of which related to CoE support on aligning UA's regulatory framework with relevant standards, with CoE support ranging from the translation of relevant components of the normative framework (laws, regulations etc.) to the actual drafting of law/regulation.*

*However, interviewees also pointed to the relevance of other dimensions of the CoE's support, e.g. awareness raising and changing mind-sets, familiarising UA partners with good practices etc. (Report No. 52, p. 14)*

*PROMOSERV [Promoting Modernization of Hydro-meteorological Services in Vietnam] contributed to enhancement of the capacities of NHMS [National Hydro-Meteorological Service of Vietnam] in specific thematic areas that were selected through an internal assessment workshop. The thematic areas the project focussed on were: automatic weather stations, weather radars, real-time data quality control, central data management systems, integration of different meteorological data sources, lightning detection, storm cell tracking and weather forecasting. During the evaluation NHMS confirmed that these topics corresponded to their internal priority needs. (Report No. 30a, p. 30)*

*In Guatemala, key stakeholders - even if not necessarily final beneficiaries - were involved in the design process. Many of them had previously participated in the Finnfor I project, and were familiar with CATIE [Tropical Agricultural Research and Higher Education Centre] and the project approach. Forest concessionaries, for example, were able to propose activities which responded to previously identified needs in financing. (Report No. 35a, p. 8)*



*Similarly, RIA-1 [Research Institute for Aquaculture] was highly satisfied with the support received from Luke and Evira. It is evident that each phase of Finnish support has built on the achievements of the previous cooperation phase. The formulation of phases has been based on practical and pragmatic needs assessments conducted jointly by RIA-1, Luke [Natural Resource Institute of Finland] and Evira [Finnish Food Safety Authority]. This has resulted in realistic project designs with a focus on the improvement of technical and scientific capacities of RCCA [Research Centre for Coldwater Aquaculture] and RIA-1. (Report No. 30b, p. 30)*

*The available documentation suggests that Project's outputs are aligned with the MoES' [Ministry of Education and Science of Ukraine] and the education system needs identified in the document on the New Ukrainian School – Conceptual Principles of Secondary School Reform. (Report No 48, p. 20)*

*Similarly, in Tunisia, the project objectives are in line with the needs of the government. MOH [Ministry of Health] representative met in Tunis explained that the project was designed in close collaboration with the ministry as well as with other stakeholders. (Report No. 77, p. 18)*

*The replacement of the Call for Proposal (CFP) process with a flexible, demand-driven and responsive approach where projects can be solicited on an on-demand basis has facilitated improved and efficient operations of the MDTF [Multi-Donor Trust Fund], addressing of needs on the ground in a timely manner and enhancing contributions to achievement of the objectives of the MENA Regional Strategy. (Report No. 41, p. 53)*

*It came to the knowledge of the MTR team that the central-level partner organizations, particularly government partner organizations, held the opinion that the programme did not conduct a sufficient strategic planning exercise with proper assessment of the programme environment and analysis of the situation, adequate participatory consultation on the conceptual framework, or sufficient programme analysis for making strategic choices, during programme planning and design. They further added that the programme adopted a 'top-down approach' rather than a 'bottom-up approach', without participatory appraisal in processing and designing the programme. (Report No. 66, p. 23)*

*There has been limited participation of the implementing partners or beneficiaries in project formulation and it is evident that all pilot projects were actually not aware of the overall project purpose of regional dissemination and replication. Some pilot projects were not explicitly based on a need or request from stakeholders (e.g. ForInfo activities in the Pilot site in Thailand) (Report No. 35c, p. 6)*

*There were concerns that international interventions in Somalia had become "projectized" and internationalised, in ways that diminished ownership and leadership by Somalis and undermined the role of constitutionally mandated institutions such as MoIFAR [Ministry of Interior Federalism and Reconciliation] to lead this work and thus build credibility and trust with the wider population. (Report No. 60, p. 15)*

shows a range of examples for the assessment of how interventions meet the needs of the target groups.



**Box 14: Examples of reasons for the assessment of relevance related to the needs of the target groups**

*Both the survey results and Consultant's interview notes indicate that many beneficiaries, most importantly senior Government counterparts, indicated that the project established and maintained effective channels of communication with beneficiaries. Most notably, most beneficiaries indicated that the perspectives of their own agencies (Government agencies, industry/business associations, and other stakeholders) were adequately considered during the project design and implementation process (see Annex 10). (Report No. 31, Pos. 184)*

*Based on documentary evidence and information received through interviews, the capacity building by VIETADAPT [Developing and Implementing Climate Change Adaptation Measures at Local Level in Vietnam] has been highly relevant to the Vietnamese partners – responding to the identified needs in their key responsibility areas, such as water resources protection, assessment of climate change impacts on water resources, exchange and provision of water resources data and information and application of new technologies and relevant research on climate studies. In particular, VIETADAPT focused on capacity building in identifying and introducing feasible adaptation measures to national hazards and impacts of climate change, and building trust between scientists and local decision makers and stakeholders. (Report No. 30c, p. 32)*

*89% of all local staff surveyed believe that the training they received from diaspora was relevant for the host institution and professional their development (Report No. 32, p. 21)*

*UA and CoE stakeholders confirmed the AP interventions' clear alignment with partner needs and priorities. Interviewees provided a wealth of examples to underline relevance, many of which related to CoE support on aligning UA's regulatory framework with relevant standards, with CoE support ranging from the translation of relevant components of the normative framework (laws, regulations etc.) to the actual drafting of law/regulation.*

*However, interviewees also pointed to the relevance of other dimensions of the CoE's support, e.g. awareness raising and changing mind-sets, familiarising UA partners with good practices etc. (Report No. 52, p. 14)*

*PROMOSERV [Promoting Modernization of Hydro-meteorological Services in Vietnam] contributed to enhancement of the capacities of NHMS [National Hydro-Meteorological Service of Vietnam] in specific thematic areas that were selected through an internal assessment workshop. The thematic areas the project focussed on were: automatic weather stations, weather radars, real-time data quality control, central data management systems, integration of different meteorological data sources, lightning detection, storm cell tracking and weather forecasting. During the evaluation NHMS confirmed that these topics corresponded to their internal priority needs. (Report No. 30a, p. 30)*

*In Guatemala, key stakeholders - even if not necessarily final beneficiaries - were involved in the design process. Many of them had previously participated in the Finnfor I project, and were familiar with CATIE [Tropical Agricultural Research and Higher Education Centre] and the project approach. Forest concessionaries, for example, were able to propose activities which responded to previously identified needs in financing. (Report No. 35a, p. 8)*



*Similarly, RIA-1 [Research Institute for Aquaculture] was highly satisfied with the support received from Luke and Evira. It is evident that each phase of Finnish support has built on the achievements of the previous cooperation phase. The formulation of phases has been based on practical and pragmatic needs assessments conducted jointly by RIA-1, Luke [Natural Resource Institute of Finland] and Evira [Finnish Food Safety Authority]. This has resulted in realistic project designs with a focus on the improvement of technical and scientific capacities of RCCA [Research Centre for Coldwater Aquaculture] and RIA-1. (Report No. 30b, p. 30)*

*The available documentation suggests that Project's outputs are aligned with the MoES' [Ministry of Education and Science of Ukraine] and the education system needs identified in the document on the New Ukrainian School – Conceptual Principles of Secondary School Reform. (Report No 48, p. 20)*

*Similarly, in Tunisia, the project objectives are in line with the needs of the government. MOH [Ministry of Health] representative met in Tunis explained that the project was designed in close collaboration with the ministry as well as with other stakeholders. (Report No. 77, p. 18)*

*The replacement of the Call for Proposal (CFP) process with a flexible, demand-driven and responsive approach where projects can be solicited on an on-demand basis has facilitated improved and efficient operations of the MDTF [Multi-Donor Trust Fund], addressing of needs on the ground in a timely manner and enhancing contributions to achievement of the objectives of the MENA Regional Strategy. (Report No. 41, p. 53)*

*It came to the knowledge of the MTR team that the central-level partner organizations, particularly government partner organizations, held the opinion that the programme did not conduct a sufficient strategic planning exercise with proper assessment of the programme environment and analysis of the situation, adequate participatory consultation on the conceptual framework, or sufficient programme analysis for making strategic choices, during programme planning and design. They further added that the programme adopted a 'top-down approach' rather than a 'bottom-up approach', without participatory appraisal in processing and designing the programme. (Report No. 66, p. 23)*

*There has been limited participation of the implementing partners or beneficiaries in project formulation and it is evident that all pilot projects were actually not aware of the overall project purpose of regional dissemination and replication. Some pilot projects were not explicitly based on a need or request from stakeholders (e.g. ForInfo activities in the Pilot site in Thailand) (Report No. 35c, p. 6)*

*There were concerns that international interventions in Somalia had become "projectized" and internationalised, in ways that diminished ownership and leadership by Somalis and undermined the role of constitutionally mandated institutions such as MoIFAR [Ministry of Interior Federalism and Reconciliation] to lead this work and thus build credibility and trust with the wider population. (Report No. 60, p. 15)*



The most common reasons for the positive assessment of the extent to which interventions met the needs of the final beneficiaries were:

- Interventions were targeting particularly vulnerable groups (PWD, women, poorest etc.) (7 cases)
- Direct reach out to beneficiaries in the planning phase or conduct of a proper needs assessment (3 cases)
- Interventions were particularly responsive to demands or needs formulated by beneficiaries (11 cases)
- Interventions fit the local/country context (7 cases)
- Evaluators provided theoretical argumentation, referred to a common sense or justified the relevance for beneficiaries by results of the interventions (9 cases)

By contrast, projects were assessed as rather not or not meeting the needs of the final beneficiaries for the following reasons:

- Projects failed to address or reach out to the most vulnerable population groups (due to lack of resources, difficulties in accessing these groups (lack of infrastructure, social access, or security reasons etc.) or not recognising them in the design) (9 cases)
- The lack of a proper needs or context assessment (5 cases)
- Beneficiaries saw the intervention as not useful or not meeting their needs (according to survey/interviews) (4)

Box 15 shows a range of examples for the assessment of how interventions meet the needs of the final beneficiaries.



**Box 15: Examples of reasons for the assessment of relevance related to the needs of the final beneficiaries**

*“The Second Level Land Certification under REILA [Responsible and Innovative Land Administration] II is perceived as very relevant to the direct beneficiaries in rural areas (land-owners/ landless; women/men; old/youth; vulnerable groups). Surveyed beneficiaries state that the SLLC [Second Level Land Certification] has helped them “to build confidence”, gives them more security, reduces disputes, motivates farmers to take care of their land and makes women feel more safe.” (Report No. 5, p. 19)*

*“The types of trees planted at both sites visited are wanted by farmers. For Tra Vinh planting more mangrove trees boost shrimp production in the ecologically friendly farms. For Dak Lak, acacias are also preferred as a suitable tree for poorer soils that are not, or no longer, productive for coffee plantations.” (Report No. 21b, p.39)*

*“The AIHRC’s reports have a high degree of relevance, focusing on the most marginalized groups, 42 including violence against women, forced gynecological examinations, women employed in the defense and security sectors, child labour, causes and consequences of Bacha Bazi and people with disabilities, and on critical human rights issues such as violations of IHL (civilian casualties), torture, access to quality health services, economic and social rights, and the situation of returnees.” (Report No., p. 41)*

*“The final beneficiaries, i.e. Vietnamese start-ups, their founders, and other representatives of the local innovation ecosystems expressed appreciation towards the IPP [Innovation Partnership Programme] II and its activities. The same applies to the Finnish enterprises involved in the VMAP [Vietnam Market Access and Partnership Program], which provided them soft-landing/market entry service.” (Report No. 19, p. 4)*

*“Supporting the MOEST [Ministry of Education, Science and Technology] in the integration of Competence-based Soft Skills in the curriculum and teaching and learning practices is relevant and it is well aligned with GoN [Gov. of Nepal] and GoF [Government of Finland] priorities. It responds to the needs and priorities of the education system and its beneficiaries.” (Report No. 46, p. 49)*

*“Mainly due to the widespread insecurity in the country and a lack of resources the AIHRC’s [Afghanistan Independent Human Rights Commission] ability to provide its services to almost half the districts in the country is restricted; as many of these areas suffer the worst insecurity the residents are arguably most at risk of human rights violations. Within the resources at its disposal, the Commission reaches out to the population as much as possible.” (Report No. 25, p. 25)*

*“The 2012 estimate that the country needed 1.3m new houses to be built in order to meet the objectives of the Vision 2030, points to the continued need for a boosting the building construction sector and improving associated infrastructure development. Job needs creation is more than ever a priority as the country aspires to become a middle-income country and the ultimate beneficiaries of this Project (people who run, and work for small-scale enterprises) still need and want jobs in this sector and in the value chain of construction.” (Report No. 76, p. 28)*

*“One important concern, however, is that the program is not reaching those rural & vulnerable communities who do not have sufficient means to travel from remote villages to cities to seek care at MSA [Marie Stopes Afghanistan] clinics. The distribution of MSA health facilities in terms of location could be more equitable in terms of urban and rural underserved areas.” (Report No. 42, p. 44)*

*“One of the key challenges in the project design (intervention logic) is that in none of the countries where the project operates is the cost of health care considered as a key issue in the ability of migrants to access health care. Moreover, granting access to health care will not lead to an automatic improvement in access as there are additional barriers such as inability to communicate (language barriers); unfamiliarity with the policies and general stigmatization of this population in all five countries where the project operates.” (Report No. 77, p. 21)*

## Coherence

### Highlights of the section:

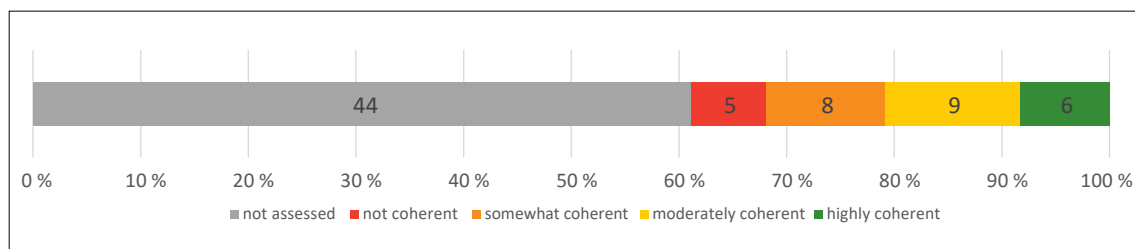
- Coherence was only assessed in 28 reports which account for only 39% of all reports.
- (8% of all) interventions were rated as highly coherent and 9 (13% of all) interventions as moderately coherent.
- 8 (11% of all) interventions were assessed as somewhat while 5 (7% of all) interventions as not coherent.



The OECD DAC criterion of coherence describes the extent to which interventions are coordinated with and complementary to other interventions of MFA (internal coherence) and with interventions of other actors active in the same region or sector (external coherence).

Figure 48 shows that coherence was assessed in 28 reports which account for only 39% of all reports. 6 (8% of all) interventions were rated as highly coherent and 9 (13%) moderately coherent. 8 (11%) interventions were assessed as somewhat while 5 (7%) interventions as not coherent.

**Figure 48: Projects’ coherence according to the evaluation reports (n=72)**

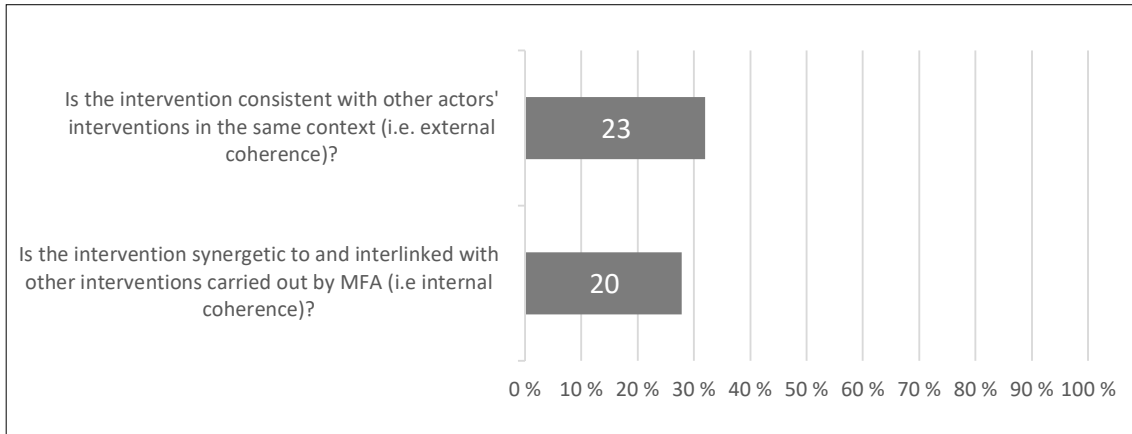


Source: own analysis of evaluation reports



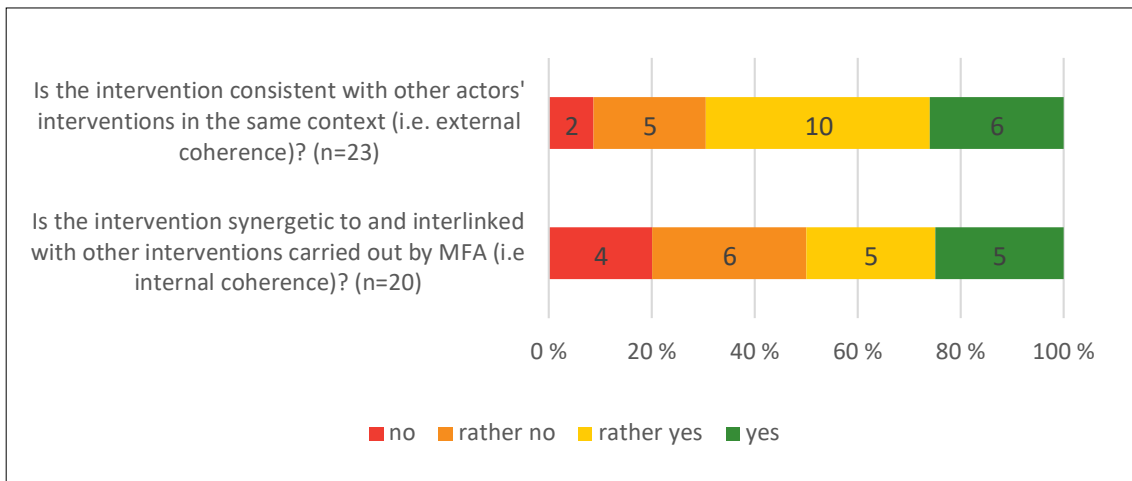
Looking at the two aspects of coherence, Figure 49 displays that internal coherence was assessed in 23 while external coherence was assessed in 20 out of 72 reports. Figure 50 displays how the reports assessed the interventions' internal and external coherence. With respect to internal coherence, 6 out of 23 reports were assessed as coherent and 10 as rather coherent, while 10 reports were assessed as rather not and 2 as not coherent. Looking at external coherence, 5 out of 20 reports were rated as coherent and 5 as rather coherent, while 6 were assessed as rather not and 4 as not coherent.

**Figure 49: Number of reports assessing different aspects of coherence (n=72)**



Source: own analysis of evaluation reports

**Figure 50: Assessment of different aspects of coherence**



Source: own analysis of evaluation reports

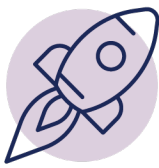




## Effectiveness

### Highlights of the section:

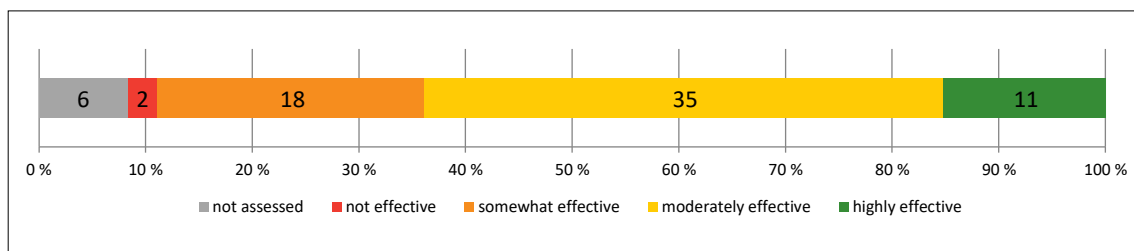
- Effectiveness was assessed in 92% of all reports.
- 11 (15% of all) interventions were assessed as highly effective and 35 (49%) as moderately effective.
- 18 (25% of all) interventions were assessed as somewhat while 2 (3%) interventions as not effective.



The OECD DAC criterion of effectiveness describes the extent to which interventions achieve their intended objectives.

Figure 51 shows that effectiveness was assessed in 66 reports which accounts for 92% of all reports. 11 (15% of all) interventions were assessed as highly effective and 35 (49%) as moderately effective. 18 interventions (25%) were assessed as somewhat while 2 interventions (3%) as not effective.

**Figure 51: Projects' effectiveness according to the evaluation reports (n=72)**



Source: own analysis of evaluation reports

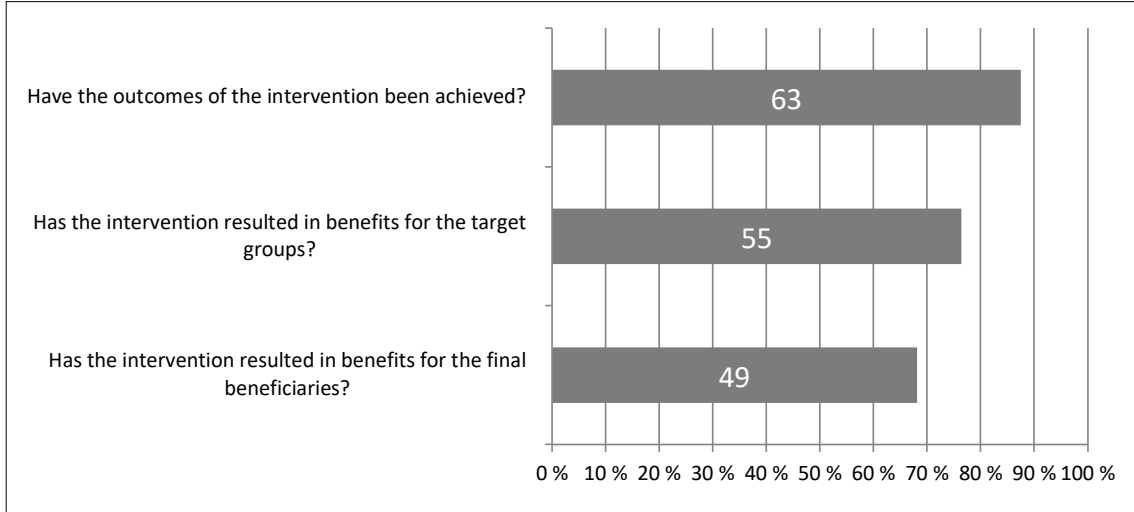
Looking at different aspects of effectiveness, Figure 52 displays that the majority of evaluators (88%) assessed the extent to which interventions achieved their outcomes. More than half of the reports also assess to what extent interventions have resulted in benefits for the target groups (76%) and the final beneficiaries (68%).

Figure 53 displays how the reports assessed the interventions' performance regarding the different aspects of effectiveness. According to the 63 reports assessing outcome achievement, 14 out of the 63 interventions have achieved, and 34 rather achieved their outcomes, while 11 have rather not, and 4 have not achieved their outcomes. 30 out of 55 interventions were assessed as resulting in benefits for the target groups and 28 as rather resulting, while 11 as rather not and 3 as not resulting. According to the 49 reports assessing the benefits for the final beneficiaries, 16 out of the 49



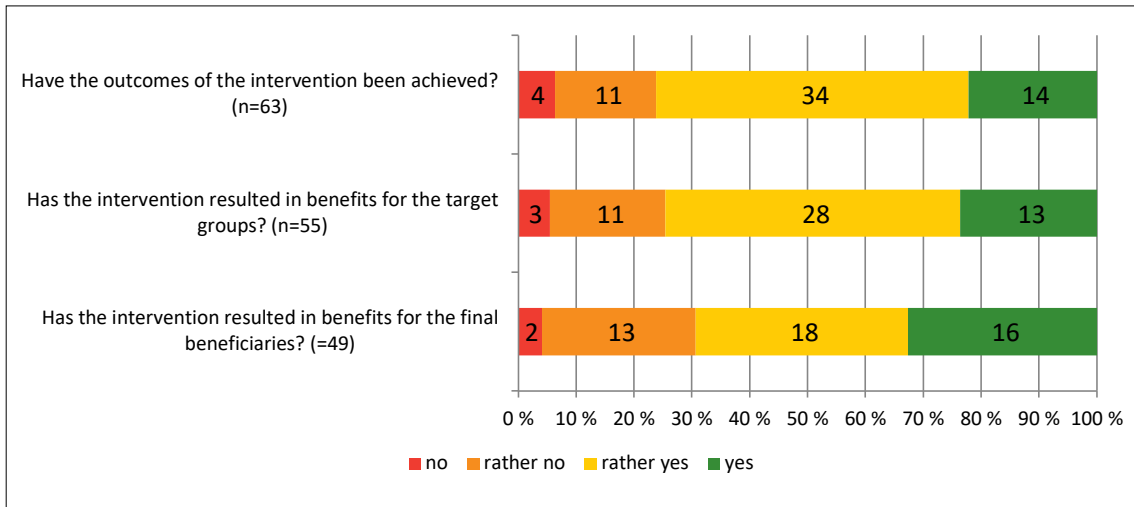
interventions were assessed as resulting in the respective benefits and 18 as rather resulting, while 13 as rather not and 2 as not resulting in benefits.

**Figure 52: Number of reports assessing different aspects of effectiveness (n=72)**



Source: own analysis of evaluation reports

**Figure 53: Assessment of different aspects of effectiveness**



Source: own analysis of evaluation reports



With respect to the achievement of outcomes, the most common reasons for the positive assessment were:

- Improved capacities of target groups (16 cases)
- Improved coordination/cooperation among stakeholders (10 cases)
- Efficient project management, flexibility and/or effective quality assurance (9 cases)
- The interventions' policy level influence or improved legal regulations (9 cases)
- Successful awareness-raising among stakeholders/change of mindset (6 cases)
- New equipment, improved IT solutions and/or improved technical set-up (5 cases)
- Good performance of local partner organisations (4 cases)
- Good communication strategies (3 cases)
- Innovative or particular effective intervention approaches (3 cases)
- By contrast, reasons, why interventions did not achieve their outcomes were:
- Poor project design (8 cases), pointing to
  - overambitious or unrealistic objectives (in 5 out of 8 cases)
  - a lack of context sensibility (in 2 out of 8 cases)
  - insufficient institutional development (in 2 out of 8 cases)
  - missing efforts towards changes in policy frameworks (in 3 out of 8 cases)
  - or a lack of or insufficient capacity building (in 3 out of 8 cases)
- M&E Shortcomings (8 cases)
- Insufficient cooperation, coordination or integration of important stakeholders (7 cases)
- Political instability in partner countries (8 cases)
- Inadequate infrastructure in partner countries (5 cases)
- Insufficient funding (3 cases)

Box 16 shows a range of examples for the assessment of how interventions meet the needs of the final beneficiaries.



**Box 16: Examples of reasons for the assessment related to outcome achievement**

*“Some IOM and THL [Finnish National Institute for Health and Welfare] staff indicated that the success of the project is linked mainly to the approach of ensuring that the diaspora and national experts work together, the willingness of the diasporas to come and impart their skills and relocate temporarily, very good working relationships and cooperation with the MoHD [Ministry of Health and Development] and other stakeholders.*

*The project also effectively supported the strengthening and capacity of various departments including MCH [Mother and Child Health], dialysis, dental, emergency and mental health departments among others.” (Report No. 2, p. 26)*

*“By bringing together a broad range of actors who are involved in providing social protection services within one programme and promoting active participation within programme activities and meetings, the UNJPSP [United Nations Joint Programme on Social Protection] facilitated enhanced coordination within the social protection sector.” (Report No. 63, p. 44)*

*“Workshops and training sessions have, according to interviews and the evaluation team’s observations, been well organized and speakers and trainers well qualified. Statements and other more analytical pieces of documentation reviewed by the evaluation team have been both well researched and well written. While most of the guidelines that will be prepared, or updated, as part of the project have not yet been finalized, the organization’s track record and quality assurance system give strong reasons to assume that these publications will meet high expectations.” (Report No., p. 20)*

*“The UNJPSP has facilitated the development of capacity of government ministries to be able to effectively develop social protection programmes and policies.” (Report No. 63, p. 43)*

*“The Synthesis Evaluation agrees with the draft BCR: SUFORD-SU– and also SUFORD and SUFORD-AF before it – had an important component providing support to legal and regulatory environment. The project has delivered a large number of inputs and outputs to policy and legal development, but the extent to which SUFORD-SU can be credited for these changes is not easy to assess. Certainly, SUFORD projects with support from WB and GoF [Government of Finland] have been instrumental in passing and implementing decrees that have led to the institutionalization of the PSFM [Participatory Sustainable Forest Management] system.” (Report No. 17a, p. 90)*

*“Two important potential target groups – the military and people in states and regions – have been reached only to a limited extent.” (Report No. 15, p. 25)*

*“Capacity development in development cooperation must target not only individuals but also organisations and how they work. The programme has been successful in reaching a number of people as representatives for organisations and institutions but we have found only few examples when an organisation or an institution has been targeted as a whole.” (Report No. 15, p. 25)*



With respect to the aspect of whether interventions have resulted in benefits for the target groups, the most common reasons for the positive assessment were:

- Improved capacities of target groups (20 cases)
- Improved coordination/cooperation among stakeholders (16 cases)
- The interventions' policy level influenced or improved legal regulations (13 cases)
- Successful awareness-raising among stakeholders/change of mindset (7 cases)
- New equipment, improved IT solutions and/or improved technical set-up (7 cases)
- Innovative or particular effective intervention approaches (3 cases)
- High ownership of target groups (3 cases)
- Projects' responsiveness to the context and needs of the target group (2 cases)

By contrast, reasons why interventions did not result in benefits for the target groups were:

- Insufficient capacity building or limited capacities of partners (11 cases)
- Insufficient funds (3 cases)
- Inadequate coordination or lack of involvement of important stakeholders (4 cases)
- Inadequate infrastructure in partner countries (2 cases)
- Poor project design (5 cases)
- Outputs were not used by target groups (4 cases)

Box 17 shows a range of examples for the assessment of whether interventions result in benefits for the target groups.



**Box 17: Examples of intervention resulting in benefits for the target group**

*“The project’s role in the capacity development has been pivotal. Government staff capacities have improved in e.g. developing forest management plans, forest certification, village forestry, forest cover assessments, and forest landscape planning. The trained staff members on all government levels have developed sufficient capacities to undertake routine tasks.” (Report No. 16a, p. 90)*

*“Capacity building for labour inspectors in Tunisia and women in workers’ organizations in Egypt serve as an initial seed that requires further attention. Interviewed beneficiaries on these streams explained that they require additional support to be able to capitalize on the knowledge and skills acquired to improve the situation of other women and play a more active role within their institutions. In Egypt, the training on role of trade unions in achieving gender equality was also delivered to men workers but with a lower percentage of participation.” (Report No. 1, p. 6)*

*“WB Task Teams support country programmatic needs and operate in a highly responsive manner which ensures the appropriateness of investments and is knowledge-based (backed by an awareness of the absence of alternative financing mechanisms). Without the MDTF [Multi-Donor Trust Fund], important needs would not be addressed, and subsequent commitments would not be made.” (Report No. 41, p. 59)*

*“The programme has facilitated the capacity building of government ministries to be able to effectively develop social protection programmes and policies through interregional lesson sharing visits to best practice sites, supporting the development of technical expertise in policy analysis and promoting cross ministerial linkages. The programme has also strengthened national level design and implementation capacity of the key line ministries through development of institutional frameworks to enhance coordination, enhanced delivery mechanisms such as ZISPIS, Integrated Framework of Basic social protection programmes and Single Window as well as improving targeting mechanisms through the use of M-tech system that greatly helped improve the coverage and reduce on the time.” (Report No. 63, p. 17)*

*“At present the provinces do not have adequate capacity to continue FLM [Forest Landscape Management] without external TA. In remote sensing and GIS [Geographic Information System], already since SUFORD the technical capacity has been relatively good at the central level. However, in the provinces and districts the capacities are not equally well developed. External support for basic GIS tasks would be required.” (Report No. 17a, p.90)*

*“In terms of qualitative achievements, such as capacity development of the state and non-state actors (SALAC, Ombudsman office and PCOs, as well as thematic CSOs) the mid-term achievements are not significant and more efforts is needed. For instance, the Ombudsman office still has weak human rights monitoring capacity and expertise; CSOs have a weak advocacy and policy influencing capacity, while SALAC needs more consistent support in increasing management capacity of the free legal aid system in Tajikistan.” (Report No. 37, Pos. 338)*



*“While the effectiveness of the IT-related capacity build-up has to be rated as low, the project and the TA could not have influenced this any further. The fact that there was no IT role, leave alone capacity, at the NGD [National Geographic Department] that could possibly handle the planned equipment and tasks should have been realised at the design stage of the SNGS. The development of a proper IT unit or section within the NGD should have been part of the project design, alongside general institutional development to prepare NGD for its expected role as a provider of national information services.” (Report No. #17b, p. 123)*

*“However, in one way, the project’s design itself limits the effectiveness of the overall project because it can be seen as overstressing. This is problematic because placing one or two experts in the Ministry of Education, as is the case in Jubaland, is a difficult situation for the experts themselves due to the workload and time needed to complete the required tasks. The geographical representation of diaspora assignments is also seen as not balanced, with more education diaspora experts placed in Mogadishu than any other target location. The disparity is even more pronounced when looking at the number of experts placed in hospitals and those placed in the universities. The evaluation team observed that the fewer experts placed in an institution, the greater the levels of dissatisfaction and frustration, which is linked to the workload, and is made worse by perceived lack of cooperation and interest from project stakeholders. A greater understanding and analysis of the institutional needs of host institutions can help determine where and how long experts should be contracted as part of the project.” (Report No. 32, p. 49)*

With respect to the aspect of whether interventions have resulted in benefits for the final beneficiaries, the most common reasons for the positive assessment were:

- Improved capacity of target groups/beneficiaries (15 cases)
- Successful awareness-raising among stakeholders/beneficiaries (change of mindset) (7 cases)
- Improved coordination/cooperation among stakeholders or beneficiaries (4 cases)
- Good project management/high flexibility (4 cases)
- Policy level influence/improved legal regulations (3 cases)
- Innovative/particular effective approaches (3 cases)
- Responsiveness to the context and needs of the beneficiaries (2 cases)
- Beneficiaries have better access to finance/credits (2 cases)
- Beneficiaries have better access to public services (2 cases)

By contrast, reasons why interventions did not result in benefits for the target groups were:

- Projects did not address/failed to reach marginalized groups (women, disability etc.) (7 cases)
- Insufficient capacity building (4 cases)
- Poor project design (3 cases)
- Overambitious or unrealistic objectives or results hypothesis (in 2 out of 3 cases)



- lack of context sensibility (in 2 out of 3 cases)
- insufficient efforts towards changes in policy framework and legal regulations (in 2 out of 3 cases)
- Insufficient funding (3 cases)
- Inadequate coordination or lack of the integration of important stakeholders (3 cases)
- M&E shortcomings (2 cases)
- Political instability in partner countries (2 cases)

Box 18 shows a range of examples for the assessment of whether interventions result in benefits for the final beneficiaries.





**Box 18: Examples of reasons for the assessment related to intervention resulting in benefits for the final beneficiaries**

*“It was indicated that capacity building activities created awareness and showed alternative business activities in the area and enhanced the knowledge of the beneficiaries” (Report No. 9, p. 30)*

*“Participants were also self-aware of needed behaviour changes. For example, one participant from Faryab province said, “Previously people would say if you use FP [Family Planning] methods you will become infertile forever, but now we know that FP is temporary.” The majority of the IDI [In-depth Interview] respondents agreed that a major shift is occurring in people’s perception regarding FR&RH [Family Planning & Reproductive Health]. An increasing number of people are accepting FP&RH. Some level of community, family and individual awareness is present, and there do exist some religious support for certain FP methods, such as condoms, injections and pills. All of these factors together with FP&RH intervention, such as the one being evaluated, have contributed to a positive change in the behaviour and attitude of people toward FP” (Report No. 42, p. 31)*

*“The poorest and/or ethnic minority households have less access to land in most project sites, and/or where land ownership conflicts exist. Livelihood models only applicable to those with land rights. FORMIS [Forest Sector Management Information System] benefits are only applicable to those with forestry land rights (Result 2). Poor ethnic minority people in the project areas are capable to make use of the information generated by FORMIS (Result 2) only in Cao Bang, where the population is entirely ethnic minority and ethnic minority people have land rights.” (Report No. 21b, p. 48)*

*“While the graduates have rated the training highly they raised some concerns related to shortages of tools and materials as well as limited practical exposure during training. In a Focus Group Discussion with the evaluation team, however, it was found that former trainees were still unemployed more than three months after they graduated.” (Report No. 76, p. 33)*

*“A large number of the respondents also highlighted the importance of providing free FP&RH services in rural communities given widespread poverty in the country. Almost all respondents agreed that MSA [Marie Stopes Afghanistan] should expand and extend its FP&RH services to other areas in order to reach more people, specifically in the rural areas. This indicates that while the awareness about services among urban population is good, it still requires improvement with regards to the awareness about the availability of services in the rural areas.” (Report No. 42, p. 29)*

*“Also, villagers have been trained to manage their development projects, although more capacity building would have been needed in grant and revolving fund management (MAF/DoF/SUFORD-SU 2019). Other evidence indicates that the villagers also need training to successfully implement their selected livelihood schemes.” (Report No. 17a, p. 90)*



## Efficiency

### Highlights of the section:

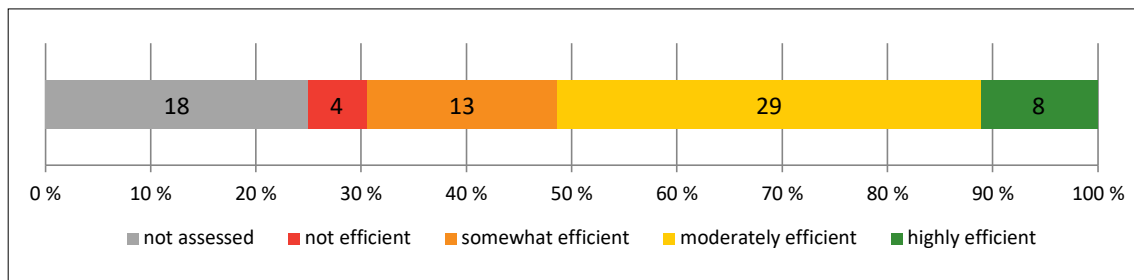
- Efficiency was assessed in 54 reports which accounts for 75% of all reports.
- 8 (11% of all) interventions were assessed as highly effective and 29 (40%) as moderately effective.
- 13 (25% of all) interventions were assessed as somewhat, and 4 (6%) interventions as not effective.



The OECD DAC criterion of efficiency refers to the efficient use of resources in relation to the respective outputs and outcomes of an intervention. It also includes aspects of management and operational efficiency.

Figure 54 shows that efficiency was assessed in 54 reports which accounts for 75% of all reports. 8 (11% of all) interventions were assessed as highly efficient, and 29 (40%) as moderately efficient. 13 (25%) interventions were assessed as somewhat, and 4 (6%) interventions as not effective.

**Figure 54: Projects' efficiency according to the evaluation reports (n=72)**



Source: own analysis of evaluation reports

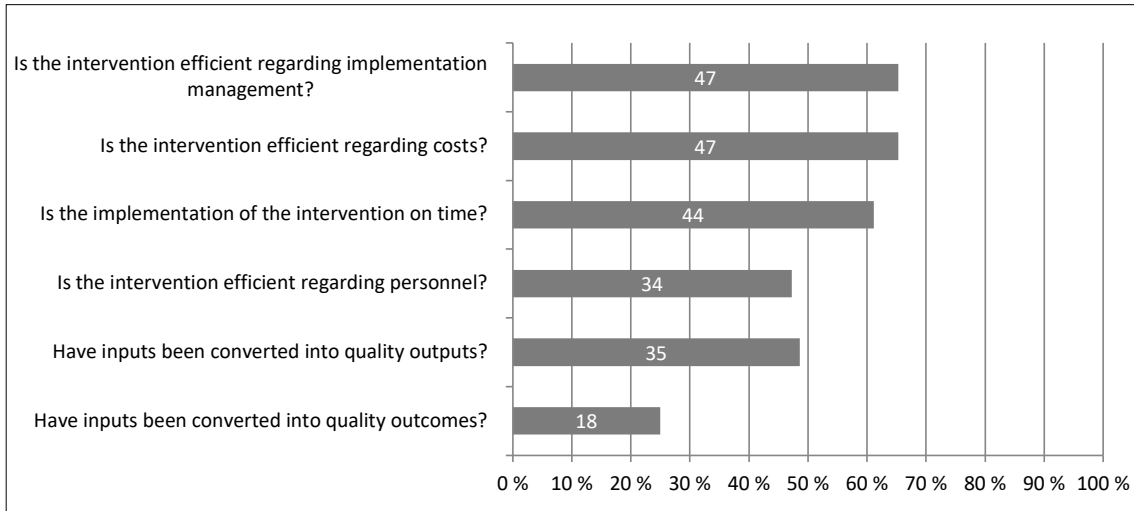
Looking at different aspects of efficiency, Figure 55 displays that more than half of the reports assessed the implementation management of interventions (65% of all reports), their cost efficiency (65%) and their timeliness (61%). Efficiency regarding personnel and with respect to the relation of inputs against achieved outputs was assessed by slightly less than half of the reports (47% and 49%, respectively). Only 25% of all reports assessed the extent to which projects converted their inputs into high-quality outcomes.

Figure 56 displays how the reports assessed the interventions' performance regarding the different aspects of efficiency. According to 47 reports, 14 interventions were assessed as efficient and 19 as rather efficient regarding implementation management. 14 were assessed as rather not, and 7 as not efficient in this respect. 47 reports also assessed cost-efficiency, concluding that 11 interventions were cost-efficient and 23 rather cost-efficient. 8 were assessed as rather not, and 5 as not efficient regarding cost. According to 44 reports, 5 interventions were on time and 16 rather on time, while 18 were rather not and 5 were not on time. 3 out of 34 interventions were assessed as efficient and 8 as rather efficient regarding personnel, while 16 were assessed as rather not and 7 as not efficient regarding personnel. According to the assessment of 34 reports, 12 interventions converted and 19 rather converted inputs into high-quality outputs. Only 2 interventions did not or did rather not



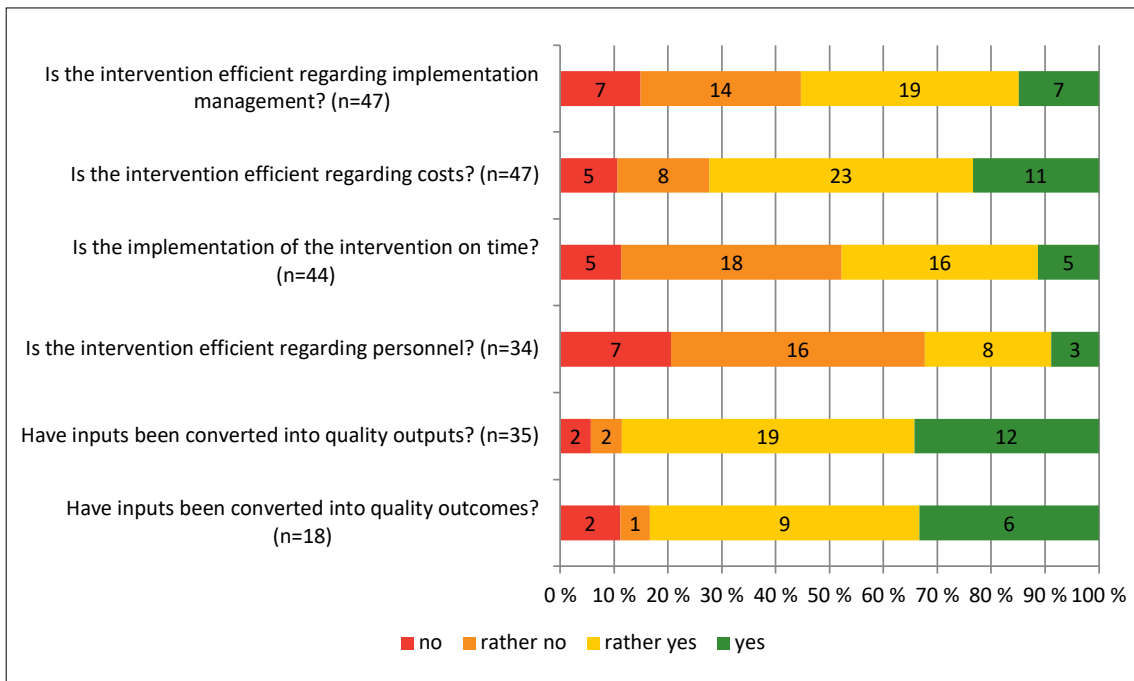
achieve high-quality outputs in relation to their inputs. With regards to the quality of outcomes, which was assessed in only 18 reports, 6 interventions did convert, and 9 rather converted their inputs into high-quality outcomes. 1 intervention did rather not, and 2 did not convert their resources into high-quality outcomes.

**Figure 55: Number of reports assessing different aspects of efficiency (n=72)**



Source: own analysis of evaluation reports

**Figure 56: Assessment of different aspects of efficiency**



Source: own analysis of evaluation reports



The vast majority of evaluations did not provide reasons for the timeliness of projects. A few reasons that were provided are:

- Fast and efficient staff recruitment (1 case)
- Flexibility and adaptability to change (1 case)
- Efficient management (3 cases)

Reasons for delays included:

- Delays or unsatisfactory results in staff recruitment (13 cases)
- Delayed disbursement of funds (10 cases)
- Unrealistic schedules and time plans (4 cases)
- High administrative burdens and overly bureaucratic or inefficient procedures (4 cases)
- External factors (beyond the project's influence/control) (3 cases)

Box 19 shows a range of examples for the assessment of whether interventions were on time.



**Box 19: Examples of reasons for the assessment related to timeliness of the intervention**

*“The implementation of the project started timely as some of the diaspora experts that were part of Phase III of the project have been recruited to work for Phase IV immediately at the start of the project. This was very crucial in reducing the time lost in identifying and onboarding diaspora professionals which usually is time taking.” (Report No. 2, p.29)*

*“Nonetheless, the project is commended for not experiencing significant delays in implementation or delivery of activities. All delays noted during the evaluation are normal and are normal in light of the context. Rather, the flexibility of the implementation strategies and the ability of the project to adapt to challenges has enabled it to complete activities and deliverables in a timely fashion.” (Report No. 33, p. 26)*

*“The stakeholders raised the effects and impacts of COVID-19 pandemic to their projects. The pandemic affects both Finland and Ukraine and travel between the countries has not been possible for quite some time. This has had implications on the work of ongoing projects.” (Report No. 33, p. 34)*

*“School grants targeting cluster inclusive education resource centres are allocated, items procured and distributed but with delay, caused by external factors beyond the control of the TA team.” (Report No. 3, p.18)*

*“Although trainings to community members to strengthen their skills in their adopted livelihood scheme were planned, most of the planned trainings could not take place due to lack of project resources. Overall, the process from planning to approval and to transfer of funds took a long time. Senior DoF [Dept. of Forestry] and TA staff acknowledged that the delays were caused partly by changed conditions on behalf of WB that needed to be reflected in the project guidelines, and partly by capacity issues with the Government financial staff. According to the draft BCR [Borrower’s Completion Report], financial management remained a major problem throughout implementation of SUFORD-SU, even after DoF had hired additional staff.” (Report No. 17, p. 96)*

*“Occasional delays in funds disbursements and complete withdrawal of funds intended for specific activities interrupted implementation and in some cases some activities could not be completed within the lifetime of the joint programme.” (Report No. 63, p. 18)*

*“There has been extensive disruption to project implementation due to the delays in financial disbursements and short funding cycles. The delay in financial disbursements, due to the move within UN Environment to a new administrative system—UMOJA, has had serious effects on activities in countries.” (Report No. 68, p. 63)*

*“The process of purchasing the materials has been time consuming due to the delay of GE-QIP-E fund disbursement, long procurement processes and limited availability of the items on the market. Thus, it was not possible to accomplish this task within the planned time.” (Report No. 3, p. 19)*



*“Recruitment delays have further frustrated the intent of the project and hampered the relations. (...) furthermore, recruitment delays, non-availability of a suitable candidate for the position of International Technical Advisor, participant selection, identification of capacity development areas for job placement, assessment of required skill sets etc which were not included in project activities have consumed the productive time.” (Report No. 4, p. 6)*

*“Once the Project got properly started, there may have been some delays in fielding some missions (i.e. in implementing some activities). These have occurred due to the unavailability of WCO Accredited Experts, which are very limited in number in the region and not always readily available from outside the region either.” (Report No. 12, Pos. 222)*

*“Considering the overall structure of the PRF, it appears that the timeframe for engaging in reforms and implementing them at different levels of the education system was not fully adequate.” (Report No. 6, p. 3)*

*“The project outcomes are generally considered to be attainable if some key challenges are met, despite the fact that the project was delayed by the coup d’état of late 2014.” (Report No. 68, Pos. 311)*

Interventions were assessed as cost-efficient for the following reasons:

- The reasonable or low cost of purchased items (7 cases)
- The share of administrative cost/overhead of overall cost was reasonable (3 cases)
- Synergies or cost-sharing with other interventions or partner organizations (6 cases)
- Solid administrative management systems (5 cases)
- The leverage of additional funds (2 cases)

Reasons for assessing interventions as rather not or not cost-efficient included:

- The share of overhead/administrative cost against the overall budget was too high (5 cases)
- Inefficient administrative structures (4 cases)
- A shortcoming in the interventions’ designs (3 cases)
- Unrealistic resource planning or the overall budget of interventions being too low (2 cases)
- Too high cost of international staff (2 cases)
- Insufficient documentation did not allow a proper assessment (2 cases)
- Unfavourable funding modalities or a delay of disbursement of funds (2 cases)
- External factors (not to be influenced by the project) (2 cases)
- Insufficient monitoring (2 cases)

Box 20 shows a range of examples for the assessment of cost-efficiency.



#### Box 20: Examples of reasons for the assessment related to cost efficiency of the intervention

*“As far as the evaluation team has been able to observe, it also appears that costs for individual items and services purchased are reasonable.” (Report No. 7, p. 35)*

*“In overall, the project budget is tight to cover distinct technical areas of work in two target countries with a large number of activities. To complement the project’s resources, the DW II project team successfully established synergies with other ILO projects (the Skills project and the YE [Youth Employment] project) through cost-sharing arrangements and/or in-kind contributions.” (Report No. 47, p. 36)*

*“The project implemented workplans in a timely fashion and had sound and efficient financial management systems. The outputs of the project justify the costs and the distribution of administrative costs vs project activities is efficient.” (Report No. 1, p. 26)*

*“The estimated (one-time) development costs for the FRMS database were EUR 0.10 per forestry plot, and it is expected that the data of any given plot will be updated every seven years at a cost of EUR 0.24 on average. Given the complexity of the plot attribute data (and its value for users), these figures suggest that the FRMS is cost efficient. The training costs per participant correspond a few hundred EUR, with a decreasing trend over time.” (Report No. 37, p. 53)*

*“The share of funds allocated directly to field activities is higher than in other comparable projects, and the proportion of general and TA costs is reasonable.” (Report No. 9, p 33)*

*“According to the PMU, the programme underspending is caused by the fact that many activities (especially under the Partnership for Innovation component) were done with no costs or shared costs with other partners.” (Report No. 19, p. 80)*

*“This causes delays in report delivery as reported by interviewees and a possibly important increase of transaction costs due to duplication of workload when draft reports bounce back and forth between the three capitals.” (Report No. 11, p. 20)*

*“However, there certainly is duplication of efforts and workload in the form of sending report drafts back and forth between the Consortium members causing higher transaction costs in the form of increased use of time – the transaction costs understood here as the cost of transforming the total financial contribution into activities in favour in the final beneficiaries.” (Report No. 11, p. 35)*

*“The evidence from the financial data provided appears to show that the resources available have been spread thinly across a complex programme, with a very wide range of outputs, and were overambitious in terms of what could be achieved, working in a difficult context with partners with limited capacity. The evidence from the budget and expenditure data for 2014 to 2016, shows a significant underspend for the programme as a whole and for each of the thematic areas for 2014 and 2015. Only in 2016 does expenditure more closely match what the budget proposed at the start of the year, as UN Women focused efforts on improving the management of the programme.” (Report No. 27, p.21)*

*“However, at the outcome level, because of the resources spreading thin, while there are and will be results, more could have been achieved with an intervention designed to be more focused.” (Report No. 12, Pos. 226)*



Common reasons for assessing the interventions' efficiency regarding personnel positively were:

- The high commitment and dedication of staff (5 cases)
- A low staff turnover and long-term commitment of staff (3 cases)
- High competencies and qualification of staff (3 cases)
- The engagement of experts with high contextual, sectoral, or institutional knowledge (3 cases)
- The engagement of local experts familiar with culture and context or with important local networks (2 cases)
- Regional offices increased across-country coordination (2 cases)
- The recruitment of international experts (2 cases)

Common reasons for assessing the interventions' efficiency regarding personnel negatively were:

- Project were understaffed (11 cases)
- A high staff turnover (10 cases)
- Problems regarding timely recruitment of staff (4 cases)
- Lack of a particular expertise or experience (3 cases)
- Too much international staff instead of local staff produced too high costs (3 cases)
- Inefficient decision-making structures (3 cases)
- The inefficient distribution of tasks among staff (3cases)

Box 21 shows a range of examples for the assessment of efficiency regarding personnel.





### Box 21: Examples of reasons for the assessment related efficiency regarding personnel

*“The added value the experts brought with regards to knowledge of the institutions, context and the project has been crucial as it helped to capitalize on their skills and knowledge to the advantage of realizing the project outcomes efficiently. This helped to achieve tangible results from the first year of implementation.” (Report No. 2, p. 29)*

*“The project costs were reduced since ActionAid conducted in the project in sites where they had other ongoing activities. Thus, the project staff had already established relationships with local residents and officials. Project monitoring trips from the AAV office in Hanoi were more efficient, as they could cover several different activities.” (Report No. 21, p. 55)*

*“Overall, the MTR considers that the TA support to regions is an appropriate modality in particular when local experts, who are familiar with the culture and language (except Somali) are engaged. The knowledge of the official language has a positive effect in general and increases the efficiency of the TA performance.” (Report No. 3, p. 23)*

*“Furthermore, the existing staffing level of the TSC [Teacher Service Commission] is insufficient to cover the extent of its mission in a satisfactory manner.” (Report No. 6, p. 89)*

*“Within UNFPA, a high staff turnover in 2018 and financial uncertainty regarding a next phase could potentially result in a loss of institutional memory and negatively affect relationships and continuity.” (Report No. 44, p. 7)*

*“Staff contracts are for one year or less. In some cases this has resulted in the loss of good staff (e.g., well qualified national coordinators have taken up longer term opportunities with other projects).” (Report No. 68, Pos. 470)*

*“However, no experts were employed to advise the Project in the thematic areas, particularly PES [payments for environmental service] and value chains. This “hands off” approach with thematic areas instead of a more engaged methodological advice led to a considerable isolation of experiences, given the relatively short duration of local projects.” (Report No. 35, p. 11)*

*“There is a strong feeling among the Government counterparts that a larger amount of project funding has been spent won benefits and entitlements of project staff than for project plan expenditure. In a project under national implementation modality, huge amounts of salaries and allowances (48.11% of total expenditure) to staff should have not been paid for such a small project covering one province. Instead, available National Talents should have been hired and trained by the project. Thus, the project has been blamed on its inefficient delivery. It is only after midterm evaluation and with new AWP [Annual Work Plan] in 2019 with one output, the project delivery sped up. The efficiency of the project takes a Not Satisfactory grade for its efficiency.” (Report No. 4, p. 35)*



The implementation management was assessed as efficient for the following reasons:

- Good coordination and communication with partners, donors or other stakeholders (18 cases)
- The solid management structures in place (13 cases)
- A solid M&E System in place (6 cases)
- The high flexibility and adaption to the context or to changes of framework conditions (5 cases)
- The learning from former M&E or evaluation results (2 cases)

The implementation management was assessed as not efficient for the following reasons:

- Insufficient coordination or inappropriate communication with partners or stakeholders (14 cases)
- The lack of a solid M&E System (10 cases)
- Contradictory plans or structures or unclear responsibilities among staff and/or partners (8 cases)
- An overly complex and bureaucratic steering structure (5 cases)
- Insufficient oversight and strategic steering (4 cases)
- Insufficient capacity or the absence of qualified staff (3 cases)
- Inadequate or no risk analysis and risk management (3 cases)
- Inefficient decision-making structures (2 cases)
- Projects were understaffed (2 cases)

Box 22 shows a range of examples for the assessment of implementation management.



**Box 22: Examples of reasons for the assessment related to management efficiency**

*“Harmonized reporting standards have saved time and made the reporting to the donors more efficient;” (Report No. 28, p. 47)*

*“District Programme Coordinators (assigned by the partners) were trained in results- based management, which helped increase the efficiency of programme implementation.” (Report No. 66, p. 37)*

*“The progress is monitored on an ongoing basis by the project staff through partners’ meetings, field visits to the targeted sites and stakeholders. The collected data from the monitoring are captured in the progress reports, which are rather action-based than the changed-oriented, i.e. reports mostly describe what has been done than what has been changed as the result of what has been done. It worth noting that, the progress reporting is linked to the existing set of indicators, but there is a difference in the reporting format in the two yearly progress reports (2018 and 2019), as the results of the reporting format change agreed in 2019. Thus, the first one is linked to the outputs and outcomes, while the second one - mostly to the outcomes.” (Report No. 37, Pos. 299-300)*

*“The Evaluation finds that the above LOTFA-SPM [Law and Order Trust Fund for Afghanistan-Support to Payroll Management] governance and oversight mechanisms have served to effectively manage risk during the period mid-2015 to end-2017.” (Report No. 26, p. 58)*

*“The project institutional, operational and financial system resulted in an efficient implementation of the project. UN Women institutional and management systems ensured the participation of relevant and competent CSOs in the process of the NAP [National Action Plan] implementation as well as the small projects implemented in each country.” (Report No. 29, p. 47)*

*“At times there was also a disparity between the perceived needs of Federal Member States Ministries and those positions proposed by Federal Government. To mitigate these challenges, IOM coordinated several discussions between line ministries and host institutions to clarify recruitment procedures and to discuss individual positions. Host institution directors have also been invited to stakeholder meetings, where they have a chance to voice their concerns and discuss the project implementation in detail with IOM and line ministries. Moreover, bi-annual stakeholder meetings bring all government ministries together and promote open discussions, which has so far ensured the successful implementation of activities” (Report No. 32, p. 52)*

*“In addition, IESE [Institute of Social and Economic Studies] has carried out, during the past two-three years, internal reorganisation processes to respond to management recommendations from previous evaluations and challenges identified by its staff and leadership. This has consumed time and energy of all involved but led to stronger management systems and procedures.” (Report No. 54, p.11)*



*“A common view is that the Global Governance structure is positive. PAGE [Partnership for Action for a Green Economy] provides an example of effective cooperation among UN agencies. The 5 agencies work well together, there is a good level of cooperation and genuine commitment and the Management Board is very strong. Many expressed concern however regarding the frequent turnover of staff at all levels, which slows down implementation as new staff need time to get up to speed and existing staff are required to provide repeat briefings. It was also suggested that at least one full time programme officer position should be created per PAGE agency at the global level to coordinate and implement PAGE at the country level. At present programme officers work on more than one programme making it difficult to focus on PAGE.” (Report No. 68, Pos. 411)*

*“Together with the problem of lack of an M&E system, knowledge management activities have been limited, with only few projects having published their experiences. There was an effort of systematization of experiences which was limited to 8 local projects. There was no grouping among the 8 cases in the aim of providing wider evidence for advocacy, replication, and scaling up.” (Report No. 35b, p. 11)*

*“Secondly, a management team to support the CTA [Chief Technical Adviser] in coordination, information sharing and reporting would have improved the effectiveness and efficiency of the programme and helped in avoiding communication challenges. Also, the requirement of development of customised monitoring system so track TA inputs and their effects would have helped the use of monitoring data as a management and learning tool. Now the reports are superficially linked to the higher-level goals which cannot be assessed at this point of implementation.” (Report No. 3, p. 28)*

*“WGF [Women and Girls First] is a collection of largely fragmented activities implemented by a range of partners to address the empowerment of the most vulnerable women and girls through SRHR [Sexual and reproductive health and rights] and GBV response and prevention services in different townships of selected conflict-affected states. Despite efforts to convene partners in Yangon and to set up communities of practice, familiarity with and feeling of ownership for WGF as a whole among IPs [Implementing Partner] is low and institutionalised collaboration is limited.” (Report No. 44, p. 43:)*

*“The Evaluation team considers that TASS [Technical Assistance for Competence-Based Soft Skills Development in School Education in Nepal] management structure is complex for a relatively small Technical Assistant intervention.” (Report No. 46, p. 39)*

*“Although TASS had an extensive management structure, the evaluation team considers that the TASS had limited strategic guidance. For instance, the MC and SC did not develop mitigation measures or alternative strategies although it was evident that the TASS performance suffers from the delays in curriculum development. Also, the MC [Management Committee] and SC [Steering Committee] made limited efforts to follow-up the work of the consultants and to ensure that their inputs were effectively used.” (Report No. 46, p. 39)*



*“The interviews and survey findings reveal that there are challenges in the coordination and communication of the project. The interviews note the need for enforced communication internally and externally, and it concerns all levels. The interviews further revealed that there is a need for improved communication between the individual Clusters and also Clusters with the PMT [Project Management Team]. The reason for that could be found in the fact that cluster advisors are not involved in PMT, and this results in the lack of communication between the main roles in the Project. The clusters are perceived ‘projects within the project’, with not sufficient coordination and communication for support and finding synergies.” (Report No. 48, p. 25)*

*“In addition, according to staff interviewed, the M&E systems within IOM are relatively weak as well and needs further investment, as there is no standardized system and it is up to each project do develop their database and monitoring tools, which often falls on Project Managers and the missions whose capacities may vary from one case to the other.” (Report No. 77, p. 26)*

## Impact

### Highlights of the section addressing EQ 5

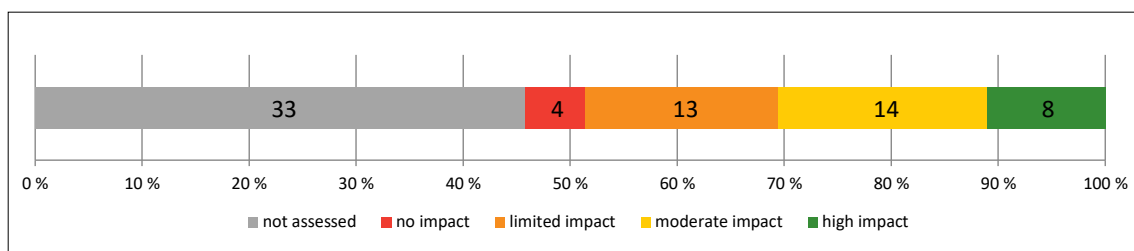
- Impact was assessed in 54 reports which accounts for 75% of all reports.
- 8 (11% of all) interventions were assessed as having a high impact and 14 (19%) a moderate impact.
- 13 (18% of all) interventions were assessed as having limited and 4 (6%) having no impact.



The OECD DAC criterion of impact refers to the positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended.

Figure 57 shows that impact was assessed in 54 reports which accounts for 75% of all reports. 8 (11% of all) interventions were assessed as having a high impact and 14 (19%) as having a moderate impact. 13 (18% of all) interventions were assessed as having limited and 4 (6%) having no impact.

**Figure 57: Interventions’ impact according to the evaluation reports (n=72)**



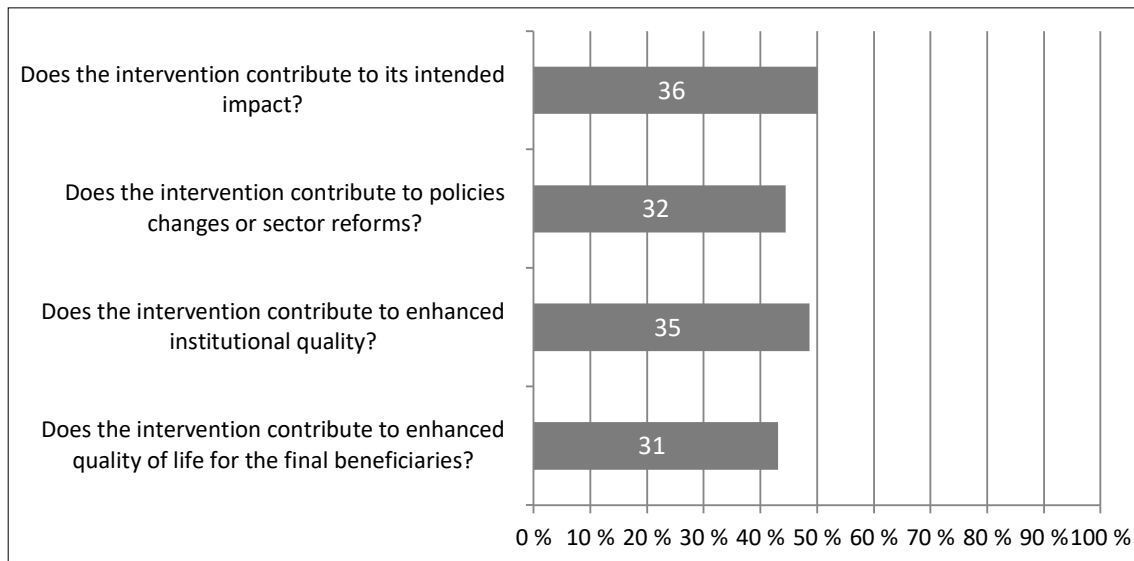
Source: own analysis of evaluation reports



Looking at different aspects of impact, Figure 58 displays that half of the reports assessed the extent to which interventions achieved their intended impacts (50% of all reports). Slightly less than half of the reports assessed the extent to which interventions contributed to policies changes or sector reforms (44%), to enhanced institutional quality (49%) and to enhanced quality of the life of final beneficiaries (43%).

Figure 59 displays how the reports assessed the interventions' impact regarding different aspects. According to the 36 reports assessing interventions' contribution to their intended impacts, 14 interventions were contributing and 8 rather contributing to their intended impact. 10 interventions did rather not, and 4 did not contribute to their intended impacts. 8 out of 32 interventions were assessed as contributing and 13 as rather contributing to policy changes or sector reforms. 5 and 6 interventions were assessed as not or rather not contributing to sector reforms, respectively. Drawing on 35 reports that assessed the contribution to enhanced institutional quality, 9 interventions contributed, 18 rather contributed and 8 rather not contributed to improved institutions. 7 out of 41 interventions were assessed as having contributed, and 14 as having rather contributed to enhancing the quality of life of the final beneficiaries. 8 interventions were assessed as having rather not and 2 as having not enhanced the situation of the final beneficiaries.

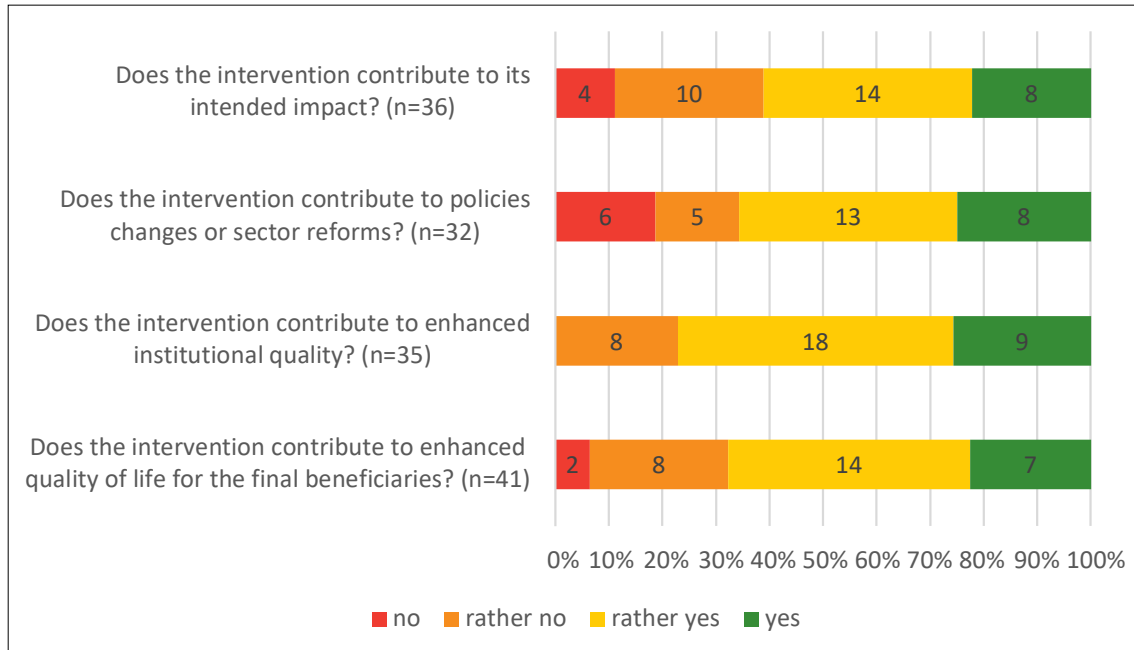
**Figure 58: Number of reports assessing different aspects of impact (n=72)**



Source: own analysis of evaluation reports



**Figure 59: Assessment of intervention quality on different aspects of impact**



Source: own analysis of evaluation reports

13 reports also found unintended positive impacts of interventions. They include:

- Unexpected economic benefits for beneficiaries, including additional income (3 cases), reduced vulnerability (1 case) and positive employment effects (1 case)
- Health benefits or improved nutrition of beneficiaries (2 cases)
- Changes in attitude or mindset of target groups or beneficiaries (2 cases)
- Better communication between or the establishment of new (business) connections among stakeholders (3 cases)

Further unintended positive impacts that only appeared once are related to:

- More sustainable land use through changes in agricultural practices
- Improved communication and access to information through new IT infrastructure
- Arousing the interest of diaspora experts to engage in the development of their country of origin



Box 23 shows a range of examples for the assessment of positive unintended impacts.

**Box 23: Examples of unintended positive impacts**

*“One of the objectives of the CELIM project was to increase nutrition of its beneficiaries through promoting inclusion of Moringa in the diet. Although some beneficiaries did indeed change their diet to include Moringa, many others sold the Moringa products like leaves and powder to make some income. Once this was clear the project tried to further promote this through contacts with a Lusaka company interested in Moringa but this eventually failed.” (Report No. 8b, p. 37)*

*“Besides the monetary income, beneficiaries also highlighted the improvement in the quality of their (work) life: people feel healthier when they inhale less smoke, they have to work less hard and don’t have to get up at night to monitor the charcoal burning process.” (Report No. 35, p. 15)*

*“PFG [People Participation in Improvement of Forest Governance and Poverty Alleviation in Vietnam Project ] also has unintended positive impacts. Its SmartPhone Application (“PFG app”) is leading to income increases from forestry, agricultural and aquacultural products due to better information regarding prices and communications with potential buyers via social media. Moreover, being able to use the internet for many core members, especially women, improves their status within communities - especially when villagers approach them to have more information about forest management and market information for key agricultural and aquacultural products.” (Report No. 21b, p. 46)*

*“One of the key unintended impact of the MIDA FINNSOM was linked to the diaspora-national engagement approach that the phase IV adopted. The project has contributed to improvements in work ethic and attitude change, which is very crucial for any profession especially for health service delivery. As per a national expert that is benefiting from the project, the diaspora-national expert work engagement is effective not just on the skills transfer, but also improving the national staff work ethic and contributing toward attitudinal change. The national health experts reported that the day to day coaching, capacity building and mentoring by the diaspora experts helped them to work hard and increase their professionalism in delivering the health services in their community” (Report No. 2, p. 33)*

Five reports also found unintended negative impacts of interventions. They appeared only in one case each and were related to:

- Fraud and corruption
- Environmental and health risks connected to fish pond farming
- Conflicts within communities
- Extension of the duration of a teacher education programme led to gaps in the availability of graduated teachers; as a result, not enough new teachers were available in the transition years.
- A new law on prosecuting violence against women and new shelters for women rather increased stigmatisation and exclusion of affected women in/from their communities





Box 24 shows all assessments of unintended negative impacts.

**Box 24: Examples of unintended negative impacts**

**Fraud and corruption**

*“Villagers lacked information on many aspects related to the Project and the service provided. This had the unexpected negative impact of leaving them at the mercy of unscrupulous individuals and organisations. This lack of information expressed in many ways, such as:*

- *The moment the funds were approved or even during the previous project, politicians or individuals who claimed to have political influence started visiting some of the villages. In some places, they requested funding contributions and the payment of fees to finance surveys and technical studies of doubtful use. For example, the evaluation team was told that in El Chaparral, a member of parliament (Doña Gladis) requested 60,000 lp (USD 2,300) in exchange for assistance to receive electricity.*
- *Once the village was already selected for electrification, some of the companies subcontracted by New Mark to install the equipment threatened villagers that if they did not offer unskilled labour and housing and food for free, they would not receive electricity.*
- *Once electricity is installed, electricity users receive a first bill for 500 lp (approximately 19 USD) to pay for the installation of the meter. This amount is to be reimbursed if the user decides to cancel their account, subject to keeping the receipt as a proof of payment. Villagers neither knew this, nor what it was that they were paying for with this first bill.” (Report No. 13, p. 26)*

**Environmental/Health risks**

*“While the economic benefits are significant and accessible to women and ethnic minorities as well, the field visit to Sapa provided evidence that there are environmental and fish health risks that are not yet effectively checked. The profitability of trout farming is increasingly known around the area and families construct tanks without consulting any authorities. For example, about 30 farms along Ban Khoang watershed have no water treatment facilities. Farms draw water from the creek and discharge it untreated back to the creek which is a significant risk both from the water quality and fish health point of view.” (Report No. 30b, p. 37)*

**Conflicts within communities**

*“Among the negative impacts, some conflicts among local people and disagreements on the use of water were mentioned. Also, some respondents were concerned that the growth of fish farming has been too fast and uncontrollable. However, few were concerned about the water quality.” (Report No. 30b, p. 39)*



### **Extension of the duration of teacher education programme leads to gaps in graduated teachers**

*“At the same time, the way the revised curriculum has been implemented so far results in lower outputs of graduates at the degree level, due to capacity issues at the EC [Education College] that will not be fully addressed by 2022. Moreover, without changing the teacher career development structure, it is unlikely that the required number of teacher students will specialize in primary education. In terms of the capacity to deliver the required number of quality teachers to solve the teacher shortage, the impact of STEM [Strengthening Pre-Service Teacher Education in Myanmar Project] Phase II might be negative in the first years. With the change from a two-year to a four-year degree program, for two years no students will graduate from the ECs. In addition, EC and hostel capacities in the EC limit the intake of students as the student teachers occupy the facilities for four instead of two years. This means that also the number of new teachers being qualified by the new degree program will be less than is currently qualified by the program that is phased out. Hence, the number of qualified teachers according to international standards, will initially not increase due to STEM Phase II. These two factors (lack of teacher policies and capacity of ECs) are likely to result in a considerable reduction of the number of qualified teachers in primary education, which may have to be compensated by increased enrolment of the shorter teacher education program” (Report No. 70, p. 45)*

### **Increase stigmatisation and exclusion of affected women in/from their communities**

*“However, the strongest evidence comes from the 2017 Samuel Hill research on the EVAW [Elimination of Violence Against Women] Commissions. With regard to legislation the research found that while most women interviewed stated knowing that the EVAW law exists (that is to say, had a general understanding that this law existed to protect them from violence and abuse), there was very limited impact because of social stigma, norms and restrictions. Additionally, women reported severe flaws in the implementation of the law. Flaws related to implementation included uncertainty over the sensitive handling of their cases, corruption (fear / perception that the perpetrator’s family could pay officials to decide in their favour) and as a result of corruption, the fear of backlash for bringing their domestic case in the public. With regard to protection the research found that the most unanimous finding across every single person interviewed for this study is the negative perception associated with women’s shelters. Shelters are considered by many as places where prostitution occurs, where women may be raped by the police and where women indulge in illicit activities. Many women interviewed strongly believe that once a woman goes to a shelter, even if she is guaranteed safety, she can never come back into society or join her family due to the stigma attached to these centres and misconceptions of shelters.” (Report No. 27, p. 28)*

## Sustainability

### Highlights of the section addressing EQ 5:

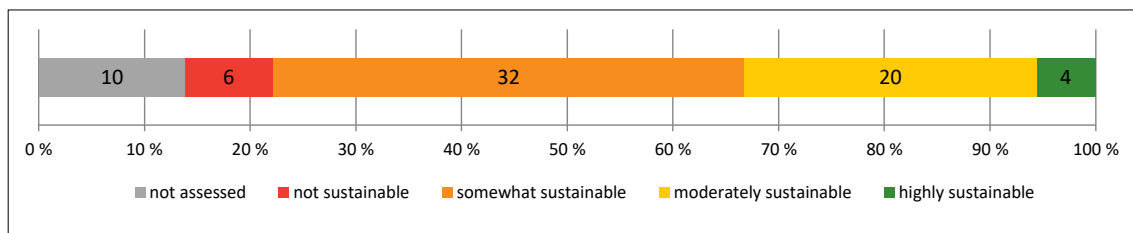
- Figure 36 shows that sustainability was assessed in 62 reports which account for 86% of all reports.
- 4 (6% of all) interventions were assessed as highly sustainable and 20 (28%) as moderately sustainable.
- 32 (44% of all) interventions were assessed as somewhat while 6 (8%) interventions were assessed as not sustainable.



The OECD DAC criterion of sustainability refers to the extent benefits of an activity are likely to continue after donor funding has been withdrawn.

Figure 60 shows that sustainability was assessed in 62 reports which accounts for 86% of all reports. 4 interventions (6% of all) were assessed as highly sustainable and 20 as moderately sustainable (28%). 32 interventions (44% of all) were assessed as somewhat and 6 interventions (8%) as not sustainable.

**Figure 60: Interventions' sustainability according to the evaluation reports (n=72)**



Source: own analysis of evaluation reports

Looking at different aspects of sustainability, Figure 61 displays that more than half of the reports assessed whether the target groups have the financial means (56% of all reports) and the capacity (67%) to make the intervention sustainable. Whether the benefits of interventions will last was also assessed in 67% of all reports. To what extent the implementing partners have the financial means and the capacity to make the intervention sustainable and whether the enabling environment allows sustainability was assessed in about one-third of all reports (28%, 33% and 31%, respectively).

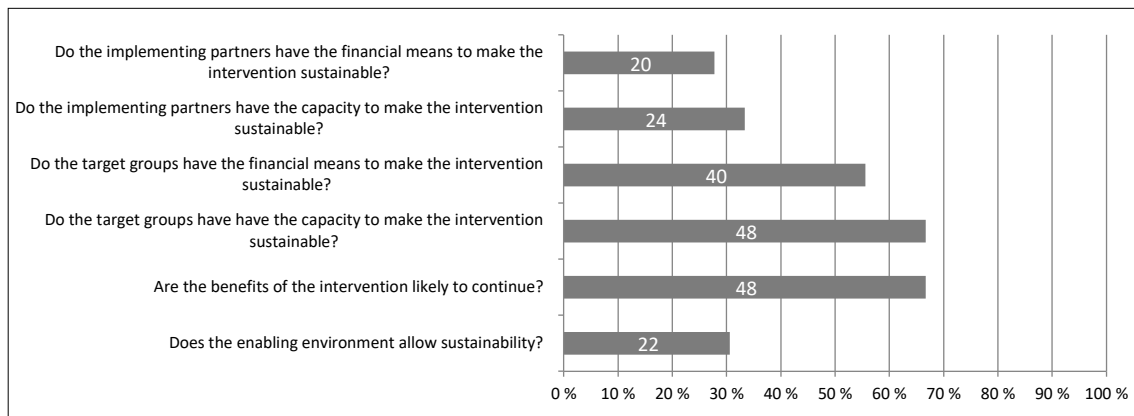
Figure 62 displays how the reports assessed the interventions' sustainability regarding the different aspects. According to 20 reports assessing the financial means of the implementing partners, only in 2 cases, implementing partners have, and 4 rather have the financial resources to make the intervention sustainable. In 7 cases each, partners were assessed to not or rather not have the required financial means. That partners have or rather have the capacity to make the intervention sustainable was assessed for 3 and 8 cases out of 24 reports, respectively. For 10 cases, partners were assessed to rather not and for 3 to not have the required capacity. According to 40 reports assessing the financial means of the target groups, only in 2 cases, target groups have, and in 7 cases, they rather have the financial resources to make the intervention sustainable. In 13 cases, target groups were assessed to rather not have and in 18 cases to not have the required financial means. That target groups have or rather have the capacity to make the intervention sustainable



was assessed for 5 and 20 cases out of 48 reports, respectively. For 10 cases, partners were assessed to rather not and for 3 to not have the required capacity.

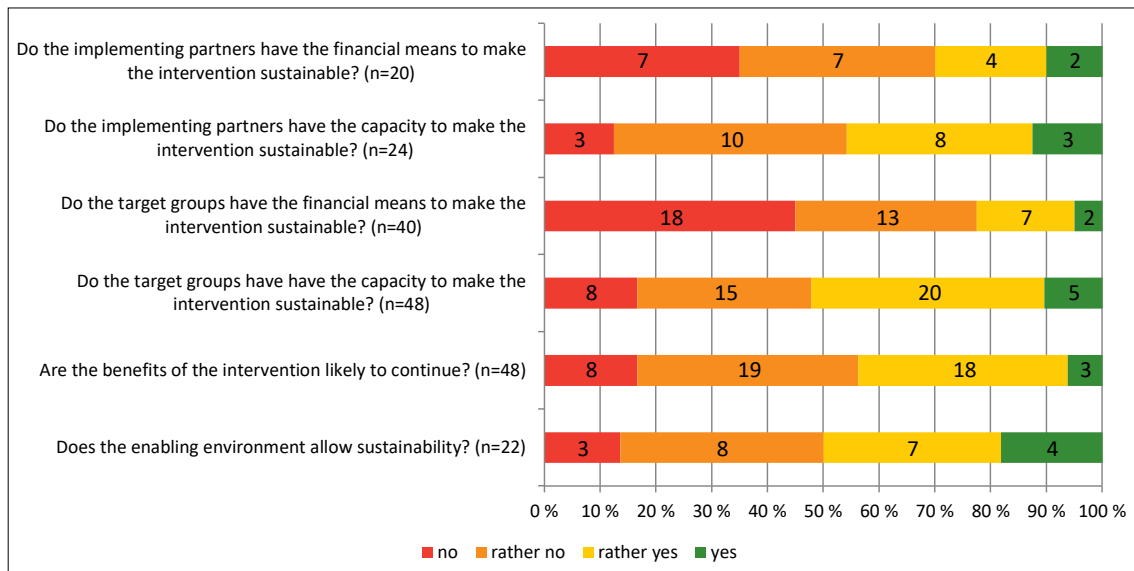
For 3 out of 48 interventions, reports assessed that the benefits are likely, and for 18 interventions, benefits are rather likely to continue. For 19 interventions, the assessment sees it as rather not likely and for 8 interventions as not likely that the benefits will continue. 22 reports further indicate that 4 interventions have an enabling, 7 a rather enabling, 8 a rather not enabling and 3 a not enabling environment for sustainability.

**Figure 61: Number of reports assessing different aspects of sustainability (n=72)**



Source: own analysis of evaluation reports

**Figure 62: Assessment of intervention quality on different aspects of sustainability**



Source: own analysis of evaluation reports



The reasons why evaluators assess that the benefits of intervention are likely to continue to include:

- Improved legal or regulatory frameworks or the successful institutionalisation of results (12 cases)
- Increased capacities of target groups, stakeholders, or beneficiaries (10 cases)
- Structures, mechanisms, approaches, or equipment is in place that ensures sustainability (10 cases)
- High commitment and engagement of government counterparts (5 cases)
- Good market demand for products or successful value chain integration (4 cases)
- High ownership among target groups or beneficiaries (3 cases)
- Capable partner organisations continue with project activities on their own (3 cases)
- Existence of a sound exit strategy (2 cases)

The reasons why evaluators assessed that the benefits of an intervention are not likely to continue to include:

- Insufficient financial resources of partners/target groups or ongoing dependency on donor funding (15 cases)
- Insufficient capacity or dependency on external technical assistance (10 cases)
- Lack of equipment and infrastructure or insufficient means for long-term maintenance of provided equipment (7 cases)
- Insufficient institutionalisation of results or lack of complementary changes in legal or regulatory frameworks (5 cases)
- Lack of ownership among target groups or beneficiaries (4 cases)
- The duration of intervention was too short (2 cases)
- High staff turnover (2 cases)

Box 25 shows a range of examples for the assessment of sustainability of interventions.



**Box 25: Examples of reasons related to the assessment of sustainability of interventions**

*“The noted outcomes of the Gender and HR Academy in Egypt, in terms of improved understanding of the role of human resources, a better understanding of labour laws and encouragement of women employment are also likely to be sustained as private sector employees become more aware of the importance of these issues in improving staff satisfaction, limiting staff turnover and ultimately increasing production and profitability.” (Report No. 1, p. 31)*

*“Livelihood support in Dak Lak is very likely to sustain after the project completes because it has a good support throughout value chain: planting new trees will be supported by potential buyers, buyers reach out to plantation owners to buy timber, timber processing cooperative established to increase value of timber through initial processing (peeling bark, cutting into required logs with certain length and width).” (Report No. 21b, p S. 58)*

*“The villager skills in using computers or smart phones to access information on market prices on agricultural, aquacultural, and forestry products will last and will be shared widely among community members. As a result, future funding for replacing the information kiosk computers or the smart phones might not be a problem, especially when the livelihood models become successful.” (Report No. 21b, p. 59)*

*“Having used low-cost / low-maintenance solutions for the early warning systems and disaster response plans, communities should be able to maintain this equipment as confirmed during the field visits. The example of Solomon Islands, where the community has replaced batteries themselves, is an encouraging example of this.” (Report No. 24, p. 46)*

*“The revised content of the teacher education curriculum can be seen as the main benefit of STEM [Strengthening Pre-Service Teacher Education in Myanmar Project] Phase II. This outcome is sustainable; so far, the Ministry has implemented the first year of the curriculum and essentially moved past the point of no return. The first cohort of teacher students already started their studies in the new curriculum. This fact will prove to be a motivating factor for the ECs [Education College] and the MoE [Ministry of Education] to keep developing the content and train the EC staff on the remaining years of the new teacher curriculum until 2022” (Report No. 14, p. 50)*

*“By using an inclusive approach for the community work, such as the establishment of National Coordination Teams involving key local and government stakeholders, broader ownership has also been created. The main implementing partner for the community work, IFRC [International Federation of Red Cross and Red Crescent Societies ], can be expected to continue to be involved at community level through its permanent presence in the countries in the form of the national RC societies, although there is a risk that the communities were the project is completed will be “forgotten” and attention will go only to replicating the model to other communities” (Report No. 24, p. 46)*

*“RIA-1 [Research Institute for Aquaculture] has strong ownership on the methods and approaches. In the field mission it became evident that the staff and management of RIA-1 feel that they have been on the driver’s seat. The cooperation was started based on Vietnamese interests and priorities expressed by MARD [Ministry of Agriculture and Rural Development] in 2001. The support from Finnish partners has helped them to set up a new centre and build up RIA-1’s capacity to the extent that it has been sharing fish breeding expertise within MARD, especially with RIA-2 and RIA-3. RIA-1 also provides training to extension staff working under MARD.” (Report No. 30b, p. 47)*



*“Since ECDC [Eden Centre for Disabled Children] is a strong local NGO with existing capacities in ECI [Early Childhood Intervention] and ECCD [Early Childhood Care and Development] and leading the advocacy with the government for improving disability laws and policy, ECDC is able to replicate and scale up ECI program within and outside the project areas in Myanmar.” (Report No. 64, p. 19)*

*“It is difficult to imagine any significant and sustainable changes being implemented without an actor exercising strong national ownership. It may be questioned whether the project in all countries will have a national partner that is interested and capable of exercising ownership, pushing for change, building capacity and generally taking the project forward.” (Report No. 7, p. 39)*

*“The Programme aims at building capacity of CEEG [Centre for Economic and Management Studies] and DEEF [Directorate of Economic and Financial Studies]. In DEEF this is mainly done by providing new or updated tools (macro model, MOZ MOD) and by training the staff. However, the financial crisis and Government’s measures to cut expenses undermine the sustainability of this capacity.” (Report No. 10, p. 48)*

*“There is wide consensus that the contributions are not enough to exit due to the magnitude of the gap in health professionals in Somaliland and heavy dependence on the diaspora and nationals recruited and deployed through the project. The evaluation revealed that, though the different phases of MIDA FINNSOM built technical capacity in place, financial resource are lacking at the MoHD [Ministry of Health and Development] to sustain the gains achieved so far.” (Report No. 2, p. 41)*

*“The National and Provincial Assemblies do not have the conditions in place for the project products and results to continue after the intervention has finished. Despite their interest, manifested by all people interviewed, financial (no money), institutional, legal (no formal relation between national and provincial assemblies), technical (few technical staff, reliance on outside experts to the detriment on in-house talent or potential talent) and political aspects (allegiance to party) are limitations. The project benefits are unlikely to continue unless the Parliamentary Centre of Study and Training and the provincial Technical Secretariats take the driver’s seat.” (Report No. 11, p. 21)*

*“Activities supporting, for example, forest landscape management were only started during SUFORD-SU. Also, village forestry was not practiced for more than a decade before it was newly incorporated in the SUFORD-SU. The implementation period has been short, and although capacities have been built, the beneficiaries struggle to carry on independently.” (Report No. 17a, p. 100)*

*“Yet the health sector is still largely dependent on external aid, with government expenditure accounting for a small fraction. MSA [Marie Stopes Afghanistan] clearly operates in a highly unstable and insecure environment. Reviewing the national budget allocated for health by the Government of Afghanistan, funding from the government of Finland is a lifeline for the project, and without this support, the program will collapse which will negatively affect the health of many communities.” (Report No. 42, p. 46)*



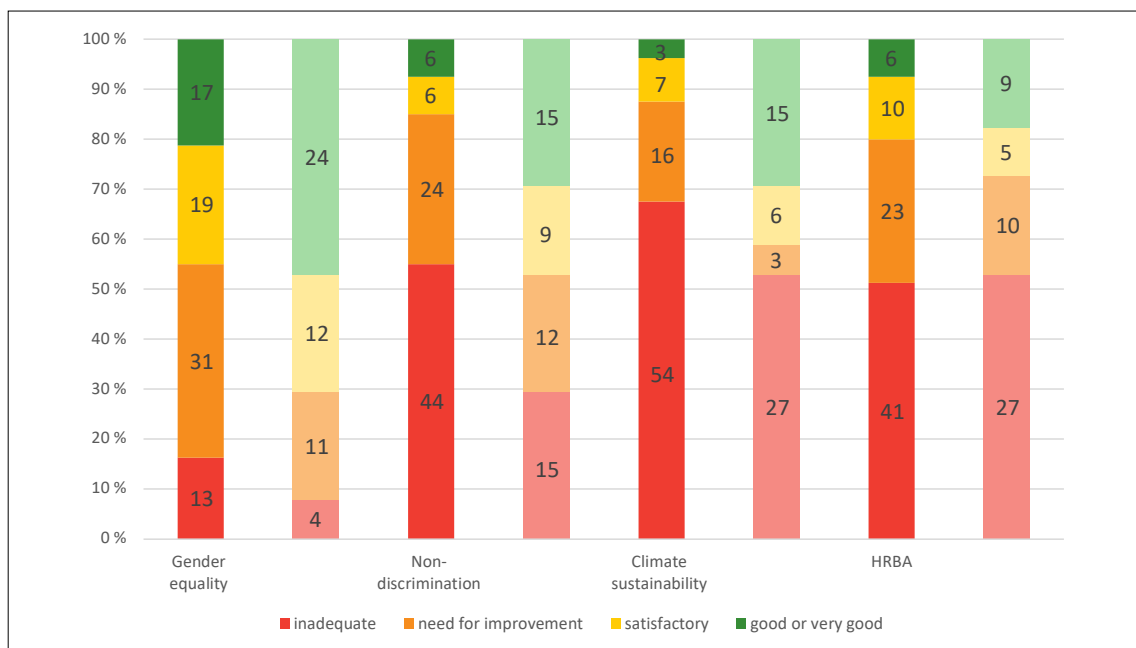
# Annex 13: Further Details on CCOs and HRBA

- No robust differences cannot between the previous and the recent metaevaluation can be assessed due to differences in methodology.
- Appropriate integration of CCOs and HRBA in analytical parts of the reports suggest deterioration since the last metaevaluation. Integration in context analyses slightly improved.
- Increased evaluation standards may have an influence.

Figure 63 displays the consideration of CCOs under the recent metaevaluation (darker colours) against the previous (light colours). A descriptive comparison suggests deterioration as the overall quality of CCO consideration is lower. However, the comparison is potentially biased due to differences in the assessment.

The recent metaevaluation has been carried out after evaluation standards have increased, and it used a partially different methodology than the previous one to assess gender equality, as requested by the MFA. Therefore, any differences between the old and new data cannot be considered robust. In addition, the quality assessments revealed that while the appropriate integration of CCOs and HRBA in the analytical parts of the reports decreased since the previous metaevaluation, there is a slight improvement regarding consideration in the context analyses.

**Figure 63: Comparison of recent (n=80) and previous (n=51) metaevaluation findings on CCOs and HRBA**



Source: own analysis of evaluation reports from recent and previous metaevaluations





In the remainder, **detailed results on the quantitative keyword analysis** are presented. The results underline that the terms “**gender**” and “**women**” are mentioned in nearly all reports in the sample (98% and 92% of the reports, respectively). Their overall frequencies are also higher than for the other keywords; see Table 8 below.

**Table 8: Frequencies of gender equality related keywords in evaluation reports**

KEYWORD	FREQUENCY	# OF DOCUMENTS	DOCUMENTS %
women	2050	60	92
gender	1172	64	98
male	386	50	77
girl*	361	24	37
female	269	45	69
GBV	187	8	12
men*	178	39	60
matern*	118	10	15
SRH	105	3	5
GESI	73	6	9
sex*	73	26	40
reproductive	59	5	8
SRHR	54	2	3
man	5	4	6
forced marriage	2	2	3
FGM	0	0	0
genital	0	0	0

Source: own analysis of 72 evaluation reports



The coloured cells’ group keywords related to maternity, sexual and reproductive health and rights. Only between 3% (3) and 15% (10) of the reports discuss the topic but with relatively high frequencies. These are most likely projects of which objectives are geared around this topic. A similar trend can be observed for gender-based violence 12% (8) of the evaluations include the word, overall frequency being 187.

The keyword analysis shows that when the word “sex” or its derivatives are mentioned, the text refers to disaggregated data in 22 % (16) of those contexts. All other contexts (78%, 57) relate to some thematic matters (such as sexual violence, sexual and reproductive health, sexual education, or sexual exploitation).

**Non-discrimination** was assessed using the keywords listed in Table 9 below.



**Table 9: Percentage of reports that mention selected keywords related to non-discrimination**

KEYWORD	FREQUENCY	# OF DOCUMENTS	DOCUMENTS %
inclusi*	785	55	85
disab*	430	27	42
excluded	120	23	35
discriminat*	90	27	42
inequality	42	18	28
exclusion	14	7	11
unequal	6	4	6

Source: own analysis of 72 evaluation reports



Among these keywords, “inclusive”, “disability”, “excluded” (and their derivatives, when relevant) appear most frequently in the evaluation reports. In 24 % (187) of the contexts in which the word “inclusive” (and its derivatives) was mentioned, the discussion is about inclusive education. No other topics emerged as clearly from the sample. Similarly, for the word “disability” (and its derivatives), there are no clear groups to which the contexts could be classified. In contexts in which the word “excluded” appears, 64% (77) refer to women.

**Climate sustainability** was also assessed using the keywords listed in Table 10 below.

**Table 10: Percentage of reports that mention selected keywords related to climate sustainability**

KEYWORD	FREQUENCY	# OF DOCUMENTS	DOCUMENTS %
climat*	464	32	49
resilien*	84	20	31
mitigation	82	27	42
DRR	60	4	6
NDC	56	4	6
emission*	40	13	20
disaster risk	21	4	6
greenhouse	18	5	8
climate adaptation	6	3	5
paris agreement	6	2	3
fossil fuels	3	2	3
sink*	3	3	5
fossil fuel	2	1	2
nationally determined contribution	1	1	2

Source: own analysis of 72 evaluation reports



Among these keywords, “climate”, “resilience”, “mitigation” (and their derivatives, when relevant) appear most frequently in evaluation reports. The word “climate” is used, for example, in the context of climate sustainability (10% of the contexts, 49 out of 80) and climate impacts (8% of the contexts, 38). For the word “resilience”, no clear trends emerge in terms of the context. However, in 22% (18) of the contexts when the word “mitigation” is mentioned, it is used

together with “adaptation”.

**HRBA** was assessed using the keywords listed in Table 11 below.

**Table 11: Percentage of reports that mention selected keywords related to HRBA**

KEYWORD	FREQUENCY	# OF DOCUMENTS	DOCUMENTS %
rights	1017	54	83
HRBA	93	11	17
duty bearers	33	13	20
rights holders	20	10	15
rights-holder*	3	3	5
duty-bearer*	2	1	2

Source: own analysis of 72 evaluation reports

Among these keywords, the term “rights” appears most frequently in evaluation reports. Expectedly, in 62% (629) of the contexts, the word “human” precedes the word “rights”. In 12% (122) of the contexts, the word “women” are mentioned. Otherwise, no clear groups of words emerge.

**Additional keywords** that are present in MFA’s Theories of Change (MFA, 2020b) and the Guideline for Cross-Cutting Objectives in the Finnish Development Policy and Cooperation (MFA, 2020a) but do not fall under any of the CCOs are presented in Table 12 below.

**Table 12: Percentage of reports that mention additional keywords from MFA’s Theories of Change**

KEYWORD	FREQUENCY	# OF DOCUMENTS	DOCUMENTS %
child	492	39	60
poor	271	46	71
poverty	214	34	52
youth	158	33	51
boy	119	10	15
young	104	31	48
old	46	20	31
adolescent	39	7	11
biodiversity	12	4	6
no harm	10	4	6
elderly	9	6	9
leave no one behind	0	0	0

Source: own analysis of 72 evaluation reports



## Annex 14: Further quality assessments by survey respondents

### Highlights of the section:

- The quality of evaluation reports is quite positively assessed.
- A bit less than half of the respondents assesses the quality of reports as (very) good. About another half assess them as satisfactory.
- The same holds true for the quality of ToRs and their appropriateness.
- Evaluation processes and quality of executive summaries are of lesser quality; about a third assess them as (very) good.




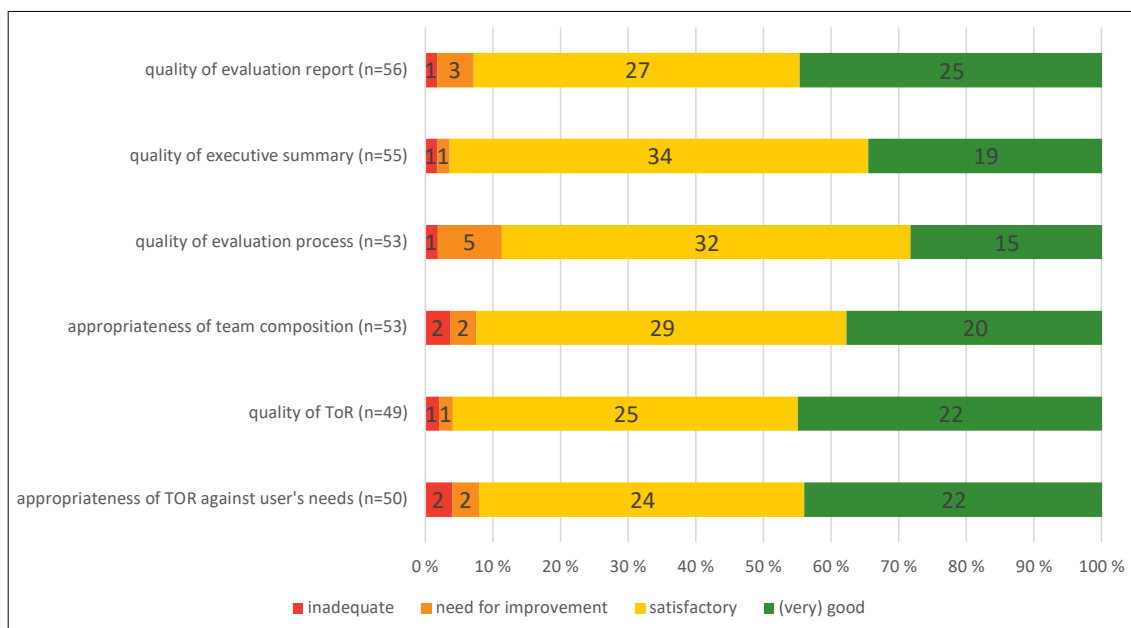
Figure 64 shows respondents' quality assessment frequency distributions on the evaluation reports and underlying evaluation processes. Again, the same 4-point rating scale was used as in the other sections, ranging from 1 = "inadequate" to 4 = "(very) good". Overall, the perceptions were positive, with a large majority of the cases being assessed of at least satisfactory quality. However, interviewees were proportionally more critical of the quality and provided recommendations  to improve the usefulness of decentralised evaluations (see Chapter 6.4.2).

Figure 64 shows that in 25 out of 56 cases, the quality of the evaluation report  was assessed as (very) good (45%), in 27 cases as satisfactory (48%), in 3 cases as in need for improvement (7%), and in one case as inadequate. The overall pattern is similar for the appropriateness of team composition and a bit lower for the quality of the executive summaries and the quality of the evaluation process. Yet, in each of these quality dimensions, roughly a third of the cases were assessed as being of (very) good quality (20, 38%, 19, 35% and 15, 28%, respectively), about a bit more than half was perceived as being of satisfactory quality (29, 55%, 34, 62%, and 32, 60%, respectively) and some of the cases were perceived as in need for improvement (2, 4%, 1, 2%, and 5, 10%, respectively) or as of inadequate quality (2, 4%, 1, 2%, and 1, 2%, respectively).

The quality of the ToR  was similarly assessed as the quality of the evaluation reports. A bit less than half of the cases (22 out of 49) was assessed as (very) good (45%), about half as satisfactory (25, 51%), one case as in need for improvement and one case as inadequate. Furthermore, the ToR were largely considered appropriate against the needs of the specific key stakeholder groups (i.e., staff at MFA headquarters and embassies and implementers (Figure 64)). Out of 49 ToR assessments, a bit less than half was perceived as of (very) good quality (22, 45%), about half as satisfactory (24, 49%) and only a single case as in need for improvement (2, 4%) and as inadequate (2, 4%).



**Figure 64: Quality assessment of evaluations and ToRs by survey respondents (case level)**



Source: own data



## Annex 15: Statistical Tests

Non-parametric statistical Mann-Whitney tests identifying group differences were applied to test for statistically significant differences on the OECD DAC criteria as assessed in the content assessment for the different PPAS. These tests did not include PPA1 as the sample size was too small. The tests for PPA2 to PPA4 revealed no differences in overall intervention quality among different PPAs (see Table 13, 14 and 15).

**Table 13: Mann-Whitney tests PPA2 vs other PPAs**

VARIABLE	$n_1$ PPA2	$n_2$ OTHER	$W$	$P$
1.2 How do the evaluators assess the relevance of the intervention in the evaluation report?	22	50	469.500	0.132
2.2 How do the evaluators assess the effectiveness of the intervention in the evaluation report?	22	50	479.500	0.771
3.2 How do the evaluators assess the efficiency of the intervention in the evaluation report?	22	50	229.500	0.058
4.2 How do the evaluators assess the impact of the intervention in the evaluation report?	22	50	162.500	0.712
5.2 How do the evaluators assess the sustainability of the intervention in the evaluation report?	22	50	302.000	0.052
9.2 How do the evaluators assess the coherence of the intervention in the evaluation report?	22	50	111.500	0.393
Overall Intervention Quality (Aggregate of the six DAC criteria above)	22	50	569.000	0.821

Source: own data



**Table 14: Mann-Whitney tests PPA3 vs other PPAs**

VARIABLE	$n_1$ PPA3	$n_2$ OTHER	$W$	$P$
1.2 How do the evaluators assess the relevance of the intervention in the evaluation report?	17	55	242.500	0.087
2.2 How do the evaluators assess the effectiveness of the intervention in the evaluation report?	17	55	283.000	0.096
3.2 How do the evaluators assess the efficiency of the intervention in the evaluation report?	17	55	242.500	0.837
4.2 How do the evaluators assess the impact of the intervention in the evaluation report?	17	55	139.500	0.584
5.2 How do the evaluators assess the sustainability of the intervention in the evaluation report?	17	55	293.500	0.291
9.2 How do the evaluators assess the coherence of the intervention in the evaluation report?	17	55	53.500	0.486
Overall Intervention Quality (Aggregate of the six DAC criteria above)	17	55	375.000	0.221

Source: own data

**Table 15: Mann-Whitney tests PPA4 vs other PPAs**

VARIABLE	$n_1$ PPA3	$n_2$ OTHER	$W$	$P$
1.2 How do the evaluators assess the relevance of the intervention in the evaluation report?	16	56	363.000	0.149
2.2 How do the evaluators assess the effectiveness of the intervention in the evaluation report?	16	56	424.000	0.305
3.2 How do the evaluators assess the efficiency of the intervention in the evaluation report?	16	56	220.000	0.471
4.2 How do the evaluators assess the impact of the intervention in the evaluation report?	16	56	150.000	0.909
5.2 How do the evaluators assess the sustainability of the intervention in the evaluation report?	16	56	300.500	0.352
9.2 How do the evaluators assess the coherence of the intervention in the evaluation report?	16	56	109.500	0.531
Overall Intervention Quality (Aggregate of the six DAC criteria above)	16	56	551.500	0.162

Source: own data



Ordinary least square (OLS) regression analysis revealed no differences in overall quality neither for commissioner nor other aspects (see Table 16).

**Table 16: OLS Regression Analysis**

	OLS
(Intercept)	1.902
	(1.269)
ToR quality	0.181
	(0.197)
Evaluation budget (log)	0.029
	(0.122)
Evaluation duration (months)	0.008
	(0.025)
Independent consultant(s) (dummy)	-0.294
	(0.181)
MFA commissioned (dummy)	-0.040
	(0.189)
Ordered geographical scope	-0.036
	(0.116)
<i>N</i>	33
<b>Model Fit:</b>	
<i>F</i> (6, 26)	1.031
<i>p</i>	0.428
<i>R</i> <sup>2</sup>	0.192
Adjusted <i>R</i> <sup>2</sup>	0.006
Note: Standard errors are heteroskedasticity robust. Ordered geographical scope: (sub-) national, national, regional, and international interventions. *** <i>p</i> < 0.01; ** <i>p</i> < 0.05; * <i>p</i> < 0.1. Source: own analysis of evaluation reports and ToRs.	

Source: own statistics based on analysis of reports

## Survey Data

Mann-Whitney test were conducted to test whether the survey responses differed between different groups of key stakeholders prior to aggregating the data. The tests were conducted, including the cases for which survey responses were available for different key stakeholder groups. The first comparison tested whether HQ and embassy staff differed in their survey responses in the sample of 32 cases in which data was available for these two perspectives. Mann-Whitney tests revealed no statistically significant differences (comparison 1, Table 17). The second comparison, testing for statistical differences between evaluation implementers and embassy staff in 34 cases, was carried out in the same way and did also reveal no differences (comparison 2, Table 18). As the case coverage between MFA headquarters staff and implementers was very low, this comparison could not be conducted. Instead, we tested in a final comparison whether MFA staff (including headquarters and embassy staff) differed in their perspective from implementers in the 38 cases for which data from both groups were available (comparison 3, Table 19).





**Table 17: Mann-Whitney test on survey responses of MFA headquarters staff vs embassy staff**

VARIABLE	$n_1$ HQ	$n_2$ EMBASSY	$W$	$P$
Q1.2: ...the timing for this evaluation?	16	16	97.000	0.776
Q1.2: ...the timeliness of informal results delivery (FCR workshop, validation meeting) against MFA HQ needs?	16	16	76.000	0.456
Q1.2: ...the timeliness of written, formal results delivery (approved report) against MFA HQ needs?	16	16	72.500	0.356
Q1.4: ...the relevance of the recommendations of this evaluation for MFA HQ staff?	16	16	106.000	0.449
Q1.4: ...the recommendations being realistic?	16	16	71.000	0.301
Q1.4: ...the implementation of relevant and realistic recommendations?	16	16	67.000	0.209
Q1.4: ...the learnings from this evaluation?	16	16	85.500	0.795
Q1.13: ...the overall usefulness of this evaluation?	16	16	107.500	0.060
Q1.13: ...the overall quality of this evaluation report?	16	16	97.000	0.209
Q1.13: ...the overall quality of the corresponding executive summary?	16	16	82.000	0.518
Q1.13: ...the overall quality of this evaluation process?	16	16	69.000	0.188
Q1.13: ...the appropriateness of the evaluation team composition?	16	16	46.500	0.431
Q1.13: ...the overall quality of the corresponding ToR?	16	16	48.500	0.327
Q1.13: ...the appropriateness of the ToR with respect to MFA HQ needs?	16	16	60.500	0.304

Source: own data



**Table 18: Mann-Whitney test on survey responses of implementers vs MFA embassy staff**

VARIABLE	$n_1$ IMPLEMENTER	$n_2$ EMBASSY	$W$	$P$
Q1.2: ...the timing for this evaluation?	17	17	91.500	0.557
Q1.2: ...the timeliness of informal results delivery (FCR workshop, validation meeting) against MFA HQ needs?	17	17	96.500	0.718
Q1.2: ...the timeliness of written, formal results delivery (approved report) against MFA HQ needs?	17	17	92.000	0.573
Q1.4: ...the relevance of the recommendations of this evaluation for MFA HQ staff?	17	17	128.000	0.287
Q1.4: ...the recommendations being realistic?	17	17	109.500	0.854
Q1.4: ...the implementation of relevant and realistic recommendations?	17	17	122.000	0.437
Q1.4: ...the learnings from this evaluation?	17	17	117.500	0.584
Q1.13: ...the overall usefulness of this evaluation?	17	17	92.000	0.559
Q1.13: ...the overall quality of this evaluation report?	17	17	98.500	0.773
Q1.13: ...the overall quality of the corresponding executive summary?	17	17	83.500	0.711
Q1.13: ...the overall quality of this evaluation process?	17	17	102.500	0.823
Q1.13: ...the appropriateness of the evaluation team composition?	17	17	82.500	0.678
Q1.13: ...the overall quality of the corresponding ToR?	17	17	60.500	0.509
Q1.13: ...the appropriateness of the ToR with respect to MFA HQ needs?	17	17	53.500	0.439

Source: own data



**Table 19: Mann-Whitney test on survey responses of implementers vs all MFA staff**

VARIABLE	$n_1$	$n_2$	$W$	$P$
	IMPLEMENTER	MFA STAFF		
Q1.2: ...the timing for this evaluation?	19	19	127.500	0.554
Q1.2: ...the timeliness of informal results delivery (FCR workshop, validation meeting) against MFA HQ needs?	19	19	144.000	0.999
Q1.2: ...the timeliness of written, formal results delivery (approved report) against MFA HQ needs?	19	19	138.000	0.829
Q1.4: ...the relevance of the recommendations of this evaluation for MFA HQ staff?	19	19	167.500	0.413
Q1.4: ...the recommendations being realistic?	19	19	150.500	0.841
Q1.4: ...the implementation of relevant and realistic recommendations?	19	19	156.000	0.691
Q1.4: ...the learnings from this evaluation?	19	19	170.000	0.371
Q1.13: ...the overall usefulness of this evaluation?	19	19	137.500	0.813
Q1.13: ...the overall quality of this evaluation report?	19	19	135.500	0.750
Q1.13: ...the overall quality of the corresponding executive summary?	19	19	120.000	0.999
Q1.13: ...the overall quality of this evaluation process?	19	19	135.000	0.985
Q1.13: ...the appropriateness of the evaluation team composition?	19	19	118.500	0.966
Q1.13: ...the overall quality of the corresponding ToR?	19	19	88.000	0.710
Q1.13: ...the appropriateness of the ToR with respect to MFA HQ needs?	19	19	89.500	0.719

Source: own data

**Table 20: Spearman Correlation between Report Quality (QA) and Overall Usefulness (Survey)**

OVERALL QUALITY (COMPONENT3)			
	$rs$	$N$	$P$
Overall quality (component 1)	-.069	53	0.621

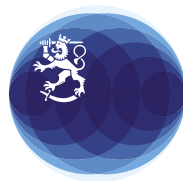
Source: own data

**Table 21: Spearman Correlation between Intervention Quality (CA) and Intervention Quality (Survey)**

OVERALL INTERVENTION QUALITY (COMPONENT 3)			
	$rs$	$N$	$P$
Overall intervention quality (component 2)	.244	36	0.152

Source: own data

# **METAEVALUATION OF MFA'S PROJECT AND PROGRAMME EVALUATION IN 2017-2020**



**Ministry for Foreign  
Affairs of Finland**