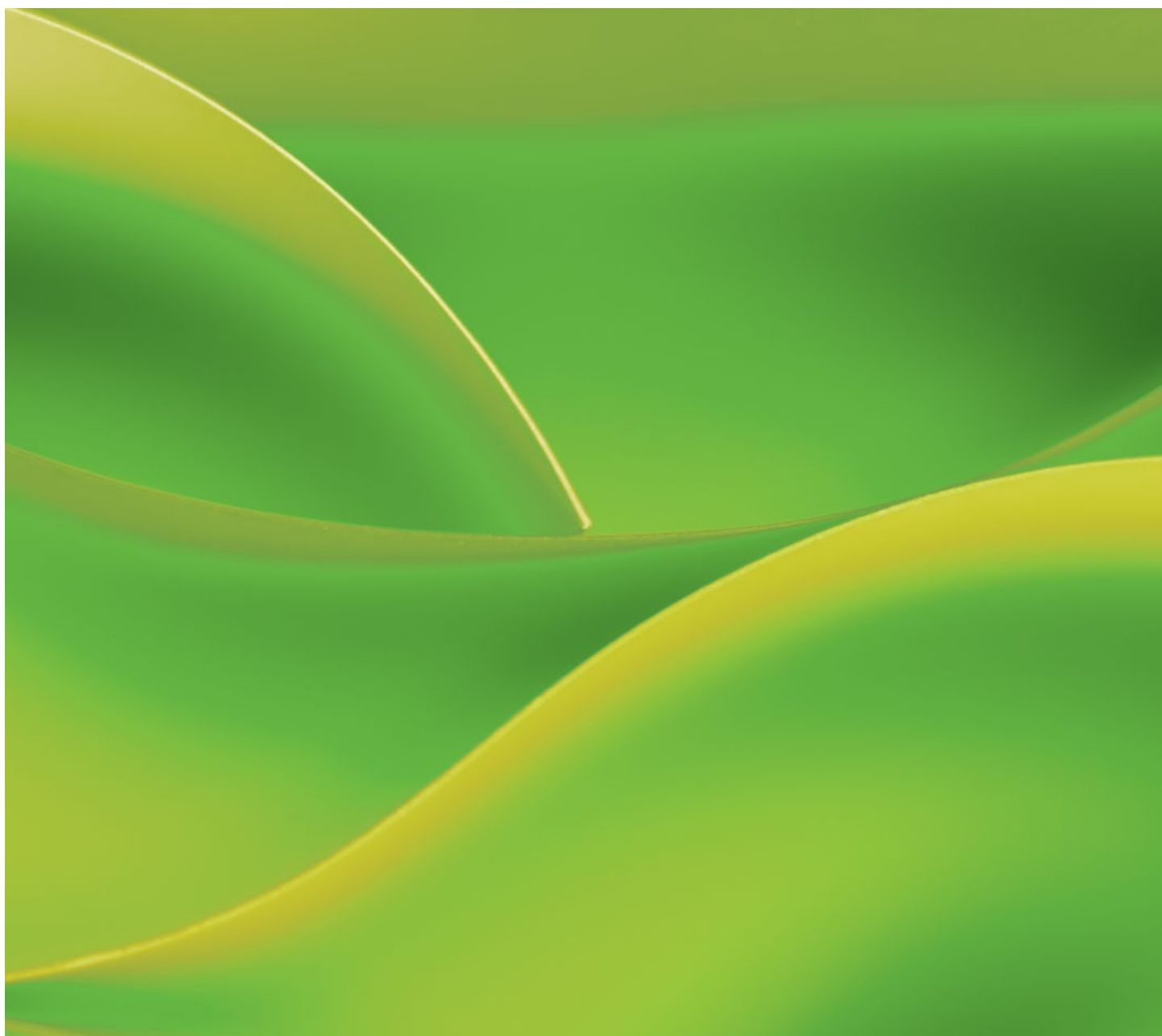




EVALUATION

Meta-Evaluation of Project and
Programme Evaluations in 2012–2014



Evaluation on Finland's Development Policy and Co-operation

2015/3

EVALUATION

META-EVALUATION OF PROJECT AND PROGRAMME EVALUATIONS IN 2012-2014

Geraldine Cooney and Katrina Rojas
Mariane Arsenault
Kelly Babcock



2015/3

This evaluation was commissioned by the Ministry for Foreign Affairs of Finland to the Universalialia Management Group Ltd. 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.

© Ministry for Foreign Affairs of Finland 2015

This report can be downloaded through the home page of the Ministry for Foreign Affairs
<http://formin.finland.fi/developmentpolicy/evaluations>

Contact: EVA-11@formin.fi

ISBN 978-952-281-407-4 (pdf)

ISBN 978-952-281-408-1 (e-pub)

ISSN 2342-8341

Cover design and layout: Innocorp Oy/Milla Toro

CONTENTS

ACKNOWLEDGEMENTS.....	IX
ACRONYMS AND ABBREVIATIONS.....	X
GLOSSARY	XI
TIIVISTELMÄ.....	1
REFERAT	2
ABSTRACT	3
YHTEENVETO.....	4
SAMMANFATTNING	7
SUMMARY.....	11
KEY FINDINGS, CONCLUSIONS AND RECOMMENDATIONS	14
1 INTRODUCTION.....	18
1.1 Overview.....	18
1.2 Evaluation Context.....	18
1.3 Organisation of report.....	19
2 APPROACH, METHODOLOGY AND LIMITATIONS	20
2.1 Approach.....	20
2.2 Methodology	20
2.3 Limitations	21
3 FINDINGS: QUALITY OF EVALUATION REPORTS.....	22
3.1 Overview.....	22
3.2 Overall Quality of Reports	22
3.3 Quality of Reports by Category of Review Tool.....	24
3.4 Summary.....	29
4 FINDINGS: QUALITY OF TERMS OF REFERENCE.....	30
4.1 Overview.....	30
4.2 Overall Quality of TORs	30
4.3 Quality of TORs, by Categories of the Review Tool	32
4.4 Summary.....	37
5 FINDINGS: EFFECTIVENESS OF FINLAND'S DEVELOPMENT CO-OPERATION	38
5.1 Overview.....	38
5.2 Overall Ratings of Development Co-operation	38
5.3 Ratings of Development Co-operation, by Category of the Review Tool.....	40
6 FINDINGS: MFA EVALUATION MANAGEMENT PRACTICES AND CAPACITY DEVELOPMENT	58
6.1 Overview.....	58
6.2 Capacity Building Services	58
6.3 Findings on the Evaluation Management Cycle.....	62

7 OVERALL OBSERVATIONS ON MFA EVALUATIONS AND META-EVALUATIONS	66
7.1 Overview.....	66
7.2 Contextual issues	66
7.3 Linkage between the Quality of TORs and Reports.....	68
7.4 Adequacy of Evaluation Resources	70
7.5 Meta-Evaluation Process.....	71
7.6 Comparison of current and previous meta-evaluations	72
7.7 Summary.....	72
8 CONCLUSIONS.....	74
8.1 Introduction.....	74
8.2 Quality of MFA Decentralised Evaluation Reports and TORs	74
8.3 Assessment of Finland’s Development Co-operation	76
8.4 Overall observations on MFA evaluations and meta-evaluations	77
9 RECOMMENDATIONS.....	80
REFERENCES.....	87
THE EVALUATION TEAM	88
Annex 1 Terms of Reference.....	90
Annex 2 People Interviewed	95
Annex 3 Documents Consulted	96
Annex 4 Methodology.....	103
Annex 5 Meta-Evaluation Matrix.....	113
Annex 6 List of Reports Reviewed: Sample and Sub-Sample.....	115
Annex 7 Quality Assessment: Review Tool for Evaluation Reports	120
Annex 8 Quality Assessment: Review Tool for TORs	126
Annex 9 Quality Assessment: Summary Ratings of Evaluation Reports.....	132
Annex 10 Quality Assessment: Summary Ratings of TOR	136
Annex 11 Development Co-operation Assessment: Review Tool.....	140
Annex 12 Development Co-operation Assessment: Summary of Ratings	154
Annex 13 Consultation Protocol on Capacity Building Services and Evaluation Management Cycle.....	158
Annex 14 Training Offered	160
BOXES	
Box 1 OECD/DAC Definition of Meta-evaluation.....	19
Box 2 Terminology	19
Box 3 Evaluation Report Review Tool	22
Box 4 Observations on Very Good and Excellent Reports.....	22
Box 5 Examples of Very Good Overall Ratings	23
Box 6 Evaluand - Example of a Very Good Rating.....	24
Box 7 Purpose - Examples of Very Good Ratings	25
Box 8 Methods - Examples of Very Good Ratings	26

Box 9	Substance - Examples of Very Good Ratings.....	28
Box 10	Presentation - Examples of Very Good Ratings.....	29
Box 11	TORs Review Tool Categories.....	30
Box 12	Example of a “Good” Overall Rating.....	31
Box 13	Example of a “Good” Rating for Overall Background Information	32
Box 14	Example of a “Very Good” Rating Overall for Rationale, Purpose, and Objectives	33
Box 15	Example of a “Poor” Rating Overall for Evaluation Scope	33
Box 16	Example of a “Good” Rating Overall for Evaluation Questions.....	34
Box 17	Example of a “Poor” Rating Overall for Methodology	35
Box 18	Example of a “Good” Rating Overall for Evaluation Process and Management Structure.....	35
Box 19	Example of a “Good” Rating Overall for Deliverables	36
Box 20	Example of a “Very Good” Rating Overall on the Logic of TORs.....	37
Box 21	Development Co-operation Review Tool	38
Box 22	Examples of reviewers’ comments on Effectiveness.....	42
Box 23	Comparison with 2012 Meta-Evaluation	43
Box 24	Example of reviewers’ comment on Poverty Reduction	43
Box 25	Comparison with 2012 Meta-Evaluation	44
Box 26	Example of reviewers’ comment on Relevance.....	45
Box 27	Comparison with 2012 Meta-Evaluation	46
Box 28	Example of reviewers’ comment on Efficiency	47
Box 29	Comparison with 2012 Meta-Evaluation	48
Box 30	Example of reviewers’ comment on Sustainability.....	49
Box 31	Example of reviewers’ comments on HRBA.....	51
Box 32	Incorporating a Human Rights-Based Approach in development projects	51
Box 33	Comparison with the 2012 Meta-Evaluation.....	52
Box 34	Example of reviewers’ comment on Gender Equality.....	52
Box 35	Example of reviewers’ comment on Reduction of Inequality.....	54
Box 36	Example of reviewers’ comment on Climate Sustainability	55
Box 37	Comparison with the 2012 Meta-Evaluation	56
Box 38	Key MFA Evaluation Capacity Needs in 2010	58
Box 39	MFA Evaluation Training since 2010: An Overview	61
Box 40	Changes made by EVA-11 to Evaluation Training Sessions in response to participant feedback ...	62
Box 41	European Commission 2014 report on the uptake of strategic evaluation recommendations ...	65
Box 42	Evaluability assessment	67
Box 43	Practices in other agencies	80
Box 44	Practices in other agencies	81
Box 45	Quality Assurance and Evaluability Practices in other agencies	82
Box 46	Evaluability	83
Box 47	Practices in other donor agencies	83
Box 48	Evaluation Capacity Building Needs Assessments	84
Box 49	The use of evaluation results	85
Box 50	Questions about limited coverage in MFA evaluations	85

TABLES

Table 1	Description and feedback on EVA-11 capacity building services	59
Table 2	EVA-11 Annual Expenditures for Capacity Development (in Euros).....	60
Table 3	Overall quality rating of TORs vs evaluation reports (based on 57 sets of TORs/reports)	69
Table 4	Comparing the overall quality rating of TORs with the corresponding evaluation reports (based on 57 sets of TORs/reports).....	69
Table 5	Quality of reports compared to adequacy of evaluation budget (based on 41 reports).....	71

FIGURES

Figure 1	Summary of Ratings – Main Questions on Quality of Reports	23
Figure 2	Summary of Ratings – Main Questions on Quality of TORs.....	31
Figure 3	Overall ratings of main criteria for development co-operation	39
Figure 4	Effectiveness – Ratings of main criterion on 5-point scale (n=34)	40
Figure 5	Comparison between Mid-term and Final evaluations for Q1: Achievement of expected results and project purposes	41
Figure 6	Effectiveness – Ratings of sub-criteria on 3-point scale (n=34).....	41
Figure 7	Overall objectives – Ratings of main criterion on 5-point scale (n=34).....	43
Figure 8	Overall objectives – Ratings of sub-criterion on 3-point scale (n=34).....	43
Figure 9	Relevance – Ratings of main criterion on 5-point scale (n=34)	44
Figure 10	Relevance – Ratings of sub-criteria on 3-point scale (n=34)	45
Figure 11	Efficiency – Ratings of main criterion on 5-point scale (n=34).....	46
Figure 12	Efficiency – Ratings of sub-criteria on 3-point scale (n=34)	47
Figure 13	Sustainability – Ratings of main criterion on 5-point scale (n=34).....	48
Figure 14	Comparison of Mid-term vs. Final evaluations for Q5: Overall sustainability	48
Figure 15	Sustainability – Ratings of sub-criteria on 3-point scale (n=34).....	49
Figure 16	HRBA – Ratings of main criterion on 5-point scale (n=34).....	50
Figure 17	HRBA – Ratings of sub-criteria on 3-point scale (n=34).....	50
Figure 18	Gender equality – Ratings of main criterion on 5-point scale (n=34).....	52
Figure 19	Gender equality – Ratings of sub-criteria on 3-point scale (n=34).....	53
Figure 20	Reduction of inequality – Ratings of main criterion on 5-point scale (n=34).....	53
Figure 21	Reduction of inequality – Ratings of sub-criteria on 3-point scale (n=34).....	54
Figure 22	Climate sustainability – Ratings of main criterion on 5-point scale (n=34).....	54
Figure 23	Climate sustainability – Ratings of sub-criteria on 3-point scale (n=34).....	55
Figure 24	Risk management – Ratings of sub-criteria on 3-point scale (n=34)	56
Figure 25	Evaluation Management Cycle	63

ACKNOWLEDGEMENTS

This report was prepared by Universalia Management Group Ltd., a Canadian consulting firm that has evaluated development co-operation projects for the past 35 years. A team of evaluation consultants, reviewers, peer-reviewers, and editors collaborated on this final product between October 2014 and March 2015.

Universalia would like to thank MFA and particularly EVA-11 representatives for the support provided throughout the meta-evaluation. In addition, we would like to thank those who provided feedback on the findings and recommendations contained in the draft report in March 2015. Their feedback and comments informed the final version of this report.

ACRONYMS AND ABBREVIATIONS

AHA	MFA's electronic project management system
CEDAW	Convention to Eliminate all Forms of Discrimination Against Women
DFATD	Department of Foreign Affairs, Trade and Development (Canada)
DFID	Department for International Development (UK)
EU	European Union
EVA-11	Development Evaluation Unit, Ministry for Foreign Affairs
HRBA	Human Rights-based Approach
ITT	Instructions to Tenderers
MFA	Ministry for Foreign Affairs (Finland)
MTE	Mid-Term Evaluation
NGO	Non-Governmental Organisation
OECD-DAC	OECD Development Assistance Committee
QA	Quality Assurance
Sida	Swedish International Development Agency
TOC	Theory of Change
TOR	Terms of Reference

GLOSSARY

Definitions from the MFA 2013 Evaluation Manual and MFA Bilateral Manual

Decentralised evaluation

Decentralised evaluation is the responsibility of the MFA departments and units that are in charge of the development co-operation projects and programmes in specific countries, regions or international institutions. These evaluations include appraisals, mid-term, final and ex-post evaluations as an integral part of project and programme cycle management. The embassies participate and facilitate evaluations at country level.

Expected results

Concrete results that have been taken into use by immediate/final beneficiaries.

Mid-term evaluation

Performed towards the middle of the implementation of the programme and typically focuses on issues that are relevant for good performance in remaining years of implementation. It also contains recommendations on the continuation of support.

Final evaluation

Assesses the achievements of the programme in relation to its set of objectives at the end of the implementation period. Final evaluations also summarise lessons learned that may be useful for future programmes.

Ex-post evaluation

Carried out after the programme has been completed and provides evidence on the longer-term impact and sustainability of the programme.

Overall objective

Overall broader development objective to which the project will contribute (Note: In OECD terminology, this is defined as impact).

Project purposes

Specific objectives that should be achieved by the end of the project; i.e. the project's expected end-result. (In OECD terminology, these are referred to as outcomes.)

Review

Conducted when a need arises to further analyse the information collected through monitoring, address problems faced during implementation, and discuss appropriate management action to support effective implementation.

OECD-DAC Criteria for Evaluating Development Assistance

Source

<http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

Effectiveness

A measure of the extent to which an aid activity attains its objectives. In evaluating the effectiveness of a programme or a project, it is useful to consider the following questions:

-
- To what extent were the objectives achieved / are likely to be achieved?
 - What were the major factors influencing the achievement or non-achievement of the objectives?

Efficiency

Efficiency measures the outputs - qualitative and quantitative - in relation to the inputs. It is an economic term which signifies that the aid uses the least costly resources possible in order to achieve the desired results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted.

When evaluating the efficiency of a programme or a project, it is useful to consider the following questions:

- Were activities cost-efficient?
- Were objectives achieved on time?
- Was the programme or project implemented in the most efficient way compared to alternatives?

Impact

The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators. The examination should be concerned with both intended and unintended results and must also include the positive and negative impact of external factors, such as changes in terms of trade and financial conditions.

When evaluating the impact of a programme or a project, it is useful to consider the following questions:

- What has happened as a result of the programme or project?
- What real difference has the activity made to the beneficiaries?
- How many people have been affected?

Relevance:

The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor.

In evaluating the relevance of a programme or a project, it is useful to consider the following questions:

- To what extent are the objectives of the programme still valid?
- Are the activities and outputs of the programme consistent with the overall goal and the attainment of its objectives?
- Are the activities and outputs of the programme consistent with the intended impacts and effects?

Sustainability

Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable.

When evaluating the sustainability of a programme or a project, it is useful to consider the following questions:

- To what extent did the benefits of a programme or project continue after donor funding ceased?
- What were the major factors which influenced the achievement or non-achievement of sustainability of the programme or project?

TIIVISTELMÄ

Metaevaluoinnissa arvioidaan evaluointiraporttien ja niiden tehtäväkuvauksen laatua sekä Suomen kehitysyhteistyön ja ulkoasiainministeriön (UM) kehitysevaluoinnin hallinnon ja kapasiteetinkehittämispalveluiden laatua. Evaluointi perustuu kolmen arviointityökalun käyttöön, dokumentteihin ja UM:n henkilöstön konsultointeihin.

Suurin osa evaluointiraporteista ja tehtäväkuvauksista sai arvosanan hyvä, mutta vain harvat arvioitiin erinomaisiksi. Suomen kehitysyhteistyö sai positiivisimmat arviot hankkeiden tarkoituksenmukaisuudesta eli relevanssista sekä tuloksellisuudesta. Heikoimmat arviot liittyivät tehokkuuteen ja kestävyteen.

Evaluointien laatuun vaikuttavia muuttujia ministeriössä ovat vahvan evaluointikulttuurin puuttuminen sekä puutteelliset tulosperustaiset johtamis- ja hallintokäytänteet ja laadunvarmistus evaluointisyklin eri vaiheissa, riittämättömät evaluointibudjetit ja joidenkin kehityspolitiikan osa-alueiden rajallinen huomiointi evaluoinneissa. UM:n johtajat, virkamiehet ja konsultit tarvitsevat tietojen päivitystä ja uuden mekanismin, kuten esimerkiksi puitesopimuksen pätevien evaluaattorien hankintaa helpottamaan. Institutionaalisoitujen työkalujen ja ohjeistusten puuttuminen rajoittaa metaevaluointien hyödynnettävyyttä eri aikoina.

Avainsanat: metaevaluointi, EVA-11, hajautettu evaluointi, ulkoasiainministeriö, evaluointihallinnon käytänteet

REFERAT

Metautvärderingen uppskattade standarden på utvärderingsrapporter och direktiv, Finlands utvecklingssamarbete och utrikesministeriets (UM) utvärderingsledning och tjänster för kapacitetsutveckling. Den baserades på tre översiktsverktyg, en översikt av dokument och konsulteringar med UM:s-personal.

De flesta av utvecklingsrapporterna och direktiven höll en hög standard, men några ansågs vara enastående. Standarden på Finlands utvecklingssamarbete utvärderades som varande mest positivt när det gällde projektens ändamålsenlighet och resultat och minst positivt i förhållande till deras effektivitet och hållbarhet.

Variabler som påverkar utvärderingars standard omfattar en avsaknad av en stark utvärderingskultur inom UM, brist på effektiva resultatbaserade ledningsmetoder och kvalitetsgarantier genom hela utvärderingscykeln, otillräckliga utvärderingsbudgetar och begränsat iakttagande av vissa UM-policies. UM-chefer, -funktionärer och -konsulter behöver vidareutbildning och en mekanism som en ramöverenskommelse skulle kunna förbättra kontrakterandet av kvalificerade utvärderare. Bristen på institutionaliserade verktyg och riktlinjer begränsar metautvärderingens användbarhet över en tidsperiod.

Nyckelord: metautvärdering, decentraliserad utvärdering, Utrikesministeriet (UM), ledningspraxis i utvärdering

ABSTRACT

The meta-evaluation assessed the quality of evaluation reports and their Terms of Reference (TORs), Finland's development co-operation, and evaluation management and capacity development services of the Ministry for Foreign Affairs of Finland (MFA). It was based on three review tools, a review of documents, and consultations with MFA staff.

The quality of most evaluation reports and TORs was good, but few were considered excellent. The quality of Finland's development co-operation was rated most positively in terms of the relevance and effectiveness of projects and least positively in relation to their efficiency and sustainability.

Variables that affect the quality of evaluations include the absence of a strong evaluation culture within MFA, lack of effective results-based management practices and quality assurance throughout the evaluation cycle, inadequate evaluation budgets, and limited attention to some MFA policies. MFA directors, officers and consultants require skill upgrading and a mechanism such as a framework agreement could improve the contracting of qualified evaluators. The lack of institutionalised tools and guidelines limits the utility of meta-evaluations over time.

Keywords: meta-evaluation, decentralised evaluation, Ministry for Foreign Affairs, evaluation management practices

YHTEENVETO

Tarkoitus ja tavoitteet: Suomen ulkoasiainministeriön (UM) kehitysevaluoinnin yksikkö (EVA-11) tilasi metaevaluoinnin tammikuun 2012 ja elokuun 2014 välillä teetetyistä hajautetuista hankearvioinneista. Tavoitteina oli arvioida evaluointiraporttien, niiden tehtävänkuvausten, sekä Suomen kehitys yhteistyön, kehitysevaluoinnin hallinnon ja kapasiteetin kehittämispalveluiden laatua.

Metaevaluoinnin toteutti Universal Management Group Ltd:n konsulttitiimi lokakuun 2014 ja maaliskuun 2015 välisenä aikana. Metaevaluointi perustui evaluointityökaluun, joka kehitettiin arvioimaan 57 hajautetun evaluointiraportin ja tehtävän kuvauksen laatua sekä Suomen kehitys yhteistyön vaikuttavuutta 34 laadultaan hyvän UM:n tilaaman evaluointiraportin pohjalta. Kehitysevaluoinnin hallintokäytänteiden ja kapasiteetin kehittämispalveluiden laatua arvioitiin UM:n henkilöstön konsultoinnin ja asiaankuuluvien dokumenttien pohjalta.

Tulosten ja päätelmien yhteenveto

Raporttien laatu: Ulkoasiainministeriön hajautettujen evaluointiraporttien laatua arvioitaessa yleisimmin annettu arvosana oli hyvä tai parempi. Yleensä raportit tuottivat tilaajilleen näyttöön perustuvia tuloksia ja hyödyllisiä suosituksia, vaikka selkeää kohdeyleisöä suosituksille ei aina ollut. Merkittävimmät heikkoudet liittyivät ihmisoikeusperustaisuuden ja läpileikkaavien tavoitteiden epätasaiseen huomiointiin sekä joihinkin menetelmällisiin rajoitteisiin.

Tehtävänkuvausten laatu: Useimmat evaluointien tehtävän kuvaukset olivat hyvin kirjoitettuja ja ilmaisivat evaluoinnin vaatimukset selkeästi. Huomattavimpia heikkouksia olivat puutteet metodologian kuvauksessa, liian suuri evaluointikysymysten määrä sekä UM:n keskeisiin painopisteisiin, kuten ihmisoikeusperustaisuuteen ja läpileikkaaviin tavoitteisiin liittyvien kysymysten puute.

Suomen kehitys yhteistyö: Yleisesti Suomen kehitys yhteistyötä arvioitiin positiivisimmin tarkoituksenmukaisuuden ja tuloksellisuuden suhteen, ja vähiten positiivisesti tehokkuuteen ja kestävyys liittyen. Silloin kun riskien hallinta oli mainittu, se arvioitiin heikoksi. Läpileikkaavien tavoitteiden ja ihmisoikeusperustaisuuden arviot olivat vaihtelevia.

Arviointien laatuun vaikuttavat muuttujat: Vaikka hyvin kirjoitetun evaluoinnin tehtävän kuvauksen ja laadukkaan evaluointiraportin välillä on havaittavissa korrelaatiota, ei ole selvää, onko kysymys syy-seuraussuhteesta vai myös mahdollisista muista muuttujista. Muista arviointiraporttien laatuun vaikuttavista muuttujista havaittiin mm. seuraavia: vahvan arviointikulttuurin puute, tulosperustaisten hallintokäytänteiden puutteet, henkilöstön vaihtuvuus, vaihtuvan kehityspolitiikan vaatimat muutokset ja riittämättömät evaluointi-

budjetit. (Metaevaluointitiimi totesi, että 43 % tehtävänkuvauksista oli riittämättömät resurssit toimeksiannon laajuuteen nähden.)

Metaevaluointien vertailua: Institutionalisoituneiden metaevaluointiohjeistuksen ja -työkalujen puute rajoittaa nykyisellään UM:n metaevaluointien mahdollista vaikuttavuutta, tehokkuutta ja hyödynnettävyyttä. Eri ajanjaksojen metaevaluointien tulosten vertailukelpoisuus lisäisi tehokkuutta ja vähentäisi uusien työkalujen ja ohjeistusten kehittelyyn sekä testaukseen käytettävää aikaa. Se mahdollistaisi myös paremman tulosten vertailun metaevaluointien välillä sekä laadun parantamiseen tähtäävien toimien vaikutusten tarkastelun ajan kuluessa (esim. kapasiteetinkehittämisen lisääminen tai muutokset evaluointien hallintokäytänteissä).

Laadunvarmistus: Vaikka suurin osa tehtävänkuvauksista ja evaluointiraporteista oli laadultaan hyviä tai parempia, lähemmin tarkasteltuna ne saivat enimmäkseen arvosanan hyvä, ei erittäin hyvä tai erinomainen. Jos ministeriö haluaa parantaa evaluointiraporttien ja tehtävänkuvauksen yleistä laatua, on laadunvarmistus huomioitava paremmin koko evaluointisyklin aikana - jo hankkeen suunnitteluvaiheesta alkaen (on varmistettava esimerkiksi läpileikkaavien tavoitteiden sisällyttäminen, tiedonkeruujärjestelmä, kumppaneiden osallistaminen, budjetin ja tehtävänkuvauksen laajuuden yhdenmukaisuus jne.). Tehtävänkuvauksia kehitettäessä on myös huomioitava riittävä taloudellisten ja muiden resurssien allokointi evaluoinnin seurantaan, toteutukseen ja tulosten hyödyntämiseen.

Evaluointien hallinnon kapasiteetti: Konsultoitaessa UM:n virkamiehiä evaluoinnin hallintoon ja kapasiteetin kehittämispalveluihin liittyvissä kysymyksissä, ilmeni, että UM:n henkilöstön puutteet evaluointiin liittyvissä tiedoissa ja taidoissa häiritsevät evaluaattoreiden valintaa ja evaluointien hallinnointia.

Vaikka kehitysevaluoinnin yksiköllä EVA-11:lla on olemassa puitesopimuksia muita toimeksiantoja varten, vastaava järjestelmä puuttuu hajautetuista hankkearvioinneista. Suurin osa haastatelluista UM:n edustajista ilmaisi, että puitesopimuksen kaltainen järjestelmä lisäisi hankintaprosessien tehokkuutta vähentämällä hallinnon työtaakkaa ja tarjoamalla lisätakeita evaluaattoreiden laadusta.

Evaluointikapasiteetin kehittäminen: ministeriö suunnittelee evaluointikapasiteetin kehittämispalveluiden kattavaa arviointia tulevana vuonna. Tämä metaevaluointi sisälsi niistä alustavan katsauksen. Siinä havaittuja aukkoja voidaan selvittää tulevassa jatkoselvityksessä.

Evaluointien seuranta ja jatkotoimet: Siinä missä joissakin evaluointiraporteissa ja tehtävänkuvauksissa ilmenneitä puutteita voidaan korjata päivittämällä UM:n johdon, virkamiesten ja konsulttien taitoja, UM:n johdolla on kuitenkin tärkeä rooli evaluoinnin laatustandardien sekä evaluointiraporttien tulosten ja suositusten hyödyntämisen seurannassa. Metaevaluoinnissa käsiteltyjen evaluointiraporttien ja tehtävänkuvauksen puutteet paljastivat, että toistaiseksi vain harvoilla UM:n johdon edustajilla on ollut tilaisuus hankkia evaluointiprosessin johdossa tarvittavia tietoja ja taitoja. Haastatteluiden mukaan tilannetta vaikeuttaa osaltaan myös vahvan arviointikulttuurin puuttuminen ministeriöstä.

Kehityspolitiikan ja direktiivien huomioiminen: Osan ministeriön toimeksi antamien ja hajautettujen hankkeiden arvioinnin laatua laski joidenkin kehityspolitiikan osa-alueiden ja direktiivien laiminlyöminen. Jotkut niistä ovat uusia tai muuttuneet ajan myötä (esim. ihmisoikeusperustaisuus esiteltiin virallisesti vasta vuoden 2012 kehityspoliittisessa ohjelmassa), jotkut taas ovat hyvin tunnettuja ja puutteet on nostettu esille jo aiemmissa metaevaluoinneissa (esim. läpileikkaavat tavoitteet ja sukupuolten tasa-arvo). Näiden seikkojen systemaattisen huomioinnin laiminlyönti laski kehitysyhteistyön vaikuttavuuden arviota. Nämä rajoitteet ovat EVA-11:n mandaatin ulkopuolisia ja kuvaavat ministeriön hajautettujen hankkeiden hallintokäytänteiden puutteita kehityspolitiikan ja direktiivien toteutuksessa.

Suositukses

1. Parantaakseen ministeriön evaluointien käyttöä jatkossa, EVA-11:n tulisi kehittää muutosteoria linjaamaan ja linkittämään evaluointien odotetut tulokset näyttöön perustuvaan päätöksentekoon.
2. Metaevaluointien pitkäaikaisempaa hyödyntämistä ajatellen, EVA-11:n tulisi harkita säännöllisen metaevaluoinnin sekä siinä käytettävien arviointivälineiden ja ohjeistusten institutionalisoimista.
3. Mikäli ministeriössä tahdotaan parantaa hajautettujen evaluointien ja tehtävänkuvausten laatua, on syytä lisätä huomiota ja resursseja laadunvarmistukseen kaikissa evaluointisyklin vaiheissa.
4. Hankintaprosessia ja pätevien evaluaattoreiden rekrytointia tehostaakseen ministeriön tulisi harkita puitesopimusten solmimista hajautettuja evaluointeja varten. Varmistaakseen riittävän pätevien evaluaattoreiden määrän, ministeriön tulisi mahdollistaa sopimukset sekä yksittäisten asiantuntijoiden että suurempien yritysten kanssa.
5. EVA-11:n tulisi viipymättä toimeenpanna suunnittelemansa selvitys, jossa huomioitaisiin evaluointikapasiteetin kehittämiseen liittyvät kysymykset. Selvityksen tulisi keskittyä metaevaluoinnissa havaittuihin puutteisiin.
6. Parantaakseen hajautettujen arviointiraporttien ja - prosessien hyödyntämistä jatkossa, ministeriön tulisi selvittää ja institutionalisoida johdon roolia ja vastuuta evaluointisyklin eri vaiheiden laadunvarmistuksessa, mukaan lukien seurannassa ja evaluointitulosten hyödyntämisessä.
6. Ministeriön tulisi selvittää metaevaluoinnissa ilmenneiden raporttien ja tehtävänkuvausten ihmisoikeusperustaisuuteen, läpileikkaaviin tavoitteisiin, kestävyys- ja riskien hallintaan liittyvien puutteiden syyt ja ryhtyä toimenpiteisiin niiden korjaamiseksi.

SAMMANFATTNING

Bakgrund och målsättning: Finlands Utrikesministeriets (UM) enheten för utvärdering av utvecklingssamarbetet (EVA-11) beställde den här Metautvärderingen av Decentraliserade Projekt och Programutvärderingar som genomfördes från januari 2012 till augusti 2014. Målsättningen var att utvärdera: standarden på utvecklingsrapporter och utvecklingsdirektiv, Finlands utvecklingssamarbete och UM:s metoder för utvärderingsledning och tjänster för kapacitetsutveckling.

Metautvärderingen genomfördes mellan oktober 2014 och mars 2015 av ett konsultteam från Universal Management Group Ltd. Den baserades på översiktsverktyg som hade utvecklats för översikt av standarden på UM:s 57 decentraliserade utvecklingsrapporter och direktiv; Finlands utvecklingssamarbetes effektivitet, baserat på analys av 34 utvärderingsrapporter av hög standard som har beställts av UM. Utvärderingen av UM:s metoder för utvärderingshantering och tjänster för kapacitetsutveckling baserades på konsultationer med UM-personal och en översikt av relevanta dokument.

Sammanfattning av upptäckter och slutsatser

Rapporternas standard: Generellt bedömdes standarden på de flesta av UM:s decentraliserade utvecklingsrapporter som hög eller ännu bättre. Rapporterna tillhandahöll generellt sett UM:s klienter med bevisbaserade upptäckter och rekommendationer, men dessa hade inte alltid en tydlig målgrupp. Påtagliga brister omfattade ojämn täckning av bevakning av tillvägagångssätt grundade på mänskliga rättigheter, tvärgående mål, hållbarhet och vissa metodiska begränsningar.

Direktivens standard: De flesta direktiv som granskades var välskrivna och klargjorde vad som krävdes av utvärderingen. Påtagliga brister omfattande bristande information rörande metodiska aspekter, ett orimligt stort antal utvärderingsfrågor och frånvaron av utvärderingsfrågor rörande UM:s prioriteringar, inklusive tillvägagångssätt grundade på mänskliga rättigheter och problem med tvärgående mål.

Finlands utvecklingssamarbete: Generellt sett bedömdes standarden på Finlands utvecklingssamarbete mest positivt när det gällde de utvärderade projektens ändamålsenlighet och resultat och minst positivt i relation till deras effektivitet och resultatens hållbarhet. Då den bedömdes ansågs riskhanteringen vara svag. Omdömena var blandade rörande tvärgående mål och tillvägagångssätt grundade på mänskliga rättigheter.

Variabler som påverkar utvärderingars standard: Medan det finns vissa tecken på ett samband mellan välskrivna direktiv och utvärderingars standard, så är det inte klart huruvida detta är ett tillfälligt förhållande eller om det finns andra variabler. Andra identifierade variabler som skulle kunna påverka

utvärderingars standard omfattar: rapporterad brist på en stark utvärderingskultur inom UM, brist på effektiva resultatbaserade ledningsmetoder, personalomsättning, behovet av anpassning till nya policier och otillräcklig utvärdering av bland annat budgetar. (Metautvärderingsteamet ansåg att 43% av direktiven hade tilldelats otillräckliga resurser med tanke på utvärderingens omfattning.)

Jämförelse av metautvärderingar: Bristen på institutionaliserade verktyg för metautvärdering plus riktlinjer begränsar UM:s rådande process för metautvärderingars potentiella effektivitet, rationalisering och användbarhet. En möjlighet att jämföra metautvärderingens upptäckter över en tidsperiod skulle öka effektiviteten och minska den tid som tillbringas med att utveckla och testa nya verktyg och riktlinjer. Det skulle också göra det möjligt för UM att göra mer meningsfulla jämförelser av omdömen över en tidsperiod och granska effekterna av dess handlingar för att åtgärda standardbrister (t.ex. ytterligare support för kapacitetsutbyggnad, eller förändringar av metoder för utvärderingsledning).

Standardgaranti: Medan de allra flesta av direktiven och rapporterna ansågs hålla hög eller ännu bättre standard, antydde en närmare granskning att de stora flertalet av dessa vara bra snarare än mycket bra eller enastående. Om UM strävar efter att öka den allmänna standarden på utvärderingsdirektiv och -rapporter, kräver detta att större uppmärksamhet riktas mot standardgaranti under utvärderingens hela hanteringscykel - med början vid den tidpunkt när ett projekt utformas (t.ex. att se till att integrationsmål verkligen har integrerats, att system är på plats för att samla in data rörande resultat, att Finlands utvecklingspartners är inblandade, att direktivens budget och omfattning är ändamålsenlig, och så vidare). Denna uppmärksamhet måste sedan fortgå under utvecklingen av direktiven, avsättandet av ekonomiska och andra resurser för genomförande och övervakning av utvärderingen, utvärderingens implementering och utvärderingens uppföljning.

Kapacitet för utvärderingsledning: Konsultationer med UM-anställda rörande tjänster för kapacitetsutbyggnad och utvärderingsledning uppdagade att de anställdas egna begränsade kunskaper och färdigheter rörande utvärdering hade haft en negativ inverkan på deras förmåga att välja och leda utvärderare.

Medan enheten för utvärdering av utvecklingssamarbetet (EVA-11) har etablerat en ramöverenskommelse för att stödja den när den beställer företagsutvärderingar, finns det ingen liknande mekanism för decentraliserade utvärderingar. Det stora flertalet av de UM-representanter som intervjuades antydde att en mekanism som en ramöverenskommelse skulle öka effektiviteten under processerna för urval och kontrakterande, ge viss garanti rörande utvärderarnas standard och minska arbetsbördan för utvärderingsledarna.

Utvärdering av tjänster för kapacitetsutbyggnad: Eftersom UM under det kommande året planerar en djupgående översikt av dess tjänster för kapacitetsutbyggnad, har metautvärderingen för 2014 omfattat en anspråkslös översikt av dessa tjänster. Översikten identifierade ett antal luckor som skulle kunna utforskas i den här djupgående översikten.


Överblick av utvärderingen och uppföljning: Medan vissa av de påpekade bristerna hos utvärderingsdirektiv och -rapporter kan rättas till genom vidareut-

bildning av UM-chefer, -personal och -konsulter, har UM-chefer också en viktig övervakningsroll när det gäller att se till att utvärderingsdirektiv och -rapporter motsvarar en acceptabel kvalitetsstandard och att utvärderingars upptäckter och rekommendationer utnyttjas väl. Metautvärderingen uppdagade att till dagens datum har få UM-chefer haft en möjlighet att utveckla tillräckliga kunskaper och färdigheter för att kunna utföra denna övervakningsroll på ett effektivt sätt, vilket de påpekade begränsningarna i vissa utvärderingsdirektiv och -rapporter vittnar om. Dessutom antyder intervjuade personer inom UM att denna situation förvärras genom avsaknaden av en stark utvärderingskultur inom Ministeriet.

Uppmärksamhet på UM-policies och direktiv: Standarden på UM:s implementering av decentraliserade projekt liksom UM-beställda utvärderingsrapporter påverkas negativt av begränsat iakttagande av vissa UM-policies och direktiv. Medan vissa av dessa är relativt nya eller har utvecklats över tid (t.ex. introducerades HRBA officiellt i Utvecklingspolicyprogrammet för 2012 är andra välkända och har identifierats som begränsningar i tidigare metautvärderingar (t.ex. genomgående mål relaterade till kön). Bristen på konsekvent bevakning av dessa frågor hade en negativ inverkan på omdömena rörande UM:s utvecklingseffektivitet. Dessa begränsningar överskrider EVA-11 mandat och speglar brister i UM:s decentraliserade projektlednings metoder för att se till att nya UM-policies och praxis observeras.

Rekommendationer

1. För att vägleda framtida stöd och handlingar som syftar till att förbättra standarden på och användningen av UM-utvärderingar, bör EVA-11 överväga att utveckla en förändringsteori som sammanfattar och kopplar samman förväntade resultat och slutsatser från genomförande av utvärderingar med utvärderingsanvändning och bevisbaserad beslutsfattning.
2. För att förbättra den långsiktiga användbarheten av metautvärderingar bör EVA-11 överväga att institutionalisera en regelbunden metautvärderingsprocess och medföljande översiktsverktyg och riktlinjer.
3. Om UM önskar öka decentraliserade utvärderingsdirektivs och -rapporters generella standard, bör det öka den uppmärksamhet som ägnas åt och resurser som avsätts för kvalitetsgaranti under hela utvärderingens ledningscykel.
4. För att öka den potentiella effektiviteten hos processer som används för att rekrytera och kontraktera kvalificerade utvärderingsexperter, bör UM överväga att etablera en/ flera ramöverenskommelse(r) för decentraliserade utvärderingar. För att se till att det har snabb tillgång till ett tillräckligt antal kvalificerade utvärderare, bör UM överväga sätt för att underlätta både deltagande av individuella utvärderare och större kontraktörer.
5. Med tanke på den sorts frågor rörande tjänster för utbyggnad av utvärderingskapacitet som uppkom under den här metautvärderingen, bör EVA-11 genomföra den planerade översikten av UM:s stöd för utbyggnad av utvärderingskapacitet så snart som möjligt. Översikten bör fokusera på de svaga områden som har identifierats i den här metautvärderingen.

-
- 
6. För att öka decentraliserade utvärderingsrapporters och processers framtida användbarhet, bör UM klargöra och institutionalisera högsta ledningens roller och ansvar för kvalitetsgaranti under utvärderingens hela ledningscykel, inklusive uppföljning och användning av utvärderingsresultat.
 7. För att rätta till påpekade brister i direktiv och utvärderingsrapporter i relation till bevakning av tillvägagångssätt grundade på mänskliga rättigheter, tvärgående mål, hållbarhet och riskhantering, bör UM försöka identifiera orsakerna för dessa brister och åtgärda dem.

SUMMARY

Background and Purpose: The Development Evaluation Unit (EVA-11) of the Ministry for Foreign Affairs (MFA) of Finland commissioned this Meta-Evaluation of Decentralised Project and Programme Evaluations conducted from January 2012 to August 2014. The objectives were to assess: the quality of evaluation reports and evaluation Terms of Reference (TORs), Finland's Development Cooperation, and MFA evaluation management and capacity development services.

The meta-evaluation was carried out between October 2014 and March 2015 by a team of consultants from Universal Management Group Ltd. It was based on review tools that were developed to review: the quality of 57 MFA decentralised evaluation reports and terms of reference (TORs); the effectiveness of Finland's development co-operation, based on an analysis of 34 good quality reports on evaluations that had been commissioned by the MFA. The assessment of MFA evaluation management practices and evaluation capacity development services was based on consultations with MFA staff and a review of relevant documents.

Summary of Findings and Conclusions

Quality of reports: Overall, the quality of most decentralised MFA evaluation reports was rated good or better. The reports generally provided MFA clients with evidence-based findings and useful recommendations, although these did not always have a clear targeted audience. Notable shortcomings included uneven coverage of human rights-based approaches and cross-cutting objectives and certain methodological limitations.

Quality of TORs: Most TORs reviewed were well written and articulated the requirements of the evaluation. Notable shortcomings included lack of information on methodological aspects, an unreasonable number of evaluation questions, and the absence of evaluation questions on MFA priorities including human rights-based approaches and cross-cutting issues.

Finland's development co-operation: Overall, the quality of Finland's development co-operation was rated most positively in terms of the relevance and effectiveness of evaluated projects and least positively in relation to their efficiency and sustainability of results. When addressed, risk management was rated poorly. Ratings were mixed on cross-cutting objectives and human rights-based approach.

Variables that affect quality of evaluations: While there is some evidence of a correlation between well-written TORs and the quality of evaluation reports, it is not clear if this is a causal relationship or if there are other variables. Other variables identified that could affect the quality of evaluations include: reported absence of a strong evaluation culture within MFA, lack of effective results-based management practices, staff turnover, the need to adapt to new policies,

and inadequate evaluation budgets among others. (The meta-evaluation team considered that 43% of TORs had inadequate resources allocated given the scope of the evaluation.)

Comparison of meta-evaluations: The lack of institutionalised meta-evaluation tools and guidelines limits the potential effectiveness, efficiency, and utility of MFA's current meta-evaluation process. Being able to compare the findings of meta-evaluations over time would increase efficiency and reduce the time spent developing and testing new tools and guidelines. It would also allow MFA to make more meaningful comparisons of ratings over time, and examine the effects of its actions to address quality shortcomings (e.g. additional capacity building support, or changes in evaluation management practices).

Quality assurance: While the majority of TORs and reports were considered of good or better quality, a closer review indicates that the bulk of these were considered good rather than very good or excellent. If MFA wishes to increase the overall quality of evaluation TORs and reports, greater attention would need to be paid to quality assurance throughout the evaluation management cycle - starting from the point when the project is being designed (e.g. to ensure that cross-cutting objectives have been integrated, that systems are in place to collect data on results, that Finland's development partners are involved, that budget and scope of TORs are congruent, and so forth). Attention then needs to be continued during the development of the TORs, the allocation of financial and other resources to conduct and monitor the evaluation, evaluation implementation, and evaluation follow up.

Evaluation management capacity: Consultations with MFA officers on capacity building services and evaluation management revealed that officers' own limited knowledge and skills in evaluation have adversely affected their ability to select and manage evaluators.

While EVA-11 has established a framework agreement to support it in commissioning corporate evaluations, no similar mechanism exists for decentralised evaluations. The majority of MFA representatives interviewed indicated that a mechanism such as a framework agreement would increase the efficiency of the selection and contracting processes, provide some assurance regarding the quality of evaluators, and reduce the workload of evaluation managers.

Evaluation capacity building services: As MFA is planning an in-depth review of its evaluation capacity building services in the coming year, the 2014 meta-evaluation included a modest review of these services. The review identified a number of gaps that could be explored in this in-depth review.

Evaluation oversight and follow up: While some of the noted shortcomings in evaluation TORs and reports can be addressed through skills upgrading of MFA Directors, Officers and Consultants, MFA Directors also have an important oversight role in ensuring that evaluation TOR and reports meet acceptable quality standards and that evaluation findings and recommendations are well utilized. The meta-evaluation revealed that to date, few MFA Directors have had the opportunity to develop sufficient knowledge and skills to provide this oversight role effectively, evidenced by the noted limitations with some evaluation TORs and reports. Moreover, interviewed persons within MFA suggest that this

situation is exacerbated by the absence of a strong evaluation culture in the Ministry.

Attention to MFA policies and directives: The quality of MFA decentralised project implementation as well as MFA-commissioned evaluation reports is negatively affected by limited attention to some MFA policies and directives. While some of these are relatively new or have evolved over time (e.g. HRBA was officially introduced in the 2012 Development Policy Programme), others are well-known and have been identified as limitations in previous meta-evaluations (e.g. cross-cutting objectives related to gender). The lack of consistent attention to these matters negatively affected MFA's development effectiveness ratings. These limitations go beyond the mandate of EVA-11, and reflect shortcomings in MFA's decentralised project management practices overall in ensuring that new MFA policies and practices are addressed.

Recommendations

1. To guide future support and actions intended to improve the quality and use of MFA evaluations, EVA-11 should consider developing a theory of change that outlines and links expected results and assumptions of conducting evaluations with evaluation use and evidence-based decision making.
2. In order to enhance the long term utility of meta-evaluations, EVA-11 should consider institutionalising a regular meta-evaluation process and accompanying review tools and guidelines.
3. If MFA wishes to increase the overall quality of evaluation TORs and reports on decentralised evaluations, it should increase the attention paid and resources allocated to quality assurance throughout the evaluation management cycle.
4. To increase the potential efficiency of processes used to recruit and contract qualified evaluation experts, MFA should consider establishing a framework agreement(s) for decentralised evaluations. To ensure that it has ready access to sufficient numbers of qualified evaluators, MFA should consider ways to facilitate participation by individual evaluators as well as larger firm contractors.
5. Given the kinds of questions on evaluation capacity building services that emerged in this meta-evaluation, EVA-11 should undertake the planned review of MFA's evaluation capacity building support as soon as possible. The review should focus on areas of weaknesses identified in this meta-evaluation.
6. To increase the utility of decentralised evaluation reports and processes in the future, MFA should clarify and institutionalise senior management's roles and responsibilities for quality assurance throughout the evaluation management cycle, including follow-up and use of evaluation results.
7. To address noted shortcomings in TORs and evaluation reports related to coverage of human rights-based approaches, cross-cutting objectives, sustainability, and risk management, MFA should try to identify the causes of these shortcomings and address them.

KEY FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

Findings	Conclusions	Recommendations
Quality of TORs and reports and variables that affect quality of evaluations		
<p>The majority of the 57 reports reviewed (77%) were given an overall quality rating of Good or higher. Evaluation reports generally provided MFA clients with evidence-based findings and useful recommendations. Notable shortcomings included uneven coverage of HRBA and cross-cutting objectives and certain methodological limitations.</p> <p>Most TORs reviewed were well written and articulated the requirements of the evaluation. Notable shortcomings included lack of information on methodological aspects, an unreasonable number of evaluation questions, and the absence of evaluation questions on MFA priorities including human rights-based approaches and cross-cutting issues.</p> <p>Variables that could affect the quality of evaluation reports include: reported absence of a strong evaluation culture within MFA, lack of effective results-based management practices, staff turnover, the need to adapt to new policies, and inadequate evaluation budgets among others. (The meta-evaluation team considered that 43% of evaluations conducted had inadequate resources allocated given the scope of the evaluation outlined in the TORs.)</p>	<p>There is some evidence of a correlation between well-written TORs and the quality of evaluation reports; however, it is not clear if this is a causal relationship, or if quality of reports may be explained by other variables as well.</p> <p>In developing the TORs for meta-evaluations over time, MFA has repeatedly noted its interest in understanding the links among various variables such as: the quality of evaluation TOR and reports, the quality of evaluation reports and evaluation budgets, the quality of evaluation reports and the competencies of evaluation teams and so forth. As noted in the 2012 meta-evaluation, it is often difficult to answer such questions as there are many associated variables and assumptions (known, unknown or assumed).</p> <p>This meta-evaluation identified several factors that could affect the quality of evaluation reports. For instance, the absence of a strong evaluation culture and the lack of effective results-based management practices in MFA may contribute to noted limitations in the quality of decentralised evaluation reports. Staff turnover and the need to adapt to new policies were also identified as institutional limitations to evaluation capacity building. This explains why the MFA has invested in capacity building for decentralised evaluations and has provided a broad range of activities to increase the evaluation skills of staff.</p> <p>Another contributing factor might be the inadequate evaluation budgets (often too small given the scope of the evaluations) that could adversely affect the quality of MFA decentralised evaluation reports.</p> <p>Mapping such variables, associated assumptions, and how they relate to one another and ultimately to the quality and use of evaluation results in decision making could be a valuable and exemplary exercise for EVA-11's consideration.</p>	<p>1. To guide future support and actions intended to improve the quality and use of MFA evaluations, EVA-11 should consider developing a theory of change that outlines and links expected results and assumptions of conducting evaluations with evaluation use and evidence-based decision making.</p>

Findings	Conclusions	Recommendations
Comparison of meta-evaluations		
<p>While differences in approach preclude direct comparison, the ratings of Finland’s development co-operation based on OECD/DAC criteria were fairly similar in the 2012 and 2014 meta-evaluations and areas for improvement were much the same in both studies.</p>	<p>The lack of institutionalised meta-evaluation tools and guidelines limits the potential effectiveness, efficiency, and utility of MFA’s current meta-evaluation process. Being able to compare the findings of meta-evaluations over time would increase efficiency and reduce the time spent developing and testing new tools and guidelines. It would also allow MFA to make more meaningful comparisons of ratings over time, and examine the effects of its actions to address quality shortcomings (e.g. additional capacity building support, or changes in evaluation management practices).</p>	<p>2. In order to enhance the long term utility of meta-evaluations, EVA-11 should consider institutionalising a regular meta-evaluation process and accompanying review tools and guidelines.</p>
Quality assurance		
<p>MFA’s overall evaluation management cycle is in keeping with good evaluation practices in other foreign service and development organisations.</p> <p>Checks and balances tend to be concentrated in the evaluation preparation/design phase; there are few in the rest of cycle.</p> <p>The main concerns were lack of follow-up on evaluation recommendations and the inconsistent involvement of Finland’s development partners throughout the evaluation cycle.</p> <p>Inadequate evaluation budgets could be adversely affecting the quality of reports.</p>	<p>While the majority of TORs and reports were considered of good or better quality, a closer review indicates that the bulk of these were considered good rather than very good or excellent. If MFA wishes to increase the overall quality of evaluation TORs and reports, greater attention would need to be paid to quality assurance throughout the evaluation management cycle – starting from the point when the project is being designed (e.g. to ensure that cross-cutting objectives have been integrated, that systems are in place to collect data on results, that Finland’s development partners are involved, that budget and scope of TORs are congruent, and so forth). Attention then needs to be continued during the development of the TORs, the allocation of financial and other resources to conduct and monitor the evaluation, evaluation implementation, and evaluation follow up.</p>	<p>3. If MFA wishes to increase the overall quality of evaluation TORs and reports on decentralised evaluations, it should increase the attention paid and resources allocated to quality assurance throughout the evaluation management cycle.</p>

Findings	Conclusions	Recommendations
Evaluation management capacity		
<p>MFA staff indicate that they invest considerable time in identifying, recruiting and contracting qualified evaluators to carry out decentralised evaluations.</p>	<p>Consultations with MFA officers on capacity building services and evaluation management revealed that officers' own limited knowledge and skills in evaluation have adversely affected their ability to select and manage evaluators.</p> <p>While EVA-11 has established a framework agreement to support it in commissioning corporate evaluations, no similar mechanism exists for decentralised evaluations. The majority of MFA representatives interviewed indicated that a mechanism such as a framework agreement would increase the efficiency of the selection and contracting processes, provide some assurance regarding the quality of evaluators, and reduce the workload of evaluation managers.</p>	<p>4. To increase the potential efficiency of processes used to recruit and contract qualified evaluation experts, MFA should consider establishing a framework agreement(s) for decentralised evaluations. To ensure that it has ready access to sufficient numbers of qualified evaluators, MFA should consider ways to facilitate participation by individual evaluators as well as larger firm contractors.</p>
Evaluation capacity building services		
<p>Evaluation training provided by EVA-11 is considered good quality, is highly valued, and is updated to remain relevant to participants' basic needs.</p> <p>Some staff expressed the need for more advanced skills training in evaluation for officers.</p> <p>Some staff requested that MFA Directors be provided with the skills upgrading they need to carry out their role in overseeing the quality of evaluation deliverables and processes and the use of evaluation results.</p>	<p>As MFA is planning an in-depth review of its evaluation capacity building services in the coming year, the 2014 meta-evaluation included a modest review of these services. The review identified a number of gaps that could be explored in this in-depth review.</p>	<p>5. Given the kinds of questions on evaluation capacity building services that emerged in this meta-evaluation, EVA-11 should undertake the planned review of MFA's evaluation capacity building support as soon as possible. The review should focus on areas of weaknesses identified in this meta-evaluation.</p>

Findings	Conclusions	Recommendations
Evaluation oversight and follow up		
<p>One main concern linked to the evaluation management cycle was the lack of follow-up on recommendations of final decentralised evaluations.</p> <p>There are currently no clearly defined processes for how recommendations should be used. Responsibilities for sharing associated lessons learned are also unclear.</p>	<p>While some of the noted shortcomings in evaluation TORs and reports can be addressed through skills upgrading of MFA Directors, Officers and Consultants, MFA Directors also have an important oversight role in ensuring that evaluation TOR and reports meet acceptable quality standards and that evaluation findings and recommendations are well utilized.</p> <p>The meta-evaluation revealed that to date, few MFA Directors have had the opportunity to develop sufficient knowledge and skills to provide this oversight role effectively, evidenced by the noted limitations with some evaluation TORs and reports. Moreover, interviewed persons within MFA suggest that this situation is exacerbated by the absence of a strong evaluation culture in the Ministry.</p>	<p>6. To increase the utility of decentralised evaluation reports and processes in the future, MFA should clarify and institutionalise senior management's roles and responsibilities for quality assurance throughout the evaluation management cycle, including follow-up and use of evaluation results.</p>
Attention to MFA policies and directives		
<p>Many evaluation reports provided limited information on human rights-based approaches (HRBA), cross-cutting objectives, sustainability, and risk management.</p> <p>Ratings were mixed on cross-cutting objectives and human rights-based approach. In the majority of projects evaluated, it was often uncertain that overall benefits would continue after project completion. Sustainability was uncertain in 19 of 34 projects. Risk management, when addressed, was rated as Poor.</p>	<p>The quality of MFA decentralised project implementation as well as MFA-commissioned evaluation reports is negatively affected by limited attention to some MFA policies and directives. While some of these are relatively new or have evolved over time (e.g. HRBA was officially introduced in the 2012 Development Policy Programme), others are well-known and have been identified as limitations in previous meta-evaluations (e.g. cross-cutting objectives related to gender).</p> <p>The lack of consistent attention to these matters negatively affected MFA's development effectiveness ratings.</p> <p>These limitations go beyond the mandate of EVA-11, and reflect shortcomings in MFA's decentralised project management practices overall in ensuring that new MFA policies and practices are addressed.</p>	<p>7. To address noted shortcomings in TORs and evaluation reports related to coverage of human rights-based approaches, cross-cutting objectives, sustainability, and risk management, MFA should try to identify the causes of these shortcomings and address them.</p>

The meta-evaluation assessed:

- the quality of evaluation reports and TORs

- Finland's Development Co-operation

- MFA evaluation services

1 INTRODUCTION

1.1 Overview

Universalialia is pleased to present this report on the Meta-Evaluation of Decentralised Project and Programme Evaluations conducted from January 2012 to August 2014.

The evaluation was commissioned by the Development Evaluation Unit (EVA-11) of the Ministry for Foreign Affairs (MFA) of Finland and, as noted in the Terms of Reference (Annex 1), there were two purposes of this study. The first was to help the MFA improve the quality of evaluations, evaluation management practices, and evaluation capacity development services. The second was to identify issues and lessons learned that could help the MFA improve development co-operation.

The objectives were to:

- assess the quality of evaluation reports and evaluation TORs,
- assess Finland's Development Co-operation, and
- assess MFA evaluation management and capacity development services.

EVA-11 commissions meta-evaluations approximately every two years. Previous meta-evaluations were conducted in 1996, 2007, 2009 and 2012. These are planned to inform decisions on Finland's policy for development co-operation. The meta-evaluations also inform the capacity development products and services provided by EVA-11, including revisions to the guidance provided to MFA staff and partners through evaluation manuals and training. Thus, the primary users of the meta-evaluation in MFA are: EVA-11, regional and policy departments, and other stakeholders involved in shaping Finnish development co-operation.

The meta-evaluation was carried out between October 2014 and March 2015 by a team of consultants from Universalialia Management Group Ltd. The evaluation was managed by EVA-11 with the support of an evaluation Reference Group made up of representatives from MFA's regional and policy departments.

1.2 Evaluation Context

Evaluation of development co-operation is part of the internal control and monitoring system of the Ministry for Foreign Affairs. According to the MFA Evaluation Manual, the purpose of evaluation is to improve the quality and promote accountability in development co-operation and development policy (Ministry for Foreign Affairs, Finland, 2013). MFA applies the OECD/DAC and European Union principles, criteria, and standards for development evaluation.

The Development Evaluation Unit (EVA-11) of the MFA is responsible for comprehensive, strategic evaluations focused on themes, sector, or country pro-

grammes. EVA-11 also regularly commissions meta-evaluations in order to summarise the quality of decentralised evaluations and assess the findings of evaluations of Finland's development co-operation.

As defined by the OECD/DAC, meta-evaluations aggregate findings from a series of evaluations and may also judge the quality of evaluations (see Box 1). This MFA meta-evaluation covered both aspects of this definition.

Box 1. OECD/DAC Definition of Meta-evaluation

- The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators.

OECD, Glossary of Key Terms in Evaluation and Results Based Management 2002, p. 27

MFA commits to evaluate every project and programme. These project and programme evaluations, or decentralised evaluations, are the responsibility of the MFA departments and units that are in charge of the development co-operation programmes in specific countries, regions or international institutions. Decentralised evaluations include appraisals and mid-term, final and ex-post evaluations that are conducted as an integral part of project or programme and programme cycle management as defined by MFA Manual for Bilateral Programmes (2012).

1.3 Organisation of report

Following this chapter, the report is organised as follows.

- Chapter 2: Approach, methodology and limitations
- Chapter 3: Quality of evaluation reports submitted to MFA (2012-2014)
- Chapter 4: Quality of TORs for evaluations commissioned by MFA (2012-2014)
- Chapter 5: Effectiveness of Finland's development co-operation
- Chapter 6: MFA evaluation management practices and capacity development
- Chapter 7: Overall observations on MFA evaluations and meta-evaluations
- Chapter 8: Conclusions
- Chapter 9: Recommendations.

The annexes include the methodology materials (TORs, meta-evaluation matrix, etc.), details on the evaluations reviewed and their ratings, and a list of references.

Box 2. Terminology

- For simplicity, this report uses the term 'project' to refer to projects, programmes, and any other types of interventions covered by the evaluation sample.
- On occasion, an evaluation report covered more than one project. As these were usually closely related, they were treated as one project in the review tool.

Meta-evaluations aggregate findings from a series of evaluations.

The meta-evaluation assessed the quality of 57 evaluation reports (33 mid-term evaluations, 20 final evaluations, 2 ex-post evaluations, and 2 reviews).

2 APPROACH, METHODOLOGY AND LIMITATIONS

2.1 Approach

The meta-evaluation included: a review of the quality of MFA decentralised evaluation reports and terms of reference (TORs); a subsequent review of the effectiveness of Finland's development co-operation, based on an analysis of good quality reports on evaluations that had been commissioned by the MFA; and a review of MFA evaluation management practices and evaluation capacity development services. Please see Annex 4 for a full description of the methodology.

The meta-evaluation was carried out between October 2014 and March 2015 and was guided by a framework (Annex 5) that included the key questions as presented by the Terms of Reference.

2.2 Methodology

MFA provided 64 reports on decentralised evaluations carried out between January 2012 and August 2014 along with their associated Terms of Reference, and in some cases Instructions to Tenderers (ITT) or their equivalents (e.g., information extracted from the ITTs by EVA-11). These were screened by Universalialia in collaboration with EVA-11 to remove appraisals and evaluations of initiatives funded exclusively through the multilateral channel. The final population of 57 evaluation reports included 33 mid-term evaluations, 20 final evaluations, two ex-post evaluations and two reviews. The list of evaluations reviewed is shown in Annex 6.

The meta-evaluation team first assessed the quality of the 57 evaluation reports and TORs and rated 44 of these as good or higher in terms of quality. Subsequently, the team analysed Finland's development co-operation based on a review of the 34 evaluation reports that had been rated good or better in terms of quality and that had been commissioned by MFA (i.e. 10 of the 44 evaluations had not been commissioned by the MFA). Of the 34 reports, 18 were for mid-term evaluations.

Three review tools (instruments) were developed to assess the quality of TORs and Evaluation Reports and Finland's Development Co-operation (see Annexes 7, 8 and 11). These were based on OECD/DAC and EU standards and principles, Finnish Development Co-operation policy, MFA Manual for Bilateral Programmes (2012), MFA Evaluation Guidelines (2007), and the MFA Evaluation Manual (2013).

Different tools used different rating scales and these are shown in each section of the report. When an element to be rated in an evaluation report or in the TORs was not addressed at all or was not clearly addressed, the element was rated as 'not addressed/not clear'.

2.3 Limitations

The methodology for the meta-evaluation had certain limitations.

- The **subjective nature of the review process** – based on document review and analysis by individual reviewers to judge the quality of the evaluation reports and TORs.
- The **scope** of the meta-evaluation did not include a review of evaluation Inception Reports or consultations with the MFA officers who commissioned evaluations or the evaluators who conducted the studies, which could have helped validate judgements and/ or address information gaps.
- The 2012 and 2014 **meta-evaluations used different methodologies and different rating scales**.
- The meta-evaluation reviewers **relied on information contained in the evaluation reports**.
- **Coverage of development effectiveness criteria in evaluation reports** was sometimes an issue.
- Due to the **small number of reports in the development co-operation sample** (34 reports), it was not possible to analyse data by sub-unit or to develop conclusions on how MFA is doing in specific sectors, countries or regions.
- The sample of reports used for the assessment of Finland's Development Co-operation included **only those that were commissioned by the MFA**, although there were others that were rated high quality.
- Including **mid-term evaluation reports** in the review of development effectiveness may not be appropriate as final evaluations are more likely to demonstrate tangible results.

3 FINDINGS: QUALITY OF EVALUATION REPORTS

3.1 Overview

This chapter provides the key observations and findings derived from the review of 57 MFA evaluation reports in terms of their compliance with OECD-DAC quality standards. The results of the review of the associated TORs are presented in Chapter 5.

The evaluation report review tool has five categories (see Box 3) and one overall rating section in which the reviewers summarised the most salient observations from the assessment.

Box 3. Evaluation Report Review Tool

- 1. Evaluand: Description of the object, context, and intervention logic
- 2. Purpose: Description of the purpose, evaluation objectives, and scope
- 3. Methods: Explanation of the evaluation methodology
- 4. Substance: Presentation of the findings, conclusions, lessons, and recommendations
- 5. Presentation: Clarity, logic, and evaluation report structure

Each category of the tool is further broken down into sub-questions and criteria (for a detailed breakdown, please see report review tool in Annex 7). When an element to be rated was not addressed at all or was not clearly addressed in an evaluation report, the element was rated as ‘not addressed/not clear’.

The following sections provide overall report ratings and observations on the five tool categories. Examples of qualitative comments associated with Very Good ratings are provided for each tool category. For detailed ratings on quality of reports, see Annex 9.

3.2 Overall Quality of Reports

Finding 1: The quality of the majority of reports was considered good or better.

As illustrated in the figure below, the majority of reports reviewed (77%) were given an overall quality rating of Good (i.e. the expected criteria were mostly satisfied, with no major gaps) or higher: 60% were rated Good and 17% were rated Very Good or Excellent. However, a significant 23% were rated Poor. Among the 57 reports reviewed, no reports were considered Unacceptable overall.

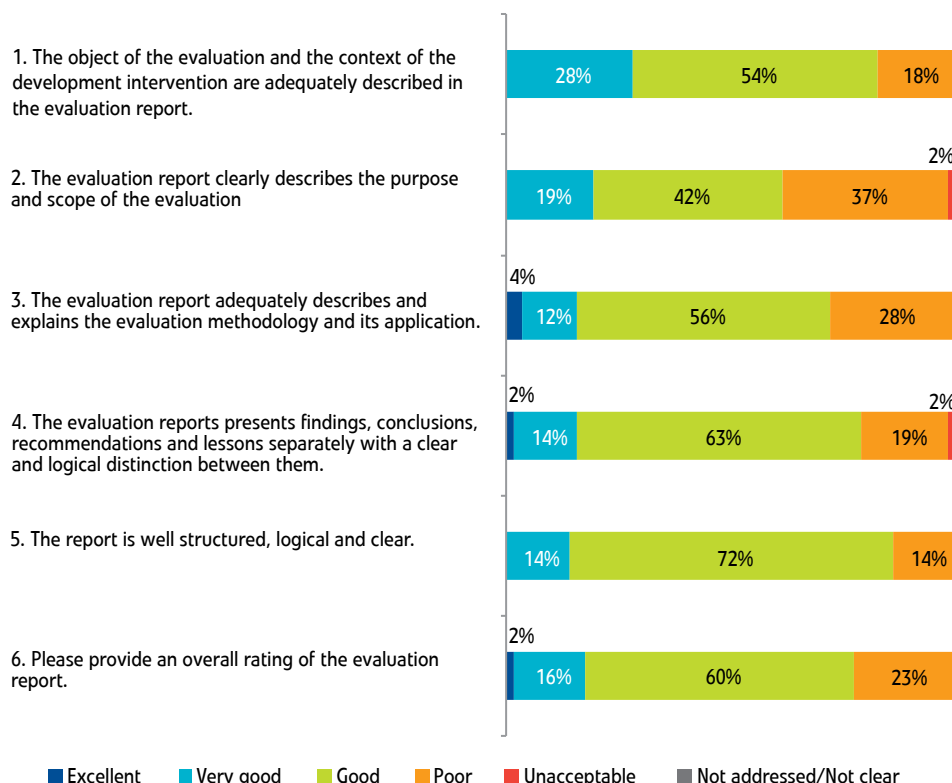
Box 4. Observations on Very Good and Excellent Reports

- Of the 57 evaluation reports reviewed for quality, 10 received overall ratings of Very Good (9) or Excellent (1). Five of the 10 evaluations considered high quality were commissioned by Finland and five were not (and were therefore not included in the review of Finland’s development co-operation). Although this sample is too small

44 of the 57 evaluation reports were rated good for quality, but only 10 were rated very good or excellent.

to draw definitive conclusions on factors that make for the highest quality reports, in the 10 reports that were rated very good or excellent, the methodology section tended to be particularly strong. The description of methodology was thorough and provided a clear rationale for methodological choices as well as good triangulation of data sources. Annexes typically provided further rich detail. These aspects, plus transparency in the report about methodological limitations faced, contributed to the strength and credibility of these sections.

Figure 1 Summary of Ratings – Main Questions on Quality of Reports



Box 5. Examples of Very Good Overall Ratings

- Very high marks for consistency of quality and the readability of the report, a very professional job. Good balance also on addressing an array issues – highly technical and social and environmental and institutional.

Comment on: **Final Review of the Restoration/Rehabilitation of Waterways Project Phase 1 (Philippines)**

- This evaluation contains many areas of strength: the presentation of the assignment, the methodology used, the recommendations and lessons are sufficiently detailed, logical and coherent. Some additional detail could have been given on the project and its context, and this may have helped understand some statements in the findings section. The conclusions of the report are concise and help highlight important areas of strength and weakness in the programme. Recommendations directly link to these and aim to overcome them.

Comment on: **PRORU/RAL Inluyente (2010-2014) Evaluación de Medio Término: Enfoque por Resultado y Sectorial (Nicaragua)**

Most reports described the evaluated object; the context was often less detailed.

3.3 Quality of Reports by Category of Review Tool

3.3.1 Evaluand: Description of the Object, Context, and Intervention Logic

Finding 2: Many evaluation reports accurately described the evaluated object; however, the context that surrounded it was often less evident or detailed within the narrative.

Reports were generally quite successful in providing a coherent overview of the evaluated object (including details on its components, budget, and key stakeholders). However, reviewers identified a number of reports that did not mention Finnish development policy (and explicitly link it to the evaluated intervention) or provide a description of the country/institutional context surrounding the evaluated object.

In total, 82% of reports reviewed were given a Good or Very Good rating for the overall description of the evaluated object and its context.

In describing the **context** surrounding the evaluated object, 65% of reports received a rating of Good or higher. Remaining reports either omitted the context section entirely or did not provide sufficient contextual detail to clearly situate the evaluated object. Further, at least 22 reports (39%) did not mention Finland's development policy or did not discuss it in enough depth for reviewers to see clear linkages with the evaluated object.

In describing the **evaluated object**, 84% of reports were rated Good, Very Good, or Excellent. High quality descriptions of the evaluated intervention were generally characterised by a good discussion of the evaluated object's different components (noted in 72% of reviews), key stakeholders and their roles (well done in 60% of reports), as well as overall cost and budget.

In describing the **theory of change or logframe** of the evaluated object, 67% of reports were rated Good, Very Good, or Excellent. High-quality reports tended to dedicate a section to describing the intervention's logic or expected results. In some reports, annexes helped expand on this discussion by illustrating the theory of change or results framework/indicators. In others, a diagram or graph of the logic underlying the intervention was included.

Box 6. Evaluand - Example of a Very Good Rating

- The opening context of the report is relevant and logical, flowing into the description of the evaluated object, its main partners and components. The most relevant cross-cutting issues are integrated in this section, and an overview of Finnish development policy is provided. The annexes include a good overview of results achievements, which also constitutes a useful source of information for understanding programme objectives.

Comment on: **Mid-Term Review of Establishment of a Regional Flood Information System in the Hindu-Kush Himalaya (RFIS; aka HKH-HYCOS) (Nepal)**

3.3.2 Purpose: Description of the Objectives, Purpose, and Scope of the Evaluation

Finding 3: The purpose and objectives of the evaluations tended to be well explained and clear; the description of scope was often lacking.

Many reports received a Good or Very Good rating for the description of the purpose, objectives, and scope of the evaluation (61% of reports).

In describing the evaluation **purpose and objectives**, 41 reports (72%) were rated Good or Very Good. At least 40% of reports provided a coherent explanation of why the evaluation was being conducted at that point in time.

On the other hand, only 40% of reports received ratings of Good or higher for the description of the evaluation's **scope** and 60% were considered Poor or Unacceptable. While 42% of reports adequately indicated the *temporal* scope of the evaluation, there was often an insufficient description of the evaluation's *geographic* scope (in at least one-third of reports). In other cases, a section on the evaluation's scope was altogether absent.

Box 7. Purpose - Examples of Very Good Ratings

- Good section dedicated to scope and purpose of the evaluation. Contains all the required information. Information presents what was contained in the TORs in a clearer way.

Comment on: **Mid-Term Evaluation of the Wider Europe Initiative**

- Sheds more light on this than the TORs itself, which presented the team with some measurement challenges.

Comment on: **Evaluation of Finnish Rural Sector Development Co-operation in Nicaragua (Nicaragua)**

3.3.3 Methods: Explanation of the Evaluation Methodology

Finding 4: Most reports adequately indicated data sources and integrated the OECD/DAC criteria. Important gaps were noted in terms of the rationale for the selection of methods and samples, as well as the provision of a clear description of methodological limitations.

Most reports included a Good, Very Good or Excellent overall presentation of the methodology (72% of reports reviewed).

The description of **information sources** was rated Good, Very Good or Excellent in 45 reports (79%) and. The sources used were identified in 86% of reports, and 39% explicitly indicated or illustrated the cross-validation and triangulation of data.

The discussion of **methods** was rated Good, Very Good or Excellent in 37 reports (65%), Poor in 18 reports, and Unacceptable in 2 reports. Important gaps related to the description of the rationale for selecting data collection and sampling methods. This was the second most frequently identified weakness in the methodological approach of reports, after the description of limitations.

In most reports, the evaluation purpose and objectives were clear; the description of scope was often lacking.

Most reports indicated data sources and 65% integrated OECD/DAC criteria.

Gaps: rationale for methods and samples, description of methodological limitations

Methodological limitations were insufficiently described or absent from a large proportion of the reports reviewed. Only 18 reports (32%) were rated Good or higher on this criterion; 15 reports were deemed Unacceptable and 24 qualified as Poor. In these instances, limitations were either completely absent from the report's narrative or not described coherently.

Ratings for the integration of the **OECD-DAC criteria** in the evaluation approach (i.e. clearly presented at the start of the report) were mixed: Good, Very Good and Excellent ratings were given to 65% of reports for this component. Successful reports listed the criteria and evaluation questions (in the text or in an annex) and noted the criteria that were more difficult to analyse or less of a key focus in the analysis. Some of the weaknesses identified included: not listing the criteria at the start of the report, combining or not addressing criteria from the TORs (without justification), and using the criteria to structure the conclusions only (i.e. absent from the findings). At least 68% of reports did not clearly list the evaluation questions used, and 28% of reports did not explicitly justify the reason for omitting or adding other evaluation criteria.

In terms of the integration of **cross-cutting objectives** within the evaluation approach, 34 reports (60%) were rated Good or higher. Most reports mentioned gender (60%) and the reduction of inequalities (54%), but only 39% mentioned human rights. Climate change was not considered applicable in six of the reports reviewed, but only about 50% of the remaining reports adequately addressed this issue in the description of the evaluation approach.

The review team noted that **risk management** was not explicitly addressed in the evaluation approach of 70% of the reports reviewed - in many cases this may have been because the evaluation was commissioned before risk management was emphasised as an MFA requirement in its Manual for Bilateral Programming (2012).¹ Consequently, the meta-evaluation team did not consider the rating on risk management in the overall rating for this section of the tool.

Box 8. Methods - Examples of Very Good Ratings

- The evaluation methodology (sources, methods and samples) is well described, both within the report and in annex. The evaluation framework helps illustrate how sources and methods were used to answer the evaluation questions. Further, the inclusion of a dedicated section on limitations adds considerable transparency and credibility to the evaluation process and report. All of the OECD/DAC criteria are included, and cross-cutting issues are well integrated into the approach adopted.

Comment on: **Mid-Term Review of Establishment of a Regional Flood Information System in the Hindu-Kush Himalaya (RFIS; aka HKH-HYCOS) (Nepal)**

- Very transparent in its use of methodology through its description of process, plans, and triangulation methods with explanatory text in each section. Good backup info in annexes.

Comment on: **Mid-Term Review of Tanzania Information Society and ICT Sector Development Project (Tanzania)**

¹ While risk is mentioned in reviewed MFA documents on program design (1999) and the 2007 Development Policy Programme, a more formal description of risk management was introduced in the 2012 Manual for Bilateral Programming.

3.3.4 Substance: Presentation of the Findings, Conclusions, Lessons and Recommendations

Finding 5: The majority of evaluations reviewed contained good quality findings that were based on evidence. Clearer conclusions and lessons could have been drawn in many instances, and recommendations should have consistently identified target stakeholders for action.

In presenting findings, conclusions, lessons, and recommendations, 79% of the reports received ratings of Good or Very Good.

The ratings for subcategories in this section tended to reflect these overall ratings as well. Presentation of the **findings** was rated Good or a Very Good in 79% of reports and 75% did a good job of backing up findings with evidence. In 60% of reports, the findings addressed the evaluation criteria well, flowed logically, and presented key messages clearly.

The presentation of **Conclusions** tended to be on the lower end of the scale: 46% of reports were rated Good and 18% Very Good, but 30% were rated Poor and 5% were considered Unacceptable. At least half of the reports established clear linkages between the findings/evidence and the conclusions. However, in about 40% of reports, conclusions simply summarised the findings or did not add significant value and depth to the analysis.

Like the findings, presentation of **recommendations** was rated mostly Good or Very Good (77% of reports). These reports included recommendations that flowed logically from the conclusions and were based on the findings. An important percentage of reports also provided recommendations that were clearly stated, as well as realistic and appropriate (noted in 40% and 44% of reviews, respectively). Nevertheless, in at least half of the sample, the recommendations did not clearly identify the target stakeholder groups for action. Further, in approximately one-third of reports, recommendations were insufficiently prioritised to facilitate implementation.

The presentation of **Lessons learned** was generally on the lower end of the rating scale. Only 24 reports (42%) received ratings of Good, Very Good or Excellent. Lessons learned were completely absent in 40% of reports. In 32% of reports, the wider relevance of lessons learned (i.e. applicability in other contexts) was not clearly stated.

Notwithstanding the areas for improvement noted above, the great majority of reports were said to **hold together in a clear, coherent way**: 50 reports were rated Good or higher in this category.

Despite the mixed results obtained for the presentation of **OECD-DAC criteria** at the start of the report, the body of most reports generally integrated these criteria well: 82% of reports were rated Good, Very Good or Excellent on this component. In reports where gaps were noted, the missing criteria were most often Efficiency and Impact (i.e. MFA overall objectives). In terms of the analysis of these criteria, ratings were also quite high. Most reports were rated Good or Very Good for addressing all OECD-DAC criteria other than Impact, which was rated Poor in 22 of the 57 reports. This is probably to be expected since most reports were mid-term evaluations and therefore had not progressed enough to show impact level results.

The majority of evaluations contained findings based on evidence.

Gaps: clearer conclusions and lessons, recommendations that identify target stakeholders for action

86% of reports were rated good or very good for presentation.

The integration of **cross-cutting objectives** within the body of the analysis was mixed. In all, 36 reports (63%) received ratings of Good, Very Good or Excellent. Among the cross-cutting issues considered, human rights was most often absent (without justification). In terms of analysis quality, human rights also received the lowest ratings (with only 26 reports rated Good or higher). Conversely, the analysis of gender was the strongest (with 38 reports rated Good, Very Good or Excellent). Results for inequalities and climate change were more spread out.

Though not considered in the overall rating for this section, the review found analysis of risks and **risk management** in 31 of the evaluations reviewed. Thus, risk management was examined in some instances even though it had not been specified in the evaluation approach.

Box 9. Substance - Examples of Very Good Ratings

- The findings in this report rest on evidence and appear objective. They integrate cross-cutting issues and provide project-specific messages on the five evaluation criteria. Overarching messages on the criteria are reportedly included in the annexes (not available for review). Though the findings are at times quite dense, the conclusions and recommendations provided highlight key areas of weakness and suggestions to overcome them. Recommendations are targeted and specific. Lessons are well identified and formulated in a sufficiently general way for wider application.

Comment on: **PRORURAL Incluyente (2010-2014) Evaluación de Medio Término: Enfoque por Resultado y Sectorial (Nicaragua)**

- Findings, conclusions, and recommendations are clearly communicated and well organized into sections which follow logically and are substantiated. Criteria to be assessed are generally well covered.

Comment on: **External Review of Core Support under Joint Financial Agreement to Zambia National Farmers Union (Zambia)**

3.3.5 Presentation: Clarity, Logic, and Evaluation Report Structure

Finding 6: The overall structure of most reports was considered good or better; executive summaries tended to cover key messages, and the annexes provided were generally considered useful.

Most reports were situated in the middle of the rating scale as 72% of reports received a Good rating. An equal number of reports received ratings of Very Good (14%) or Poor (14%).

Ratings of specific sub-questions in this section were all focused around the mid-range of the scale as well, with small differences of note. With regard to the **executive summary**, 68% of reports received ratings of Good or a Very Good and 60% of executive summaries adequately covered report findings, conclusions, and recommendations. However, more than half made no mention of lessons learned.

The overall **evaluation structure** was rated Good or Very Good in 88% of reports, and the quality of **annexes** was rated Good or Very Good in 84%. However, annexes were sometimes incomplete, due to the absence of one or more important elements (including lists of data sources, such as informants or documents reviewed; and data collection tools, such as surveys or interview protocols).

Graphs and tables that added value to the analysis were used in 53% of reports, and 63% clearly defined specialised concepts or included a list of acronyms.

Box 10. Presentation - Examples of Very Good Ratings

- Given the questions and issues to be covered per the TORs, the team has done a good job at organizing the material in a logical and accessible manner, with good balance between narrative and annexes.

Comment on: **Mid-Term Evaluation of Project GCP/KOS/005/FIN Support to Implementation of Forest Policy and Strategy in Kosovo (Kosovo)**

- The report is structured well overall, and has taken a number of strategies to try and serve different audiences who may be interested in more or less detail. Annexes are extensive and contain all expected elements and others as well, though they are extremely lengthy.

Comment on: **Mid-Term Evaluation of the Neighbourhood Investment Facility under the European Neighbourhood and Partnership Instrument 2007-2013**

3.4 Summary

The majority (77%) of reviewed **evaluation reports** were judged to be of acceptable quality overall, providing MFA clients with evidence-based findings and useful recommendations. Notable **weaknesses** included: inadequate descriptions of context, scope and methodological limitations, the frequent absence of lessons learned; recommendations that did not always have a clear target; and weak conclusions.

Despite their overall quality, many reports did not address all aspects of development effectiveness. While coverage of the standard set of OECD/DAC criteria was strong, risk management was mentioned in only half of reports. The integration of human rights-based approaches in evaluations was absent in a majority of cases, and cross-cutting issues of gender equality, reduction of inequality, and climate sustainability were addressed in about half of the reports.

77% of reports were judged to be of acceptable quality overall.

4 FINDINGS: QUALITY OF TERMS OF REFERENCE

4.1 Overview

Each of the 57 evaluations reviewed had Terms of Reference (TORs), and Instructions to Tenderers (ITT) were also available for 35 of the evaluations. The ITTs were reviewed for any additional data not provided in the TORs. Ratings were based on the best information provided, whether in the TORs or ITT. (Typically, the only items for which the ITTs contained more information were the budget or resources required.)

The TORs review template included nine categories (see Box 11) and one overall rating section in which the reviewers summarised the key points of the assessment.

Box 11. TORs Review Tool Categories

- 1. Context: Background information including programme context, programme description, results of previous evaluations, and budget
- 2. Goals: Description of the rationale, purpose, and evaluation objectives
- 3. Scope: Description of evaluation scope
- 4. Questions: Evaluation questions
- 5. Methods: Proposed methodology
- 6. Process: Evaluation process and management structure
- 7. Deliverables: Description of deliverables
- 8. Resources: Description of resources required
- 9. Presentation: Structure and clarity of writing

The following sections provide the ratings and observations on each of the nine tool categories. When an element was not addressed at all or was not clearly addressed in the TORs, it was rated as ‘not addressed/not clear’. For a detailed breakdown of the sub-questions and criteria in each category, please see TORs review tool in Annex 8.

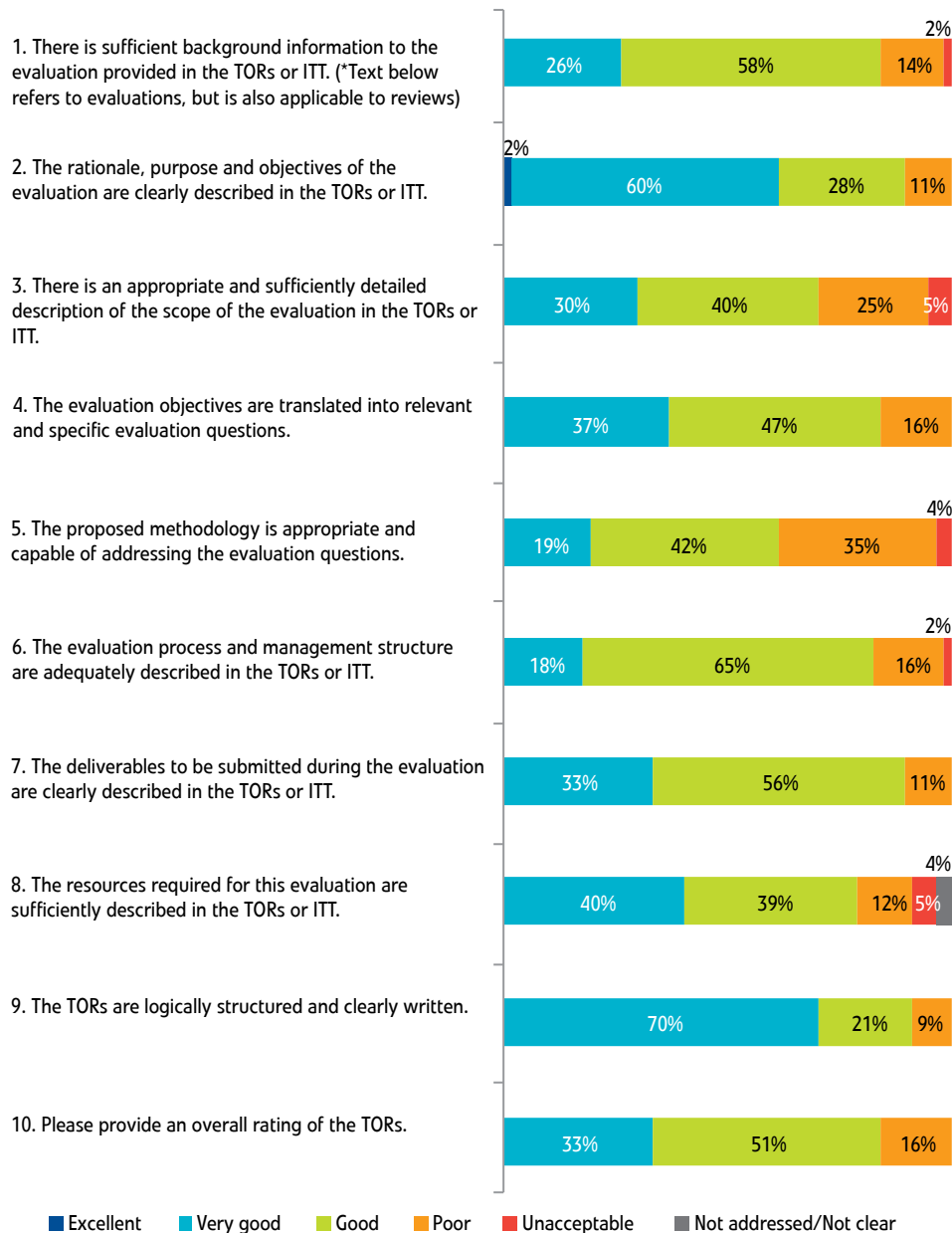
4.2 Overall Quality of TORs

Finding 7: While the TORs reviewed presented a mix of strengths and weaknesses, overall most TORs received positive ratings.

As illustrated in the figure on next page, the majority of the TORs reviewed (84%) were given an overall rating of either Good (51%) (i.e. the expected criteria were mostly satisfied, with no major gaps) or Very Good (33%), while only 16% were rated Poor (i.e. the expected criteria were barely satisfied). None were rated as Excellent or Unacceptable overall.

48 of the 57 TORs were considered good quality.

Figure 2 Summary of Ratings – Main Questions on Quality of TORs



Box 12. Example of a “Good” Overall Rating

- These TORs provide a good overview of the evaluated object, many relevant evaluation questions and consider cross-cutting issues. However, the purpose and scope of the report must be more clearly defined in order to provide sufficient guidance to the evaluators. Further, the number of evaluation questions should be considerably reduced, and the guidance provided on methodology could be more detailed (at least outlining the types of methods and samples to be used).

Comment on: **Terms of Reference – Evaluation of the Finnpartnership Programme, 2012**

Most TORs provided good descriptions of the programme. The context sections received lower ratings.

The ratings of many of the individual criteria for the quality of TORs were also positive. On 14 of the 22 criteria, 80% of the TORs were rated Good or better. However, many TORs were rated significantly lower on other criteria, including: descriptions of evaluation governance and management structures, description of programme context, and methodological guidance.

Overall, about 80% of the TORs reviewed were found to provide enough focus and direction for a good evaluation response, and **enough information to complete a proposal**. See summary of TORs quality ratings in Annex 10.

4.3 Quality of TORs, by Categories of the Review Tool

4.3.1 Background Information and Context

Finding 8: The TORs reviewed generally provided good descriptions of the programme to be evaluated. The description of programme context was not always comprehensive.

Overall, 84% of the TORs reviewed included Good (58%) or Very Good (26%) background information.

The description of the **programme context** was mixed in quality: 32% of the TORs provided descriptions that were considered Very Good, and almost the same percentage were considered Poor. Most TORs (about 65%) described stakeholders and their roles well. The context sections in more than half of the TORs were missing descriptions of the link with Finland's development policy, and background information on cross-cutting objectives almost as often. The context section was one of the four lowest rated criteria (only 58% of TORs were rated Good or better).

Box 13. Example of a "Good" Rating for Overall Background Information

- The first few pages of the TORs cover the evaluated object in some depth, but the context surrounding it (e.g. specific issues identified in the region with regard to waste management) is not highlighted.

Comment on: **Terms of Reference for Mid-Term Evaluation (MTE) of Basel Convention Regional Centre Support Programme, 2013**

The **description of the programme** was generally stronger, with 89% of TORs rated Good (54%) or Very Good (35%). Programme objectives were well described in almost 90% of TORs, but intervention logic was missing in more than one-third of the TORs.

The **results of previous evaluations** were provided in only 14 of the 57 TORs reviewed. However, only 21 of the 57 were TORs for final or ex-post evaluations, so it is possible that there were no previous evaluations conducted in many cases.

4.3.2 Goals: Rationale, purpose and objectives

Finding 9: The majority of TORs reviewed specified the objectives, rationale, and purpose of the evaluation.

Overall, in 88% of the TORs reviewed, the rationale, purpose and objectives description was considered Good (28%) or Very Good (60%) and one report was rated Excellent.

Box 14. Example of a “Very Good” Rating Overall for Rationale, Purpose, and Objectives

- The rationale and objectives of the evaluation are clearly outlined in the TORs. Objectives are numerous, but specific and expressed clearly. The use of the information is also well identified.

Comment on: **Terms of Reference- Mid-term Review. Support to Minority Communities during and after the Decentralisation process in Kosovo**

In terms of providing the **rationale and purpose**, most TORs were considered strong: 54% were rated Very Good, 28% Good, 5% Excellent; 12% were considered Poor. Most TORs (81%) specified **why the evaluation was taking place at this point in time**.

The **objectives of the evaluation** were clearly stated in 96% of the TORs, with 68% rated Very Good, 26% Good, and 2% Excellent. This was a strength frequently noted in qualitative comments.

A small number of TORs (5) did not make clear what type of evaluation they were requesting.

4.3.3 Evaluation scope

Finding 10: While most of the TORs provided good descriptions of the evaluation scope, there were some ambiguities related to the range of stakeholders and/or target groups to be consulted, and insufficient calendar time allocated for realisation of the evaluation’s objectives.

In explaining the **scope** of the evaluation, 70% of TORs were considered either Good (40%) or Very Good (30%) and supplied a sufficiently detailed description; 25% were rated Poor and 5% Unacceptable.

In the qualitative analysis, the weaknesses noted most frequently included: an inappropriate timeline for achieving objectives (in almost half of TORs) and no specification of stakeholder groups and sampling targets (about 40%). Almost one-third of TORs did not clearly define what would be included and excluded from the evaluation. MFA’s evaluation manual notes that it is important for the TORs to clearly identify what is excluded from the evaluation, but it rarely mentions what must be included.

Box 15. Example of a “Poor” Rating Overall for Evaluation Scope

- The evaluation’s scope is not clearly identified in the TORs. A dedicated section on the geographic and temporal scope, as well as the sampling targets, would add considerable insight to help guide the evaluators in responding to needs and expectations.

Comment on: **Terms of Reference for Mid-Term Evaluation (MTE) of Basel Convention Regional Centre Support Programme, 2013**

88% of TORs specified the objectives, rationale, and purpose of the evaluation.

70% of the TORs provided good descriptions of the evaluation scope.

Gaps: timeline for achieving objectives, specification of stakeholder groups and sampling targets

84% of TORs outlined evaluation questions and made use of OECD/DAC and MFA criteria.

Gaps: unreasonable number of questions, did not always include cross-cutting objectives

The methodology sections in TORs received particularly low ratings.

4.3.4 Evaluation questions

Finding 11: The majority of TORs reviewed outlined evaluation questions and made use of OECD/DAC and MFA criteria, although more than half had an unreasonable number of questions. Evaluation questions did not always include cross-cutting objectives.

Overall, 84% of the TORs were rated Good (47%) or Very Good (37%) at translating the evaluation objectives into relevant and specific **evaluation questions**. In 75% of the TORs reviewed, the quality of evaluation questions was considered good or better. A frequent weakness (in almost half of TORs) was an unreasonable number of evaluation questions. The current MFA guideline is to include a maximum of 12 questions.

Box 16. Example of a “Good” Rating Overall for Evaluation Questions

- The five standard OECD/DAC criteria are linked to sets of relevant questions, but the additional criteria provided significantly increase the number of questions to be addressed and do not provide sufficient guidelines around how this should be done within the report.

Comment on: **Terms of Reference for Evaluation of Institutional Support to Civil Society Organisations in Nicaragua, 2012**

OECD/DAC and MFA criteria were well applied in the evaluation questions in the TORs, with 93% rated Good (26%) or Very Good (67%). Over 90% of the TORs included evaluation questions on **relevance, effectiveness, efficiency, and sustainability. Impact (i.e. MFA overall objectives)** was the core OECD/DAC criteria that was most likely to be absent (missing from 15% of TORs), but given the large number of mid-term evaluations in the sample, it was premature to assess this in most cases.

Although questions on **cross-cutting themes** were included in most TORs, these varied by theme. Questions on **climate sustainability** were missing from 33% of TORs, questions on **human rights** were missing in 28%, and questions on **reduction of inequalities** were missing in 25%. **Gender equality** was the cross-cutting theme most likely to appear in the evaluation questions (missing from only 19% of TORs).

Considering all the evaluation questions together, 79% of TORs contained questions that provided a **clear focus and context** for the evaluation, (46% Good, 33% Very Good), while 21% did not (Poor).

4.3.5 Methodology

Finding 12: Methodology was an area that received particularly low ratings as many TORs provided insufficient information.

The methodology sections in the TORs reviewed received mixed ratings: 42% were rated Good and 19% Very Good, while 35% were considered Poor and 4% Unacceptable.

Guidelines for the methodology in terms of **data collection and analysis** were largely Good (39%) or Very Good (25%). However, 33% were Poor, and 2 TORs did not include any guidelines.

In 18 instances, TORs did not include any information on the **sources of data** to be analysed. Other elements were missing from the majority of TORs, including information on how **data analysis** would be conducted (missing from 43 TORs), references to **evaluability** (missing from 54 TORs), and **ethical considerations** (also missing from 51 TORs).

Box 17. Example of a “Poor” Rating Overall for Methodology

- Highly simplified methodology section that does not give the evaluation team sufficient guidance on the breadth of what is to be assessed nor the level of effort it will require.

Comment on: **Terms of Reference: Midterm Evaluation of BioFISA, 2012**

Of the four lowest rated criteria, two pertained to methodology:

Lowest rated criteria	Cumulative % of ratings Excellent/Very Good/Good
Q5 The proposed methodology is appropriate and capable of addressing the evaluation questions.	61%
Q5.1 General guidelines for the methodology are included in the TORs for data collection and analysis.	63%

4.3.6 Evaluation process and management structure

Finding 13: The majority of TORs clearly described the evaluation process, structure and phases of the evaluation, but many did not discuss the roles of MFA and other stakeholders. Information on governance and management structures was a noted gap in a significant number of TORs.

Overall, 83% of TORs reviewed were rated good or higher for the description of the evaluation process and structure - 65% Good and 18% Very Good. Only 16% were rated Poor, and one report was considered Unacceptable (missing any description of process or management structure).

Box 18. Example of a “Good” Rating Overall for Evaluation Process and Management Structure

- The evaluation process and management structure are outlined simultaneously, but greater distinction between the two (i.e. a clear list of responsibilities, outlined separately from the steps of the evaluation) would have ensured utmost clarity in the mandate and support provided.

Comment on: **Terms of Reference for Mid-Term Review of Finnish Support to the Sustainable Management of Land and Environment in Zanzibar, Phase II (SMOLE II), 2012**

In describing the **phases** of the evaluation, 44% of the TORs were rated Very Good, 44% Good, 11% Poor, and one report was rated Unacceptable (lacked any information on phases).

83% of TORs clearly described the evaluation process, structure and phases.

Gaps: roles of MFA and other stakeholders, governance and management structures

84% of TORs specified the expected deliverables.

Gap: 80% lacked a quality assurance system.

79% of TORs described the human resources required to conduct the evaluation.

Approximately half of TORs did not describe the **roles of MFA and other parties**.

Information on **governance and management structures** received mixed ratings. Slightly more than half of the reports (56%) were rated Good or Very Good, 33% as Poor, and 11% (6 reports) Unacceptable. This was the element most frequently missing from the TORs and was one of the four lowest rated criteria.

4.3.7 Deliverables

Finding 14: Most TORs reviewed clearly specified the expected deliverables.

In the sample of TORs reviewed, 84% were rated good or higher in clearly specifying **reports and other evaluation deliverables** to be submitted (30% Good, 54% Very Good, and 5% Excellent). Only six reports were rated Poor in this regard. The most commonly noted gap (approximately 80% of TORs) was the lack of a **quality assurance** system (only seven TORs explicitly requested a quality assurance system). In about half the TORs reviewed, deliverables were not only listed, but described.

Box 19. Example of a “Good” Rating Overall for Deliverables

- The TORs outline the deliverables required, based on the review stages presented, but little is said about the production of an inception report and the implementation of a formal quality assurance process (except for reviews of the findings and draft by stakeholders).

Comment on: **Terms of Reference - Mid-term Review of Finnish support to the Institute of African Leadership for Sustainable Development (Uongozi), 2012**

4.3.8 Resources required

Finding 15: TORs generally provided good descriptions of the human resources required to conduct the evaluation.

In the TORs reviewed, 79% provided a sufficient description of **resources required** for the evaluation (40% Very Good, 39% Good). Seven TORs (12%) were rated Poor, and three were missing any description.

In most TORs, the description of the desired **expertise and knowledge** of the team was rated Good or higher (about 80%), as was the description of the mix of **evaluative skills and thematic knowledge** (about 70%). The most frequently noted gap (in about two-thirds of TORs) was that **gender balance** of the consulting team was not mentioned.

4.3.9 Presentation: Structure and clarity

Finding 16: The majority of TORs reviewed were logically structured and well written.

Overall, 91% of the TORs reviewed were **logically structured and clearly written** (70% Very Good, 21% Good) and only five were considered Poor.

Box 20. Example of a “Very Good” Rating Overall on the Logic of TORs

- The TORs are divided into clearly identified subsections that follow a logical structure and order.

Comment on: **Terms of Reference – Final Independent Evaluation of the Project – From the Crisis towards Decent and Safe Jobs in Southern Caucasus and Central Asia**, 2014

In terms of having sections that hold together in a **logically consistent** way, 72% of the TORs were rated Very Good, 19% were rated Good, and 9% Poor.

In 93% of the TORs, the **style** was brief and to the point, logically structured, and easy to understand (65% Very Good, and 28% Good). The writing style of four TORs was rated as Poor.

4.4 Summary

The majority of **evaluation TORs** were judged to be of good quality overall, with 84% receiving ratings of Good or Very Good. Most TORs reviewed were well written, made use of OECD/DAC and MFA evaluation criteria, and articulated the requirements of the evaluation.

Notable **weaknesses** in the TORs mirrored some of the shortcomings of the evaluation reports: incomplete descriptions of context and lack of information on methodological aspects. In many cases, an unreasonable number of evaluation questions were present. Evaluation questions on **human rights-based approaches** were missing in one-third of the TORs reviewed, and about one-quarter of TORs did not include questions on the **cross-cutting issues** of gender equality, reduction of inequality, and climate sustainability.

91% of TORs were logically structured and well written.

The meta-evaluation reviewed 34 evaluation reports that were rated good or higher in the quality review and that were commissioned by Finland.

5 FINDINGS: EFFECTIVENESS OF FINLAND'S DEVELOPMENT CO-OPERATION

5.1 Overview

This chapter provides the key observations and findings in terms of development effectiveness derived from the review of 34 MFA evaluation reports. These were reports that were rated good or higher in the quality review and that were commissioned by Finland (see Annex 6).

A project review tool, based on OECD/DAC criteria and MFA policies and guidelines for development co-operation, provided the basis for the assessment. The project review tool has eight categories (see Box 21).

Box 21. Development Co-operation Review Tool

- 1. Effectiveness
- 2. Impact (MFA overall objectives)
- 3. Relevance
- 4. Efficiency
- 5. Sustainability
- 6. Human Rights-based Approach
- 7. Cross-cutting Objectives: Gender Equality, Reduction of Inequality, Climate Sustainability
- 8. Risk Management

Each category of the tool is further broken down into sub-questions and criteria (for a detailed breakdown, please see review tool in Annex 11).

The following sections provide a discussion of the sample, overall ratings, observations on each of the eight categories of the tool, and a comparison with the findings of the last meta-evaluation where applicable. For detailed ratings, see Annex 12.

All judgments of quality referred to in the text are those of the meta-evaluation review team, based on the information available.

5.2 Overall Ratings of Development Co-operation

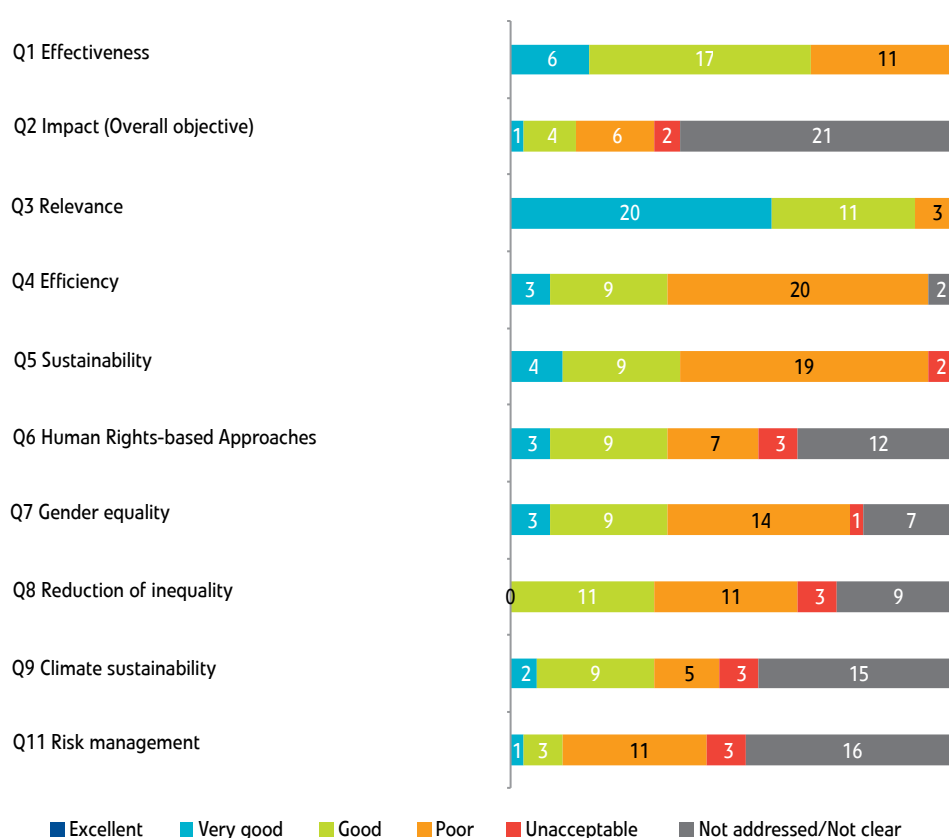
This section provides a summary of the overall ratings for the main criteria in the review tool and highlights the main strengths and weaknesses that emerge when considering the overall ratings for these criteria.

The main criteria were rated on a 5-point scale as shown below. Not all evaluation reports addressed every criterion in the review tool, or did not address them all clearly. In such cases the reviewers rated the criterion as 'not addressed/not clear'. This rating was also used to indicate criteria that were considered

not applicable: for example, it would not be reasonable to expect evidence of impact (i.e. MFA overall objectives) in a mid-term evaluation; impact was rated only for final evaluations. The areas most frequently not addressed in evaluation reports were risk management, human rights-based approaches, and cross cutting objectives.

Finding 17: Overall, the quality of Finland’s development co-operation was rated most positively in terms of evaluated projects’ relevance and effectiveness and least positively in relation to project efficiency and sustainability of results. When addressed, risk management was also rated poorly. Ratings were mixed on other elements (overall objectives, human rights-based approach and cross-cutting objectives).

Figure 3 Overall ratings of main criteria for development co-operation



Note: Q10 was intended for ‘other cross-cutting issues or themes,’ but there was insufficient data to rate any of these.

Of the OECD/DAC criteria, **Relevance** was a clear strength of these projects, receiving by far the highest combined positive overall rating (over 90%). **Effectiveness** was the next highest rated, with almost 70% positive ratings. The lowest rated criteria were **Efficiency** and **Sustainability**, with 60% or more negative ratings.

Ratings on **Impact** (i.e. MFA overall objectives) were mixed, with more negative than positive ratings. However, this was based on a small sample, as only projects with final evaluations or ex-post evaluations (41%) were rated on MFA overall objectives. Since mid-term evaluations generally take place too early in

Evaluated projects were rated most highly for relevance and effectiveness, and least positively for efficiency and sustainability of results.

23 of 34 projects were achieving expected results and making progress towards project purpose, especially projects at or near completion.

a project cycle to reliably assess MFA overall objectives, some reports did not address overall objectives sufficiently to apply a rating.

Ratings were mixed for the **human rights-based approach and cross-cutting objectives**. Roughly half of projects that could be rated were rated positively on the human rights-based approach, gender equality, reduction of inequality, and climate sustainability.

Risk management was not addressed in many reports; when addressed, it was often rated negatively.

5.3 Ratings of Development Co-operation, by Category of the Review Tool

5.3.1 Overview

This section presents ratings and observations on the main criteria and sub-criteria of the review tool, based on the quantitative ratings and the qualitative information about projects drawn from the evaluation reports.

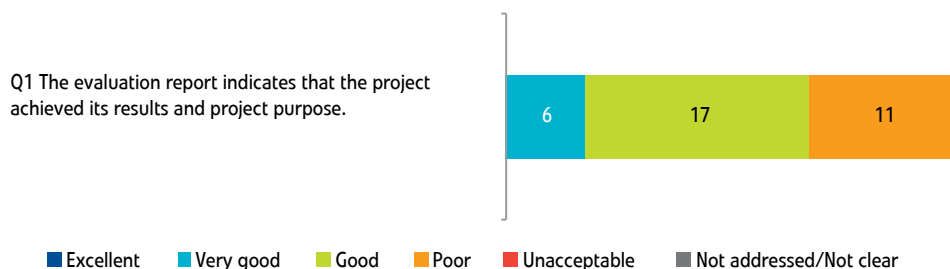
The main criteria were rated on a 5-point scale and sub-criteria were rated on a 3-point scale as shown in the charts below. Not all evaluation reports addressed every criterion and sub-criterion in the review tool, or did not address them all clearly. In such cases the reviewers rated these as ‘not addressed/not clear’.

5.3.2 Effectiveness

Finding 18: The majority of projects evaluated were achieving their expected results and had partially or fully met planned project purpose, especially projects at or near completion.

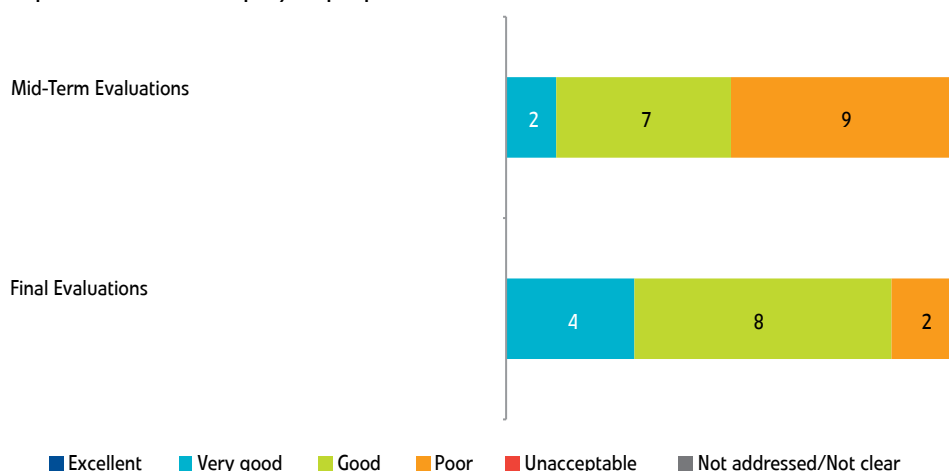
Projects were rated on one main criterion and two sub-criteria for effectiveness.

Figure 4 Effectiveness – Ratings of main criterion on 5-point scale (n=34)



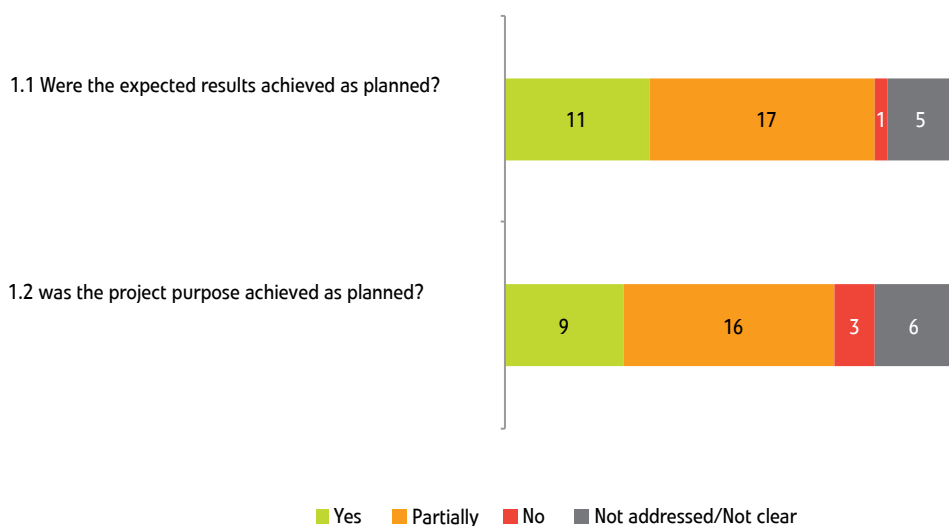
On the main criterion for effectiveness (Q1), 23 projects were rated Good or higher overall. Projects at the stage of a final evaluation were much more likely to be rated Good or Very Good in achievement of results and project purpose overall (see Figure 5 below). Although no projects were considered to have exceeded expectations, all had achieved at least some results. One-third (11) of projects were considered Poor in meeting expected results and project purposes.

Figure 5 Comparison between Mid-term and Final evaluations for Q1: Achievement of expected results and project purposes



Although all reports provided information on the overall effectiveness of projects, almost one-quarter of reports did not clearly identify specific expected results or project purposes (as shown in Figure 6 below), which made it difficult to assess whether these were being achieved as planned overall. Ratings on the specific effectiveness sub-criteria show that 28 projects were fully or partially achieving planned expected results and 25 projects were achieving planned project purposes. Expected results were somewhat more likely to be achieved than project purposes overall. Specific results and project purposes were more likely to be achieved by projects at the final evaluation stage.

Figure 6 Effectiveness – Ratings of sub-criteria on 3-point scale (n=34)



The evaluation reports noted various factors that affected the successful implementation of projects. Perhaps the key factor that emerged from these evaluations is the role of stakeholders. Their level of ownership, communication, and collaboration can affect implementation positively or negatively.

Specific factors that facilitated the successful implementation of projects were strong ownership by relevant stakeholders (9 projects), strong collaboration

Only 5 of 34 projects were rated good or higher in achieving overall objectives (impact).

and coordination amongst stakeholders (9 projects), and strong institutional and human capacities (8 projects). Less commonly mentioned was a favourable external environment (such as strong institutions or the availability of data) (6 projects), strong communication amongst stakeholders (2 projects), and good monitoring and evaluation capacities (2 projects).

Box 22. Examples of reviewers' comments on Effectiveness

- "The project succeeded in designing and implementing numerous interventions in its three priority areas that resulted in policy changes, capacity building at the institutional and individual levels, awareness raising, and behaviour changes."

Comment from: **Independent Final Evaluation Report - From the Crisis towards Decent and Safe Jobs in Southern Caucasus and Central Asia**

- "In the end, the mismatch between resources available (too limited) and project objectives (too large) made it virtually impossible to achieve the desired goals. In other words, the source of the problems here is an unrealistic project design rather than faulty implementation."

Comment from: **Mid-term Evaluation Report - Building Trade Capacity through Customs Modernization in the East and Southern Africa Region**

The most commonly cited factor hindering implementation of projects was an unfavourable external environment (16 projects), such as an unfavourable political environment. Weak institutional and human capacities, including poor management capacities, were also a barrier to implementation (9 projects), as well as lack of communication amongst stakeholders, including the sharing and disseminating of information (9 projects). Similarly, a weak sense of ownership amongst stakeholders was a common issue (9 projects). Lack of resources, whether financial, human, or material, in some cases caused by the departure of partner organisations or the abrupt end of funding for some other reason, posed further problems for projects (8 projects).

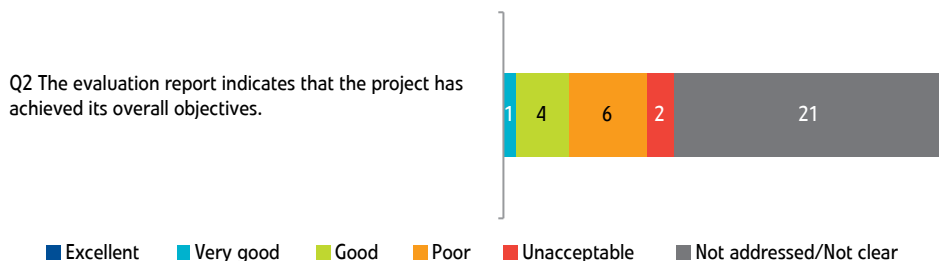
Delays in implementation or a short implementation timeline (4 projects) posed a challenge to achieving results and project purposes. Lack of data was also an issue, for example, lack of baseline data against which to measure progress (2 projects) and insufficient monitoring (1 project). These were sometimes combined with unrealistic or ambitious targets (1 project), or external factors such as legal and political changes (1 project).

5.3.3 Impact (Overall Objectives)

Finding 19: Most of the completed projects that could be assessed fell short of achieving their intended overall objectives; however, many final evaluations were conducted just before the end of projects, which is generally recognised as too early to expect achievement of long-term overall objectives.

Projects were rated on one main criterion and one sub-criterion for overall objectives.

Figure 7 Overall objectives – Ratings of main criterion on 5-point scale (n=34)



Overall, only 13 of the 34 projects addressed the main criterion (Q2), and five of these (15%) were rated Good or higher.

Box 23. Comparison with 2012 Meta-Evaluation

- This seems to be a slight improvement since the 2012 Meta-Evaluation, which found only 12% of projects to be Good or Very Good in achieving overall objectives while 66% were problematic or had serious deficiencies.

In general, the numbers reflect the challenges of achieving overall objectives, and most projects fell short of achieving their intended overall objectives. To date, MFA has commissioned few ex-post evaluations; thus there is limited information available on the extent to which MFA projects are realizing their overall objectives.

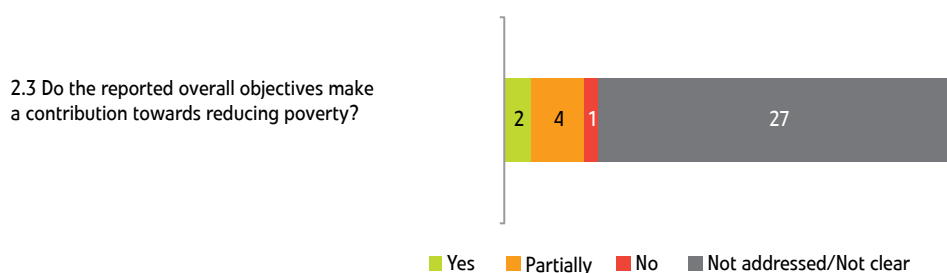
While these figures highlight the continuing challenges of achieving and demonstrating contribution to impact (overall objectives), it is also important to consider that many final evaluations were conducted just before the end of a project, which is too early to observe long-term overall objectives. The few that could identify overall objectives reported socio-economic, legal, environmental and institutional changes.

Box 24. Example of reviewers' comment on Poverty Reduction

- "It is too early to expect concrete poverty reduction impact or increases in harvests for improved seed production. But the potential impact of the project is great."

Comment from: **Final Evaluation of Nicaragua-Finland Agro-biotechnology Programme NIFAPRO**

Figure 8 Overall objectives – Ratings of sub-criterion on 3-point scale (n=34)



Most projects were relevant to beneficiary and country needs, global priorities and donor policies.

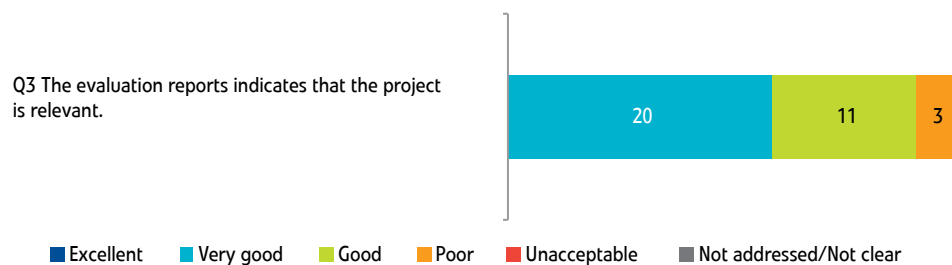
At the sub-criteria level, the **contribution towards reducing poverty** was not specifically addressed in most evaluations, so clear conclusions cannot be made. Of the seven interventions that were rated on this element, six made some contribution. Of the six projects that made contributions, the reports mentioned the following: decreasing poverty through increased agricultural production, reduction of poverty through reduced household expenditures on health, the contribution of flood management to poverty reduction, increased ability of women to start small businesses, improved income and livelihoods through improved forestry management, and contributions of food security, savings systems, and literacy to long term reduction of poverty.

5.3.4 Relevance

Finding 20: The strongest aspect of development effectiveness in projects was relevance, suggesting that the objectives of the projects were largely consistent with beneficiaries' requirements, country needs, global priorities and donors' policies.

Projects were rated on a main criterion and seven sub-criteria for relevance.

Figure 9 Relevance – Ratings of main criterion on 5-point scale (n=34)



Overall, 31 projects (91%) were rated Good or higher on the main criterion (Q3). Twenty projects fully achieved expectations for relevance and only three projects were found to be weak in relevance overall.

Box 25. Comparison with 2012 Meta-Evaluation

- Strong project relevance was also found in the 2012 Meta-Evaluation, in which more than 80% of reports were rated Very Good/Good and only 5% in Problems/Serious deficiency categories.

Figure 10 Relevance – Ratings of sub-criteria on 3-point scale (n=34)



Relevance to the **needs and/or priorities of the target groups** was very strong, with 28 projects demonstrating clear relevance, and another 3 demonstrating partial relevance.

Projects with clear relevance were often noted as responding to critical regional or country risks (such as natural disasters and climate change), addressing gaps (such as human capacity development or health sector funding), or focusing on vulnerable populations (such as the rural poor, adolescents or unemployed youth). In one project that was considered only partially relevant to target groups, the project was developed in response to the expressed priority of the country government, but the evaluators found that most stakeholders had little capacity to absorb the technical assistance provided.

Box 26. Example of reviewers' comment on Relevance

- "The project remains highly relevant as it combines sustainable management forest resources, conservation of environment, and development of livelihoods addressing the rights of women, poorest and marginalized segments of the society. The programme specifically addresses the rights of the poorest of the poor and in overall also climate and environmental sustainability through rehabilitation of degraded forest."

Comment from: **Final Evaluation of Technical Assistance for Leasehold Forestry and Livestock Development Programme**

Only 12 of 32 projects were rated good or higher for efficiency.

Projects were usually in **alignment with national development goals and policies of partner countries** with 26 aligned and 4 partially aligned. In many projects, alignment was reflected in government development targets, national poverty reduction strategies, and various aspects of national legislative and policy environments. Only two projects were not aligned with partner country goals and policies. In both cases, this was not considered a deficiency of the project, but rather a result of the project's support for strengthening civil society in a country in which the government was not committed to sharing power with civil society.

Of the 14 reports that commented on the project's alignment with **global trends and best practices**, 11 projects were in alignment and 3 projects were partially in alignment. In one project, the Ministry of Health in the country was incorporating best practices from various international agreements such as the Rome Declaration and the Monterrey Consensus Statement to develop a Code of Conduct. It should be noted that very few reports addressed global trends and best practices as the MFA did not request that evaluations address this specific criterion.

Of the reports that provided sufficient information on relevance to **Finland's priorities for development co-operation** between 75% and 80% of projects were rated fully or partially relevant to these priorities.

5.3.5 Efficiency

Finding 21: Most projects faced challenges with efficiency, often related to poor design and management, and inefficient systems and procedures for implementation.

Projects were rated on efficiency on one main criterion and three sub-criteria.

Figure 11 Efficiency – Ratings of main criterion on 5-point scale (n=34)

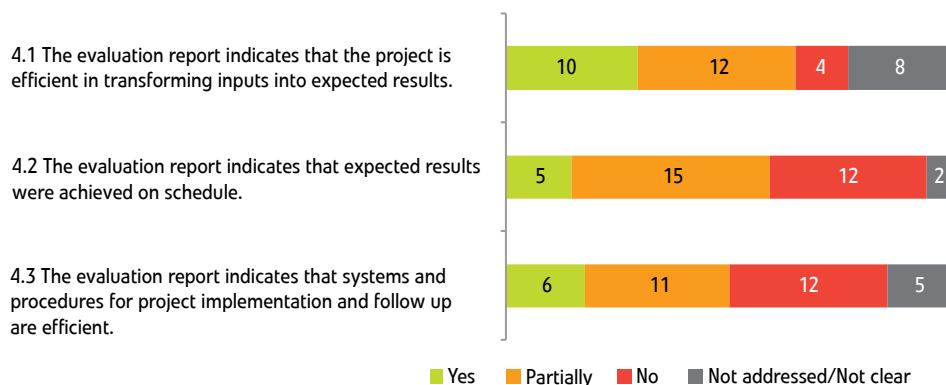


Overall, 12 of the 32 projects that provided information on efficiency were rated Good or higher on the main criterion (Q4). The factor that most commonly affected efficiency was poor design and management (18 projects). Good collaboration and coordination amongst partners and stakeholders was, on the other hand, a strength (13 projects) that contributed to achieving efficiency.

Box 27. Comparison with 2012 Meta-Evaluation

- Efficiency was also an area of concern identified in the 2012 Meta-Evaluation. Almost two out of three (63%) of reports received a score in the Problems/Serious Deficiencies category, and less than one in five reports received a score in the Very Good/Good category. Problems were noted with performance monitoring, administrative processes and human resources and technical assistance, similar to the observations of the 2012-2014 projects.

Figure 12 Efficiency – Ratings of sub-criteria on 3-point scale (n=34)



Of the three efficiency sub-criteria examined, the interventions evaluated were rated highest on **transforming inputs into expected results**. **Systems for project implementation and follow up** were rated more negatively.

Box 28. Example of reviewers' comment on Efficiency

- "Crucial management systems were not designed and put in place, including; effective oversight and decision-making arrangements and an M&E system. Of particular importance is the fact that insufficient attention was paid during inception to developing the implementation modalities of [the project]."

Comment from: **Midterm Evaluation to Support to Rural Development in Zambia Province, Second Phase (PRODEZA II): Final Report**

For projects in which objectives were largely being achieved on time, good management processes were most often given credit, especially the flexibility to adapt and react to changing circumstances. This included being able to respond to emerging issues, take timely corrective action, or re-allocate financial or human resources in response to changes external to the project. Delays in implementation due to inefficient systems and procedures were the largest factor for those projects not achieving objectives as planned. These included slow administrative processes (such as the transfer/disbursement of funds, procurement), slow hiring processes for human resources, and overambitious goals. A lack of a clear division of roles and responsibilities of stakeholders was also noted, as well as the availability of human resources required to support the implementation of projects.

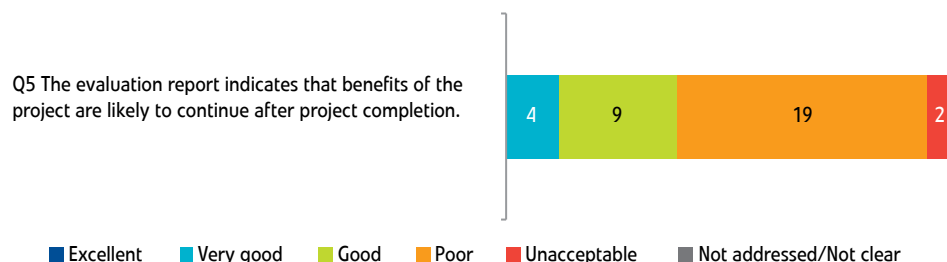
5.3.6 Sustainability

Finding 22: In most projects, it was uncertain that overall benefits would continue after project completion, although final evaluations showed more likelihood of sustainability than mid-term evaluations. The projects that could be rated were expected to be partially sustainable on the dimensions of social, environmental, technical, and institutional sustainability, with financial/economic sustainability the weakest dimension.

Projects were rated on one sustainability main criterion and five sub-criteria.

Only 13 of 34 projects were rated good or higher for sustainability.

Figure 13 Sustainability – Ratings of main criterion on 5-point scale (n=34)



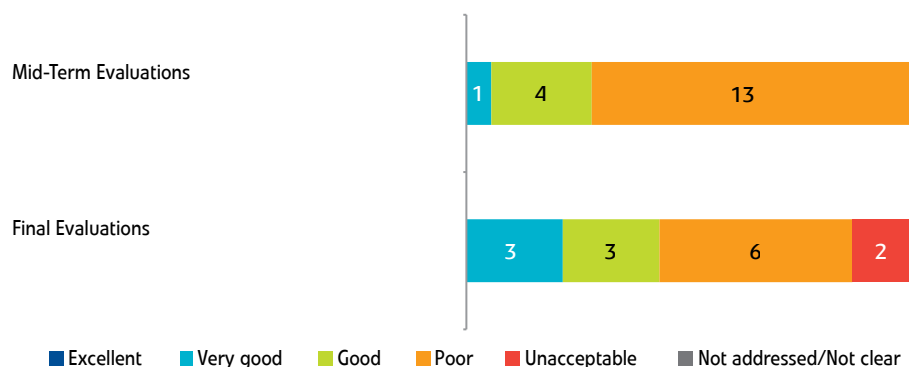
Overall, sustainability was an area of concern in most projects. On the main criterion (Q5), only 13 projects were rated Good or higher.

Box 29. Comparison with 2012 Meta-Evaluation

- Prospects for sustainability were not found to be strong for most projects in the 2012 Meta-Evaluation. Almost two out of three (63%) of reports were categorised as having problems or serious deficiencies in this area, and only 7% scored in the Very Good/Good category. As with the current sample, strong projects had good participation and ownership that helped to consolidate and sustain project results. Weaknesses were likewise found where project activities were slow to integrate into national structures, policies and strategies, and where risks had been poorly anticipated.

In comparing the ratings of overall sustainability for projects at mid-term with those nearing completion, there is an increase in sustainability over time (as shown in the figure below) but this may not be statistically significant.

Figure 14 Comparison of Mid-term vs. Final evaluations for Q5: Overall sustainability



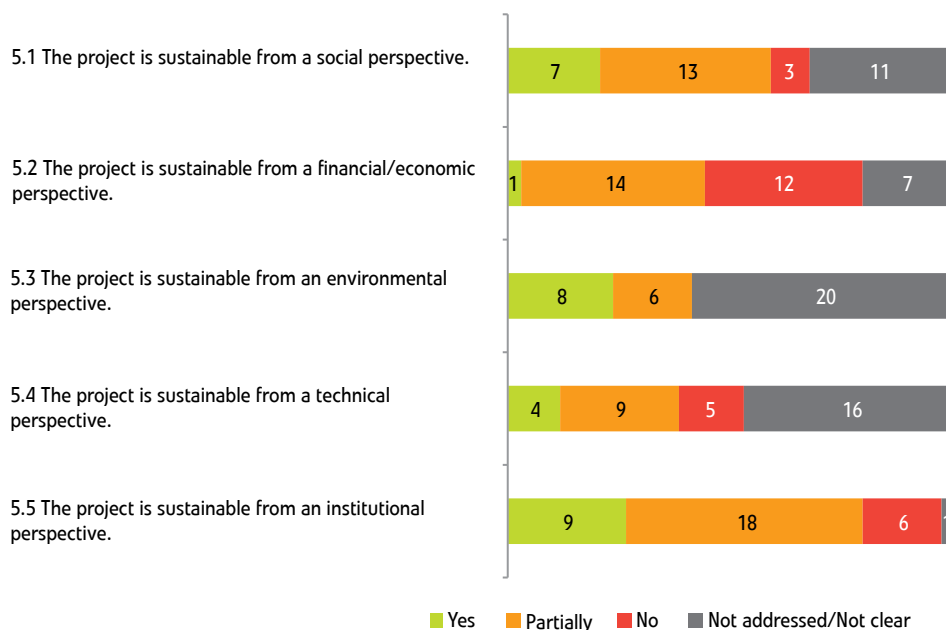
Factors affecting sustainability were very diverse, making generalisations difficult. The most recurrent issues included weak viability due to financial constraints (including reliability of external donor funding) and lack of entities to continue with the project activities after project end (15 projects). Weak organisational capacities to carry out the work (9 projects), lack of sustainability plans and strategies (7) and lack of local ownership of project (6) were recurring internal issues, accompanied by external political and environmental factors (7 projects) beyond the project's control. On the positive side, strong local ownership (9 projects) and building capacities (6 projects) boosted sustainability.

Box 30. Example of reviewers' comment on Sustainability

- "The sustainability of the results or purpose has not been sufficiently embedded in the design of the project and during the decisions taken during the project. The plan for sustainability which was made by the project towards the end was 'too little, too late'."

Comment from: **Final Evaluation of Municipal ICT Programme in Nicaragua**

Figure 15 Sustainability – Ratings of sub-criteria on 3-point scale (n=34)



As shown in Figure 15, the review also examined five specific dimensions of sustainability identified in the 2012 MFA Manual for Bilateral Programmes: **social, financial and economic, environmental, technical, and institutional.**

With the exception of the environment, most projects were considered partially sustainable on these dimensions. As the reviewed projects were designed before 2012 and since these dimensions were not all required elements of evaluations, the extent to which they were addressed in reports varied considerably.

Environmental sustainability (Q5.3) was the least likely sustainability criterion to be addressed in reports. However, of the projects that could be rated, this criterion had the highest proportion of projects that were considered either sustainable (8) or partially sustainable (6). Projects in which the report did not address this topic were usually those with no direct bearing on environmental issues, and therefore this dimension was not necessarily perceived by evaluators as relevant.

Social sustainability (Q5.1) was addressed in 23 projects and seven were considered sustainable and 13 partially sustainable from a social perspective. Social sustainability was often connected to the extent of beneficiary and community participation, such as the extent to which decision making was locally owned.

Institutional sustainability (Q5.5) was addressed in almost all project evaluations, and 27 projects were either sustainable (9) from an institutional perspective, or partially sustainable (18). The main issue for institutional sustainability that

Projects were congruent with MFA policy on HRBA, but only 12 of 22 projects were considered effective in addressing HRBA.

emerged from the evaluations was the extent to which the project was embedded in an institutional setting (such as integrated into government operations) or had co-operation from other institutions and partners with strong capacity.

Technical sustainability (Q5.4) was discussed in only about half of evaluation reports, but of those projects, four were considered sustainable and nine partially sustainable. The main issue that emerged was knowledge management - obtaining and maintaining human resources with technical knowledge and capacity, as well as technical transfer through the capturing and dissemination of knowledge.

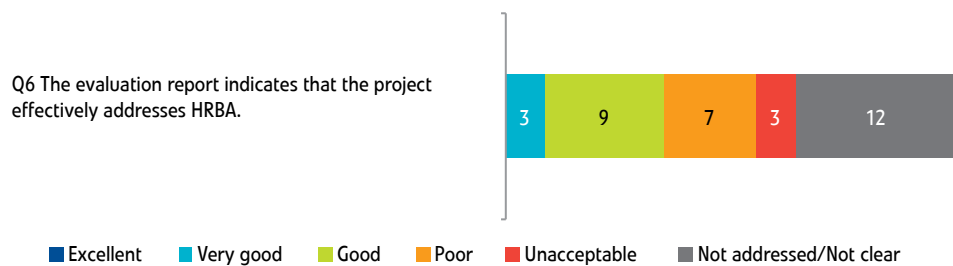
The dimension of **financial and economic sustainability** (Q5.2) was the weakest in the projects reviewed as 12 were considered not financially sustainable, which was the highest negative proportion for any of the dimensions of sustainability. Not surprisingly, a common issue was reliance on donor funding (in some cases on a single source of funding), and limited prospects of obtaining funding from other sources (such as beneficiaries or government, who usually have limited funds available).

5.3.7 Human Rights-Based Approach

Finding 23: Projects were largely congruent with MFA policy on the Human Rights-Based Approach, but had mixed effectiveness in addressing this objective and difficulty in demonstrating evidence of sustainable results.

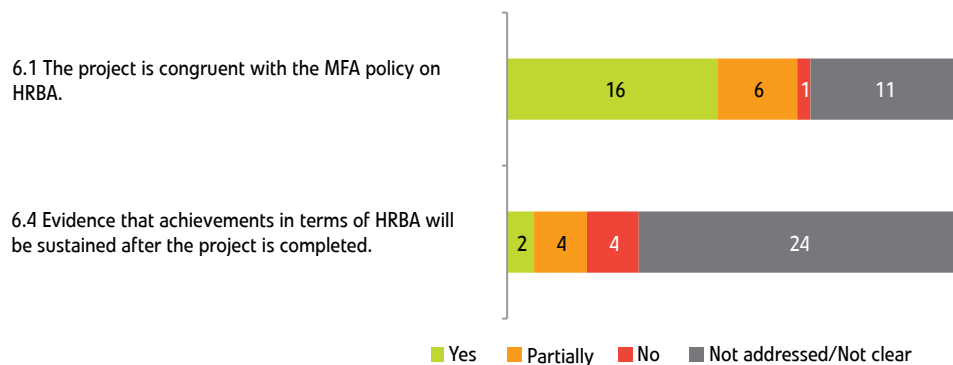
Projects were rated on the human rights-based approach (HRBA) on one main criterion and two sub-criteria.

Figure 16 HRBA – Ratings of main criterion on 5-point scale (n=34)



Of the 22 projects that could be rated for **effectiveness in addressing a Human Rights-Based Approach** (HRBA), 12 were rated Good or higher. In those with negative ratings, there was a need for more mainstreaming of HRBA (8 projects), and a more participatory approach (5 projects).

Figure 17 HRBA – Ratings of sub-criteria on 3-point scale (n=34)



Congruency with MFA policy on HRBA took into consideration whether the intent of the policy was reflected in either project design, the inherent nature of the project, or the way the project was being implemented. Congruency with this policy was high, with 16 projects in congruence, and 6 in partial congruence. A few examples of how a human rights-based approach was implemented in projects included participatory approaches (mentioned in 2 projects), and integrating HRBA into training materials (3 projects).

Only 10 project evaluations addressed **sustainability of HRBA achievements**. Of these, 2 projects were found to be sustainable and 4 partially sustainable. A few successes mentioned by more than one project included increasing beneficiaries' awareness of their rights (5 projects), increasing participation and leadership and strengthening the voice of rights holders (3 projects), and increasing access of rights holders to justice (2 projects).

Box 31. Example of reviewers' comments on HRBA

- "As regards [the project's] objective and strategic plan, impunity has not been reduced in global terms, but the citizens know more and more about their rights, the topic is hot, and human rights issues enjoy increasing visibility thanks to the work of the organisation."

Comment from: **Institutional Support to Civil Society Organisations in Nicaragua: Final Evaluation**

Box 32. Incorporating a Human Rights-Based Approach in development projects

- The Human Rights Based Approach, or HRBA, uses the language of rights, rather than needs, of citizens. It differentiates between rights holders and duty bearers in the process of development, and recognises the social contract that exists between citizens and those who represent them. Considerations for putting this principle into practice include an analysis and identification of rights issues, rights holders and duty bearers; seeking assistance from experts in the sector; incorporating an HRBA framework in the initial design of a project; and including indicators in the logframe to measure progress.

5.3.8 Cross-cutting Objectives

Finding 24: Of those projects for which a rating could be made, congruency with MFA policy on the cross-cutting objectives of gender equality, reduction of inequality, and climate sustainability was high. However, this did not usually result in a high level of effectiveness or sustainable results for those objectives.

The cross-cutting development objectives of **gender equality, reduction of inequality, and climate sustainability** were rated in terms of the project's overall achievement for each objective, congruency with MFA policy, and the extent to which results achieved in that objective were likely to be sustainable. However, only one-third of reports addressed the sustainability of these objectives. In some cases the evaluators did not attempt to comment on results, or their likely sustainability, because no results in a cross-cutting objective had ever been articulated by the project. This was often the case for climate sustainability, as this was added as a cross-cutting objective in the 2012 Development Policy Pro-

While projects were highly congruent with MFA policies on cross-cutting objectives, this was not matched by effectiveness in achieving those objectives.

Contributions to gender equality: only 12 of 27 projects were considered good or better.

gramme, well after many of the interventions included in the evaluation sample were designed.

While the projects for which a rating could be made were highly congruent with MFA policies for cross-cutting objectives, this was not usually matched by strong effectiveness in achieving those objectives.

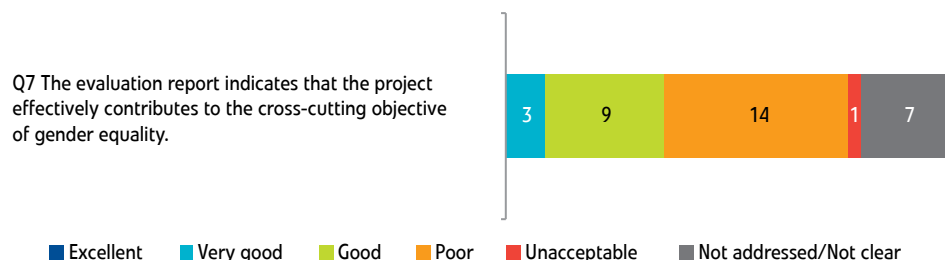
Where the sustainability of results was assessed (which was at most 38% of projects in any of the cross-cutting objectives), it was often found to be weak.

Box 33. Comparison with the 2012 Meta-Evaluation

The 2012 Meta-Evaluation found that cross-cutting objectives were not systematically integrated and addressed by projects. It noted that previous meta-analyses found that only superficial attention was given to cross-cutting objectives – if they were mentioned at all – except for cases where they represented the key priority of the project. It is difficult to conclude whether there has been an improvement since 2012, as a significant proportion of projects in the current sample could not be rated on cross-cutting objectives (between about 20-45%, depending on the objective) because these reports did not address the extent to which such issues were integrated in the projects. However, of those that could be rated, there seems to be moderate improvement in project contributions to MFA cross-cutting objectives.

Gender equality

Figure 18 Gender equality – Ratings of main criterion on 5-point scale (n=34)



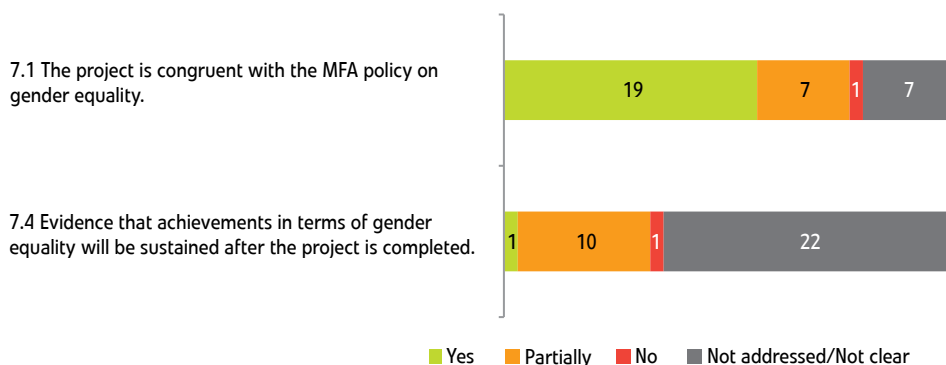
Projects were mixed in their contributions to **the cross-cutting objective of gender equality**, but were most often weak. Just over a third of the projects were found to be Good or Very Good, while about another third were Poor. The most common contributors to low ratings (15 projects) included: lack of focus on gender in general, absence of a gender strategy, or the need for more gender mainstreaming. In other cases there was low female participation in projects (7 projects). Gender disaggregated indicators, or a lack of monitoring of gender mainstreaming was also a reoccurring issue (5 projects).

Box 34. Example of reviewers' comment on Gender Equality

"Due to the planting activities the community (especially women) now saves 2-4 hours daily in fodder collection. This has freed up time for women to participate other activities as cleaning, looking after their children and additional vegetable gardening. Gardening has in turn improved the nutrition of the families in the community. The enrolment numbers of girls in the schools were reported to increase due to the saved time from household chores, mainly fodder collection."

Comment from: **Comment from: Final Evaluation for Technical Assistance for Leasehold Forestry and Livestock Development Programme**

Figure 19 Gender equality – Ratings of sub-criteria on 3-point scale (n=34)



Of those reports that addressed the **congruency of the project with MFA’s policy on gender equality**, 26 were partially or fully congruent.

Only 12 projects could be rated on the **sustainability of results for gender equality**. Of these, 10 were found to be partially sustainable, while one project was thought to have sustainable results and one project could demonstrate no evidence of sustainability. High female participation in projects was often cited as a positive factor (10 projects), along with gender mainstreaming or a gender focus (8 projects) and positive results from providing training to women (8 projects) or providing economic opportunities for women (7 projects).

Reduction of inequality

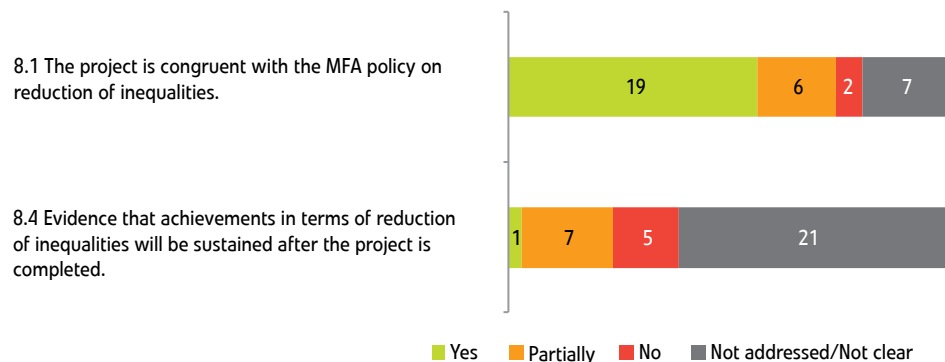
Figure 20 Reduction of inequality – Ratings of main criterion on 5-point scale (n=34)



Reduction of inequality was not effectively addressed by over half of the projects that could be rated - only 11 of the 25 were rated Good (none were rated higher). Weaknesses included a lack of focus on reduction of inequalities, or a lack of explicit linkages with this objective. Another contributing factor was a lack of data analysis, monitoring, or follow-up on vulnerable groups.

Contributions to reduction of inequality: only 11 of 25 projects were rated good.

Figure 21 Reduction of inequality – Ratings of sub-criteria on 3-point scale (n=34)



Of those reports that addressed the **congruency of the project with MFA's policy on reduction of inequality**, a strong 92% were either partially congruent (6) or congruent (19). Typically, these projects either had a specific objective related to reducing inequalities, or indirectly addressed this theme.

Most reports did not indicate the **likelihood of sustainability** for results specific to reduction of inequality. Of the 13 that provided sufficient information to rate this criterion, 7 reports indicated results would be partially sustainable, one report suggested the likelihood of full sustainability, and it was anticipated that 5 projects would not be sustainable.

A few examples of results achieved included economic empowerment for marginalised groups (including job creation, access to markets, and enhanced livelihood opportunities) and improved access to health (including access to water and sanitation).

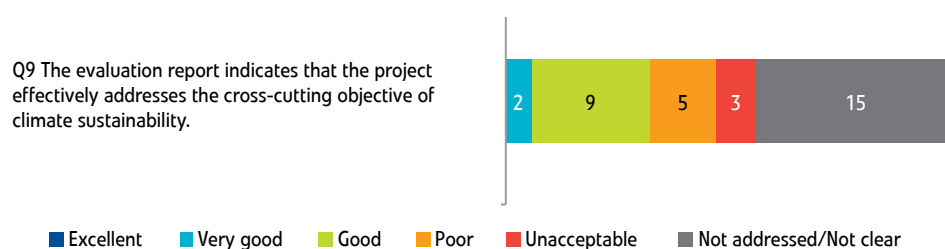
Box 35. Example of reviewers' comment on Reduction of Inequality

■ During the design phase of the project... social inclusion... [was] not considered. The document does not analyse how the data/information can influence the policy makers in improving the socio-cultural, political and economic status of the poor, vulnerable and disadvantaged community; and does not give recommendations on institutional aspects in terms of gender equality and social inclusion. Based on the document (and the consultations) it is also unclear how the project will contribute towards poverty reduction and whether it will benefit the poorest sections of the society.

Comment from: **Mid-term Evaluation of Forest Resource Assessment in Nepal 2009-2014**

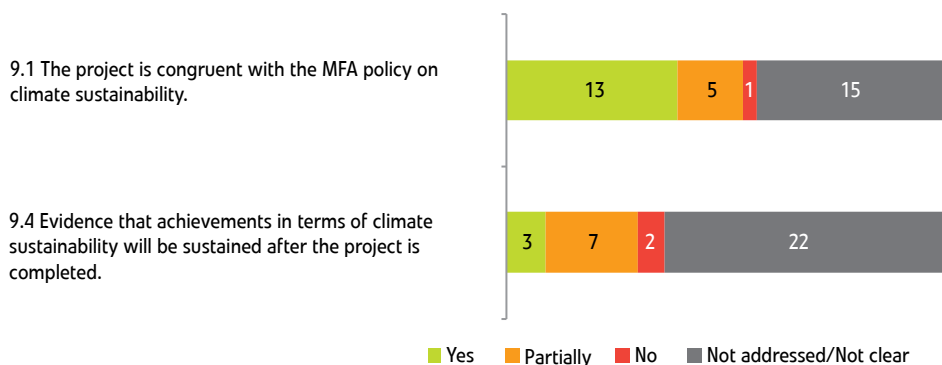
Climate Sustainability

Figure 22 Climate sustainability – Ratings of main criterion on 5-point scale (n=34)



Of the 19 projects that could be rated for this criterion, 11 were found to be **effectively addressing climate sustainability**. Many of the projects in the sample were designed before the 2012 MFA policy for climate sustainability, which resulted in poor coverage of this criterion.

Figure 23 Climate sustainability – Ratings of sub-criteria on 3-point scale (n=34)



Of the approximately half of reports that addressed the **congruency of the project with MFA's policy on climate sustainability**, 18 were partially or fully congruent. For those projects which were congruent, the activities and/or results of the project were benefitting climate sustainability, either with a specific focus on the objective, or some kind of integration of this theme, such as training on environmental issues.

Of the 12 project evaluations that addressed **sustainability of results** in this area, 10 projects were either sustainable (3) or partially sustainable (7) and two were thought to have unsustainable results.

Box 36. Example of reviewers' comment on Climate Sustainability

- “...climate change, which [is an] important cross-cutting objective, [has] not been incorporated into the work plan and road map (despite the fact that addressing climate change is included as one of the key goals of the project).”

Comment from: **Mid-term Review of the Environment Management Support Programme (EMSP) 2010-2014**

5.3.9 Risk Management

Finding 25: In those projects which could be rated, risk management was most often found to be lacking. Risks were often not taken into account during the design phase, and risk assessments were not always complete or accurate.

Risk management was an area of weakness both in terms of the extent to which it was clearly addressed in evaluation reports and the extent to which risks are being well managed by projects.

Contributions to climate sustainability: 11 of 19 projects were effectively addressing climate sustainability.

Many projects were designed before the 2012 MFA policy.

Risk management was weak: 14 of 18 projects were rated poor or unacceptable.

Figure 24 Risk management – Ratings of sub-criteria on 3-point scale (n=34)



Box 37. Comparison with the 2012 Meta-Evaluation

- The 2012 Meta-Evaluation did not comment extensively on risk management, but found that risk analysis in projects was inadequately addressed, as had been the case in the 2009 Meta-Analysis. Weak project designs were an important factor, as is the case in the 2012-2014 projects.

Of the 18 reports that specifically addressed this topic, over half (11 projects) were found to be Poor, and three received the lowest rating of Unacceptable, indicating that they were not meeting any expectations for risk management. Four projects received a positive rating.

Based on those reports that discussed risk management, the overall approach to risk management in projects seemed to be lacking. Some reports noted that risks had not been taken (or only partially been taken) into consideration during the design phase of the project. If a risk assessment was conducted, it was sometimes incomplete or inaccurate, and mitigation strategies were sometimes completely or partially lacking.

Amongst those reports that specifically discussed risk management, reports occasionally mentioned specific risks that were being well managed. These included political risks (in two projects), financial risks, including corruption (four projects), risks related to government capacities (three projects), and environmental risks (two projects). Reports were much more likely to mention risks that were not being well managed. Two types of risks that were frequently noted (in eight or nine reports) as being not properly managed include 1) financial (including lack of funding or inefficient use of resources); and 2) lack of institutional, technological, or human capacities - including high staff turnover. Several projects each mentioned the issues of inadequate country ownership, lack of coordination or overlapping of activities, political risks, and environmental/social risks.

5.3.10 Summary

The majority of projects were making **progress on results and project purposes**, particularly in the case of projects nearing completion. Many projects, however, were not able to demonstrate significant contribution to **overall objectives**. This is to be expected, since in most cases it is not reasonable to expect long term impact immediately on project completion.

As in past meta-evaluations, **relevance** emerged as a strong element. Most projects faced challenges with **efficiency**, stemming either from the initial design or from weaknesses in the implementation process. **Sustainability**, particularly financial, was an ongoing challenge, particularly in projects undergoing a mid-term evaluation. Project congruence with MFA policies on **human rights-based approaches** and the **cross-cutting objectives** of gender equality, reduction of inequality, and climate sustainability was high. However, this did not usually result in a high level of effectiveness in achieving those objectives. In addition, many evaluations did not adequately address cross-cutting objectives and so could not be assessed in this regard; this was most often the case with climate sustainability. **Risk management** was lacking in many projects.

EVA-11 is mandated to provide capacity building services to MFA (training, guidelines, tools and evaluation support services).

EVA-11 support for decentralised evaluations has been enhanced since 2010.

6 FINDINGS: MFA EVALUATION MANAGEMENT PRACTICES AND CAPACITY DEVELOPMENT

6.1 Overview

This chapter presents the meta-evaluation findings on MFA capacity building services and the evaluation management cycle.

EVA-11 is mandated to provide capacity building services to MFA, and these services have typically included training, guidelines, tools and evaluation support services. This chapter provides a cursory high level review of MFA's evaluation capacity building services, and is designed to inform a future in-depth study planned by EVA-11.

The analysis was undertaken on information collected from consultations with nine MFA personnel (see list in Annex 2 using a standard protocol (see Annex 13)) and on a review of relevant documentation (see methodology in Annex 4).

6.2 Capacity Building Services

Finding 26: EVA-11's capacity building support for decentralised evaluations has been enhanced since 2010. Consultations showed that support provided by EVA-11 is generally perceived as responding to the needs of MFA staff engaged in decentralised evaluations.

The 2010 and 2012 meta-evaluations recommended that EVA-11 strengthen its capacity building support to MFA staff for decentralised evaluations.

In response to a recommendation in the 2010 meta-evaluation, EVA-11 conducted a needs assessment to determine MFA staff capacity development needs in relation to commissioning and carrying out evaluations. The assessment identified important needs within MFA as shown in Box 38.

Box 38. Key MFA Evaluation Capacity Needs in 2010

- Clarify the rationale, use and key users of evaluations
- Improve the use of evaluation results
- Better understanding of basic evaluation concepts, criteria and international standards
- More guidelines on how to address evaluation requirements in project design (including the use of logical framework and measurable indicators)
- More information on how to manage evaluation processes
- More guidelines on how to evaluate new types of MFA interventions (including framework programmes, new instruments, budget support and sector wide approaches)
- Clarification of the roles and responsibilities of different stakeholders in the evaluation process (including MFA officials from different units and careers, embassies, advisors, consultants)

Source: EVA-11, Needs Assessment Survey (23.6.2010), AVS-KEO, EVA-11 4.10.2010.

Following the 2010 needs assessment, EVA-11 took several steps to expand its evaluation capacity building support; these included the consolidation of the help desk function, the provision of additional training sessions, the establishment of an annual evaluation event (Evaluation Day), and the development of an updated evaluation manual.

As noted in the table below, MFA staff consulted during the 2014 meta-evaluation generally reported satisfaction with the kinds and quality of capacity building services provided by EVA-11. In particular, informants noted the high quality of evaluation training workshops and experienced trainers, as well as useful manuals and guidelines provided to support staff on evaluation. In addition, all people consulted reported overall satisfaction with the valuable and accessible expertise provided by EVA-11 despite some internal constraints (e.g., few staff).

Table 1 Description and feedback on EVA-11 capacity building services

Services	Description	Overall 2014 feedback
Evaluation help-desk service	<p>The help-desk service on evaluation was established in 2009. The typical types of services offered include: reviews of TORs and reports, and general advice on evaluation issues to bilateral programs or others.</p> <p>The help-desk supported 10 requests in 2011, 19 requests in 2012, 15 requests in 2013, and 14 in 2014.</p>	<p>Consulted informants indicate that the help-desk is a useful source of information, especially when new practices are being used or put in place in MFA. Persons consulted indicate that they appreciate having a reliable source of advice.</p>
Evaluation Day	<p>Evaluation Day is an event meant to raise awareness on evaluation. It is organised once a year by EVA-11 on a specific evaluation topics.</p> <ul style="list-style-type: none"> • 2011: Human Rights and Gender in Evaluation & Evaluation Capacity Development in Partner Countries • 2012: Good Planning – Key to Evaluability and Results • 2013: Budget support –What evidence from Evaluations? • 2014: Multilateral Organisations and Results-based Evaluation <p>Evaluation Days are meant for the development evaluation community (MFA staff, consultants, universities, NGOs) and reach between 50-100 participants each year.</p>	<p>Without being prompted about Evaluation Day, two informants reported that they appreciated the Evaluation Day themes, the utility of sessions offered, and the guest speakers in particular.</p>
Training sessions	<p>Since 2010, EVA-11 has offered 28 training sessions for more than 500 participants. Additional information on training workshops is provided below.</p>	<p>Consulted informants and summaries of participants' feedback suggest that training provided by EVA-11 is generally well perceived.</p> <p>Consulted informants noted that the increased number of external participants in EVA-11 training sessions is perceived as helpful, particularly the involvement of consultants.</p>

Services	Description	Overall 2014 feedback
Manuals and guidelines	EVA-11 provides MFA staff with supporting materials on evaluation, such as guidelines on writing TORs and a checklist on integrating cross-cutting objectives. In 2013, EVA-11 updated and published a new version of the Evaluation Manual.	Feedback from informants confirms that the 2013 Evaluation Manual is good and used a lot. Other guidelines and tools are perceived as useful and relevant. However, it is sometimes unclear which versions of the templates (e.g. TORs, ITT), manuals, and guidelines are the most recent. A simple solution would be to ensure that all these tools and guidelines are dated.

To improve the quality of evaluation reporting, the 2012 Meta-Evaluation recommended that “Capacity building support should be provided to MFA project staff and consultants on how to comply with the DAC/EU Quality Standards.” The MFA management response on the implementation of the 2012 Meta-Evaluation recommendations agreed and strongly encouraged all staff assigned to managing an evaluation to participate in evaluation training. While EVA-11 made increased efforts to offer basic training sessions on a regular basis throughout the year (see finding below on training) participation remained voluntary. Some people consulted for this study argued that this was a limitation as, in their view, the staff most in need of training in evaluation tend not to attend sessions. EVA-11 may want to explore mandatory versus voluntary training offerings.

MFA has also increased resources to support evaluation capacity building:

- Human resources have been allocated to EVA-11 for evaluation capacity building: In 2010, EVA-11 established one full-time position for staff capacity building and consultants have been contracted to assist in planning and conducting training.
- The EVA-11 budget for capacity building services increased by tenfold over the past five years as shown in the table below.

Table 2 EVA-11 Annual Expenditures for Capacity Development (in Euros)

Year	Total Expenditures
2010	€4,701
2011	€15,299
2012	€80,302
2013	€72,466
2014	€50,453

The total expenditures for capacity development include the costs of training (consultant and travel costs when required), the costs of the development of the new evaluation guidelines (consultant expenses), the costs of organising the annual development Evaluation Day in December each year, and travel costs associated with the EVA-11 evaluation capacity building advisor attendance at international evaluation conferences, meetings and seminars. The salary cost of MFA staff is not included.

When asked about the adequacy of resources for EVA-11, the majority of consulted informants said that resources were adequate and well-balanced compared to other operational units within the ministry. While this preliminary review did not identify any extended delays in responding to services or backlogs in training needs, these aspects of service delivery could be assessed by EVA-11 in future studies to validate the level of resources (supply) against demand for services.

Finding 27: Consulted MFA staff report that evaluation training provided by EVA-11 is of good quality, highly valued, and updated to remain relevant to participants' basic needs. Their suggestions to expand training offerings and targets require follow-up study by EVA-11.

The MFA evaluation training programme was designed and piloted in 2010. Training courses were designed to respond to the learning needs of MFA staff as identified in the 2010 needs assessment. Following the needs assessment, EVA-11 training sessions were integrated into the MFA's project management training with a view to strengthening the connection between project planning and evaluation. The objective was to combine theory and practice and enable participants to use the knowledge acquired in their daily work. Finally, to accommodate staff's busy schedules, training sessions were packaged into modules that were offered in three afternoon sessions or in a two-day session.

An overview of the types and frequency of training sessions offered to MFA staff since 2010 is provided in Annex 14 and summarised in Box 39.

Box 39. MFA Evaluation Training since 2010: An Overview

- EVA-11 basic evaluation training package has been delivered since 2010 as 3 afternoon modules or in a 2-day training session
- Basic evaluation training has been offered primarily in Finnish and in English on some occasions. Training materials are in English.
- EVA-11 training sessions have been delivered to external consultants since 2012. And since 2013 training has been offered to mixed audiences (MFA staff from Finland and embassies, consultants, representatives of other countries' ministries, etc.)
- EVA-11 has offered specialised training on various topics (including impact evaluations, evaluation in the UN system and evaluation with a human rights-based approach)
- EVA-11 has occasionally engaged foreign evaluation experts to provide specialised and more advanced courses.

Based on feedback collected by EVA-11 from training participants collected by EVA-11 in 2012 and 2014 as well as informants consulted for this meta-evaluation, a majority of participants had positive views on session objectives, content, trainers, and the atmosphere of the training. They identified several areas where they gained new knowledge, including: evaluation principles, criteria and standards; MFA evaluation policies and culture; the project management cycle; and the logical framework approach. Participants reported they were pleased with the balance of theory and practice in the training as well as the interactive methods used.

In their feedback to EVA-11 on training sessions they attended, participants identified several areas for improvements, including a desire for:

- More concrete and practical examples, more real-life cases and more examples of good practices;
- Additional information on methodology to be included in training sessions, e.g. how to balance qualitative and quantitative methods; and
- Additional training on specific issues, e.g. evaluation cycle, results-based management, impact evaluation, human rights-based approach (HRBA) and cross-cutting objectives.

Evaluation training provided by EVA-11 is of good quality, highly valued, and relevant to participants' basic needs.

Some participants would like more advanced training in evaluation.

Some are concerned that senior managers have limited understanding of the value of evaluation.

In response to this feedback, EVA-11 made changes to training content and design as illustrated in Box 40. The training materials, course structure and invitations to attend training sessions reviewed all show the efforts made by EVA-11 to respond to participants' feedback and changing policies.

Box 40. Changes made by EVA-11 to Evaluation Training Sessions in response to participant feedback

- Cross-cutting objectives and HRBA added to basic training sessions
- Training materials contain more real-life cases and exercises
- Training session scheduling adapted to respond to participant needs. Workshops offered as afternoon sessions (3 modules) or as two-day training sessions.

When asked about other areas for improvement, consulted MFA staff made some comments and suggestions related to the training provided by EVA-11.

- While EVA-11 offers a variety of training sessions in addition to basic training in evaluation, some informants mentioned that they would welcome more advanced training in evaluation. This feedback was also reflected in summaries of previous training sessions in which some participants commented that they would appreciate further education on evaluation to deepen their knowledge.
- Some consulted informants raised concerns about: the relatively low profile of evaluation in MFA, the modest understanding of the role and value of evaluation by senior managers, and senior managers' relatively limited participation in evaluation training sessions (noting that senior MFA staff rarely attend any training).

In order to describe EVA-11 capacity building support for this study, EVA-11 representatives were required to review many files and documents manually, which was time-consuming. At present, EVA-11 does not have a system that consolidates information on the evaluation capacity building support it provides over time (e.g. the number and profile of participants, the courses it offers, or the venues of training sessions). The absence of such information limits EVA-11's ability to review and analyse the support it provides (e.g. to identify trends in participation rates, drop-out rates, and so forth) and also its ability to carry out any longer term studies of the use and effects of its training over time.

EVA-11 may wish to explore these comments and observations on training as part of its planned in-depth review of evaluation capacity building support.

6.3 Findings on the Evaluation Management Cycle

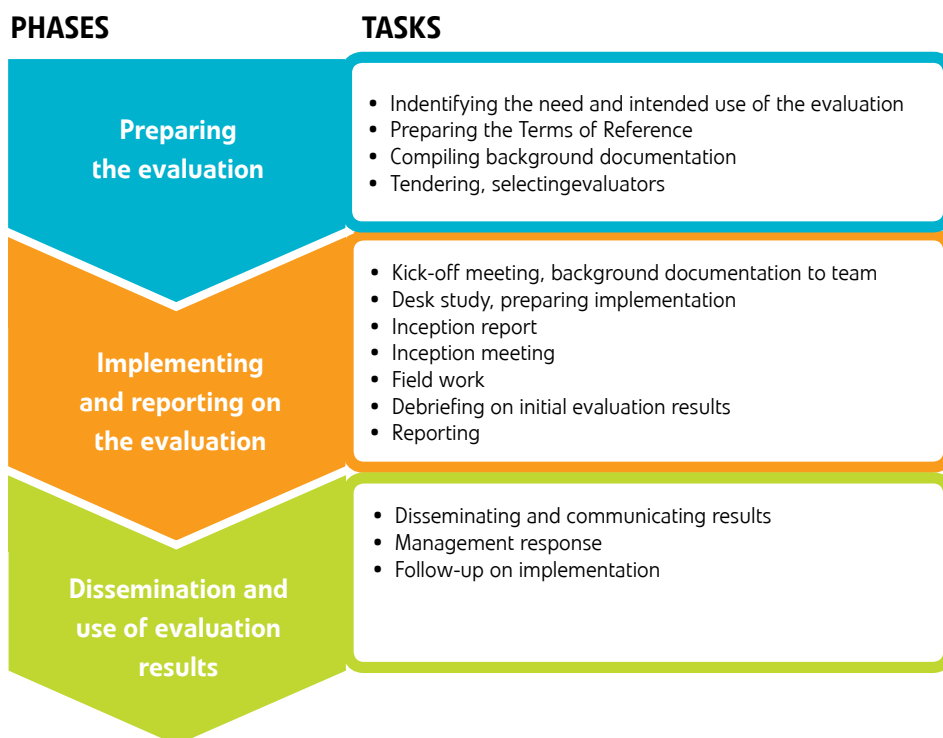
This section describes the meta-evaluation team's main observations on the evaluation management cycle of decentralised evaluations, and its strengths and areas for improvement.

Finding 28: MFA's current evaluation management cycle is in keeping with good evaluation practices in other foreign service and development organisations.

The evaluation management cycle, as depicted in the 2013 Evaluation Manual and shown in the figure below, stemmed from actual capacity building inter-

ventions and was formalised based on consultations with MFA and consultants. The cycle is clearly reflected in the training offered by EVA-11 (i.e. the basic annual training in evaluation of development co-operation).

Figure 25 Evaluation Management Cycle



Consulted informants generally agreed that the evaluation management cycle as depicted in the Evaluation Manual was complete and did not require substantive modifications. The meta-evaluation team agreed that this is a good overview of what is supposed to be done in an evaluation and in keeping with best practices of other comparable agencies known to the team.

Consulted informants provided some suggestions to improve the evaluation management cycle, which MFA may wish to consider in updating the Evaluation Manual in the future.

- In the first phase, the step “Compiling background documentation” could precede “Preparing the terms of reference”.
- In the guidance on Terms of Reference in the Evaluation Manual, it would be useful to expand the description of the scope to specify what is to be included in the TOR as opposed to focusing only on what is to be excluded.
- In the procurement and tendering processes, internal support is often required due to the lack of clear guidance. Procurement is often difficult and time consuming. Framework agreements for decentralised evaluations could be a good option to find reliable and qualified consultants rapidly.
- Some persons reported confusion with the word “Inception” noting a definition or additional guidance is needed.

Follow-up on implementation of evaluation recommendations is most problematic with final evaluations. Some respondents reported that final evaluations are not used systematically.

- In the third phase, dissemination is often quite limited, usually not extending beyond a project, and lessons learned are not always disseminated to a wider audience.
- The cycle as defined above is “front loaded” (i.e. checks and balances tend to be concentrated in the preparation phase) and there are few checks and balances in the rest of cycle to examine the progress of the evaluation over time. This has implications for the quality of evaluation reports; it is often too late to ask for a major overhaul when a draft version of the report has been submitted. One informant noted that evaluation teams are often not briefed on MFA evaluation requirements, which has repercussions on the quality of reports.
- Finally, it was noted that the MFA Evaluation Manual is designed to guide both development evaluations and management reviews – two deliverables that have distinct objectives and that do not serve the same users. Informants noted the ambiguity that currently exists between management reviews and evaluations, and suggested that reviews be taken out of the Evaluation Manual and that distinct guidance be developed for reviews.

Finding 29: MFA informants indicate that responsibility for follow-up and use of final decentralised evaluation results is ambiguous and limits their use.

The use of evaluation results is a long standing challenge, noted as far back as 2010 in the EVA-11 needs assessment. A majority of informants reported that that follow-up on final evaluation recommendations is most problematic in the cycle (third phase) mainly because roles and responsibilities are not clear (i.e. roles and responsibilities of evaluation managers in country, EVA-11, partner institutions, and management in Helsinki). Follow-up on implementation of evaluation recommendations is most problematic with final evaluations. Informants reported that final evaluations are not systematically used, and it remains up to desk officers whether follow-up is required or not. It is unclear who uses the results of final evaluations, either within the MFA or partner countries. Consulted informants reported that management responses are taken into account in the field, but not necessarily in MFA.

According to consulted MFA staff, the situation is not as problematic with mid-term evaluations as it is clear who the owners of the evaluation report are (i.e. the embassy and the Ministry) and there are established mechanisms (such as the Steering Committee) that address recommendations. In theory, the evaluation results should serve both their purposes.

The lack of follow-up has implications on the use of evaluation results in informing lessons learned. Informants recognised that this relates to the higher level issue noted above: setting evaluations as a priority within the top management of the Ministry to ensure use of results.

Box 41. European Commission 2014 report on the uptake of strategic evaluation recommendations

- The first recommendation was that the primary responsibility for promoting a learning culture and upgrading the use of evaluation should lie with the political leadership and senior management. In their view, senior management needs to provide a clear rationale for conducting evaluations and incentives for using their results.

Source: Development and Co-operation EuropeAid (2014). *Study on the uptake of learning from EuropeAid's strategic evaluations into development policy and practice: Final report*, 61 pages.

Finding 30: While the MFA Evaluation Manual requires that partner countries play an active role in the evaluation process, consulted informants indicate that this does not always happen in practice.

According to consulted MFA staff, partners in most countries are actively involved in reviewing the TORs, but joint management of the entire evaluation process (as encouraged by the Evaluation Manual) is less common. Consulted informants reported that it is not clear whether the Evaluation Manual's guidance on engaging country partners is followed or who has the responsibility for engaging the local partners. This seems to be done on a project basis and is not systematic. There are some examples of participatory processes in-country, but they require considerable time and effort.

This is an area that requires further investigation by EVA-11 to fully understand the situation and the consequences of involving or not involving partner countries in the evaluation process.

6.4 Summary

The limited review of evaluation capacity building services suggests that support provided by EVA-11 is responding to the needs of MFA staff, and is comprehensive and useful. Informants identified a need for more advanced level training sessions, and an updated training needs assessment of MFA staff could shed light on this area.

The **evaluation management cycle** described in the 2013 Evaluation Manual is complete and comprehensive, but there is ambiguity around responsibilities for evaluation follow-up, particularly with final evaluations, which limits the use of evaluation results.

Partners in most countries are involved in reviewing evaluation TORs, but less often in other phases of the evaluation.

7 OVERALL OBSERVATIONS ON MFA EVALUATIONS AND META-EVALUATIONS

7.1 Overview

The previous chapters of this report presented findings related to the quality of MFA evaluation deliverables, the effectiveness of Finland's development co-operation, and MFA's evaluation management and capacity building processes.

This chapter provides observations about issues that cut across the findings of the previous chapters with the intent of shedding additional light on the evaluation function and helping inform MFA actions to continue to enhance the quality and utility of decentralised evaluations and meta-evaluations.

The following sections include observations on: cultural and institutional issues within the MFA that may affect the quality of evaluations, the linkages between the quality of TORs and reports, the adequacy of evaluation resources, and EVA-11 meta-evaluation processes.

7.2 Contextual issues

Finding 31: Consulted stakeholders commented that the absence of a strong evaluation culture and the lack of effective results-based management practices in MFA contribute to noted limitations in the quality of decentralised evaluations. Staff turnover and the need to adapt to new policies were also identified as challenges to evaluation capacity building.

Evaluation culture

The meta-evaluation's review of evaluation reports and consultations with MFA stakeholders suggest that there are some contextual matters that adversely affect the quality of decentralised evaluations. One of these is the reported absence of a strong, supportive evaluation culture. Consulted informants feel that evaluation is not perceived as a priority by the Ministry's higher management and that efforts to improve evaluation capacity will be in vain if they are not supported by evidence of the use of evaluation results to inform decision-making in the MFA. This is something that MFA may wish to investigate and validate in a more in-depth study, to help identify ways in which the Ministry could best use the results of evaluations and become a learning organisation.

Results-based Management

Another shortcoming that goes beyond the control of the evaluator relates to the absence of effective results-based management practices within MFA projects and programmes, and the effect on the quality of evaluations. This is a continuing challenge with several implications for evaluations. These include: the absence of clear, appropriate theory of change and/or results framework against which the effectiveness of a project can be reliably measured; and a

MFA evaluation culture: Consulted informants feel that evaluation is not perceived as a priority by the Ministry's higher management.

dearth of information on the planned/actual effectiveness of projects in realising planned overall objectives, purposes, and expected results, which in turn limits the evaluator's ability to make judgments about a project's effectiveness. These factors can contribute to evaluation reports that do not adequately meet MFA evaluation requirements. Ideally, such limitations should be identified and addressed during the evaluation inception phase through, for example, an evaluability assessment which could inform more effective use of evaluation resources.

The 2012 Meta-Evaluation recommended that existing tools and guidelines related to results-based management be reviewed and improved, in particular the Manual for Bilateral Programmes published in 2012, which has not been updated since then. The modest progress since 2012 is also reflected in an internal 2013 MFA memo for directors and advisors which, among other needs, identified the need for results-based management training for directors and the use of results-based management at all levels of the organisation to enhance development co-operation. It suggested the need for MFA to review and identify how to address noted challenges with existing results-based management practices. EVA-11 might want to follow the practice of other donors by introducing evaluability assessments as part of, or prior to, commissioning an evaluation (see Box 42).

Box 42. Evaluability assessment

- Evaluability assessment is the determination of the appropriateness of conducting an evaluation. It is used to determine if the Evaluand is "ready" to be evaluated based on the existence of goals, resources, data accessibility, and how the evaluation is intended to be used.

Source: Wholey, J. S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J.S. Wholey, H. J. Hatry & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15-39). San Francisco: Jossey-Bass

MFA staff turnover and rotation

Another contextual factor is the reported high staff turnover and rotation throughout MFA, a challenge faced by foreign ministries and departments in other countries. This situation calls for MFA staff to have regular and easily accessible skills training to carry out their responsibilities, including management of decentralised evaluations. EVA-11 has partly addressed this issue by providing basic training packages on a regular basis. However, a number of informants identified a bigger challenge; according to them, it can be difficult to write good TORs if one does not have sufficient knowledge of development co-operation policies and dynamics, target countries or the substance of the project. For the same reason, it is difficult to know if the quality criteria of an evaluation report are met. Informants mentioned that this may indicate a need for capacity building in other and broader areas such as development co-operation. This argument is supported by the previously noted "Executive Summary" event that took place in 2013 for directors and advisors, during which it was mentioned that additional investments were required in development policy capacity building: i.e. recruiting senior experts specialised in adult-education and learning to address the emerging training needs and to improve the quality through capacity building of MFA staff in all levels.

RBM: The absence of a theory of change and/or results framework and lack of information on planned/actual project results limits the evaluator's ability to assess project effectiveness.

MFA staff need skills training to carry out their responsibilities for decentralised evaluations. Some suggested this should include training in broader areas such as development co-operation.

Informants consulted for this study reported that more concrete guidelines and tools are needed to implement new cross-cutting objectives.

According to informants, AHA is rarely used during the implementation of a project because it is difficult to make changes.

Evolving MFA policies

As in other foreign ministries, MFA priorities, policies and procedures continue to evolve over time. For example, over the past few years, MFA has increased the emphasis on HRBA and risk management and has further clarified the various dimension of sustainability that should be examined. A memo prepared by EVA-11 in 2013 described the challenges in implementing the 2012 Development Policy Programme. Stakeholders emphasised the absence of clear guidance to put the new policy into practice; the associated tools were often delayed and there were sometimes conflicting guidelines.

Informants consulted for this study reported that more concrete guidelines and tools would be needed to implement new cross-cutting objectives, especially climate sustainability, a human rights-based approach, and risk management in all phases of the project cycle. Since 2012, EVA-11 has supported the development of training modules to address HRBA and other cross-cutting objectives as part of its basic training package. However, a number of consulted MFA staff reported challenges in knowing how to address cross-cutting objectives, especially when their relevance is not obvious in a specific project. This issue is part of a larger project management problem that goes beyond the evaluation function, i.e. if cross-cutting objectives or HRBA are not included in project design documents, they will likely be overlooked in evaluation TORs and reports. MFA staff may require practical guidance on how to operationalise such objectives throughout the project cycle since the basis for integrating HRBA has to be specified in the planning phase of a project.

AHA

Informants also reported shortcomings with the AHA (MFA's electronic project management system). AHA includes process descriptions for evaluations for bilateral programmes, with associated templates and guidelines. Officers need to use the AHA system to get a project approved, but it is not mandatory to use it afterwards, hence the system is missing some information on evaluations among other things (including TORs, ITT, reports, budget, etc.). However, according to informants, AHA is barely used during the implementation and day-to-day management of a project because it is reportedly complicated and difficult to make changes (e.g. it is not possible to edit the text, and documents written in Word have to be copied and pasted). AHA was not considered useful for embassy staff and is mostly centred on the needs of Helsinki. These limitations affect the timeliness of the assignment and go beyond the evaluation cycle.

7.3 Linkage between the Quality of TORs and Reports

Finding 32: Based on TORs and evaluation reports reviewed for this meta-evaluation, there is some evidence of a correlation between well-written TORs and the quality of evaluation reports. However, it is not clear if this is a causal relationship, or if quality of reports may be explained by other variables as well.

Past meta-evaluations have explored the possible relationship between the quality of TORs and the quality of the related evaluation reports. The meta-analyses of 2007 and 2009 both concluded that high quality TORs were a pre-

condition for achieving high quality evaluations, although the authors experienced difficulties in accessing all the related TORs. The 2012 Meta-evaluation, which had access to all TORs, was not able to find a correlation between high quality TORs and high quality reporting, and rather suggested a need for identifying other factors that influence the quality of reporting. In examining the sample of 57 reports from 2012-2014, there is evidence of a positive relationship between the quality of TORs and reports.

First, it must be noted that the TORs in the sample received higher overall quality ratings (84% good or higher) than reports (77% good or higher), as shown in the table below. For example, TORs were twice as likely to receive a rating of Very Good, and less likely to receive a rating of Poor compared to a report.

Table 3 Overall quality rating of TORs vs evaluation reports (based on 57 sets of TORs/reports)

Quality Rating	TORs	Reports
Excellent	0%	2%
Very Good	33%	16%
Good	51%	59%
Poor	16%	23%

A significant percentage of evaluation reports were considered of lower quality than their corresponding TORs, as shown in the table below.

Table 4 Comparing the overall quality rating of TORs with the corresponding evaluation reports (based on 57 sets of TORs/reports)

Improvement (the report received a higher rating than the TORs)	No change (the report received the same rating as the TORs)	Decline (the report received a lower rating than the TORs)
14%	51%	35%

Although these data suggest that good quality TORs do not necessarily result in evaluation reports of equal quality, a more in-depth analysis suggests a correlation between higher quality TORs and higher quality evaluation reports. Of the 19 TORs (33%) rated Very Good for quality, not one had a subsequent report rated Poor. This group of Very Good TORs was also twice as likely (37%) to be followed by a Very Good/Excellent report than the overall proportion of Very Good/Excellent reports in the sample (18%). Similarly, of the 9 (16%) TORs rated Poor, none of their subsequent reports were rated better than Good.

However, these observations do not necessarily indicate a causal relationship, and there may be other factors that also affect quality of reports, such as stronger or weaker staff capacity for evaluation management overall. Thus, while good TORs are necessary, they are not sufficient to guarantee good evaluations.

While there is some evidence of a correlation between well-written TORs and the quality of evaluation reports, other factors may affect quality.

The meta-evaluation team considered that 43% of TORs allocated inadequate resources to carry out the the evaluation.

Overall, evaluation budgets represented approximately 0.89% of total Finnish financial contributions to the project budget.

7.4 Adequacy of Evaluation Resources

Finding 33: Inadequate evaluation budgets could be adversely affecting the quality of MFA decentralised evaluation reports.

The level of financial resources allocated for an evaluation affects an evaluation in several ways including the scope of data collection, the types of methods that can be used, the levels and types of expertise that can be engaged, as well as the level of effort (number of person days). The 2012 Meta-evaluation was not able to draw conclusions about the adequacy of evaluation resources in general, as most of the TORs did not include information on the evaluation budget or the budgeted level of effort.

The **adequacy of evaluation resources** was one of the aspects examined during the 2014 meta-evaluation, and was included in the TORs quality rating tool as Question 10.3: “Are the resources allocated sufficient to carry out the evaluation (alignment between scope and resources)?” With access to all related TORs, and to the accompanying ITT for 60% of the evaluations, information on evaluation budgets was available for 41 of the 57 evaluations reviewed (72%), and the TORs for 13 evaluations provided an estimated level of effort but no financial amount. This information was used by the meta-evaluation team to assess the adequacy of evaluation resources.

Of the 54 TORs/ITT that included enough information on the evaluation budget or estimated level of effort to make some assessment of the adequacy of resources, the meta-evaluation team considered that 43% (or 23 TORs) had inadequate resources allocated given the scope of the evaluation. In the team’s estimation, the level of effort was frequently deemed insufficient to support the proposed methodology, answer the evaluation questions listed in the TORs, conduct sufficient field missions, and write the evaluation report. This conclusion is supported by numerous comments found in the reviewed evaluation reports whose authors reported limitations of time and/or budget to complete various aspects of an evaluation and/or thoroughly address all evaluation questions.

Of the evaluations for which budget information was available in reviewed documents, the evaluation budgets represented approximately 0.89% of total Finnish financial contributions to the project budget. In instances when a project had other donors, their financial contributions were excluded from this calculation; if included, the percentage spent on evaluation would decline considerably. The rule of thumb used by one UN entity (UN Women) is that between one and five percent of the total annual project/program expenditures be allocated for evaluation. MFA reports that overall, its total expenditures on evaluation are in the range of 0.3 to 0.5 percent of overall annual project/program expenditures, considerably lower than UN Women’s guidelines.

To examine whether the adequacy of resources might have had an impact on the quality of evaluations in the 2012-2014 evaluations, the meta-evaluation team compared the overall quality rating of the evaluation reports with the perceived adequacy of the evaluation budget of those 41 evaluations where such information was available (see Table 5). Although an identical proportion of reports with adequate and inadequate budgets achieved a Very Good rating, a much larger proportion of reports with an adequate budget were rated Good,

and a much smaller proportion Poor. This suggests that evaluation budgets could be contributing to the quality of reports. If this is the case, a significant proportion of MFA evaluations could be affected.

Table 5 Quality of reports compared to adequacy of evaluation budget (based on 41 reports)

Report Quality Rating	Inadequate evaluation budget (n=19)	Adequate evaluation budget (n=22)
Very Good	14%	14%
Good	45%	77%
Poor	27%	9%

One of the factors affecting the quality of TORs, which was raised in previous meta-evaluations and that also emerged in the 2014 meta-evaluation, was the large number of evaluation questions in the TORs. The 2012 meta-evaluation concluded that “In a majority of TORs the number of evaluation questions was often too many and not prioritised”, and suggested that questions be reduced and prioritised to help guide the consultants to focus on the most relevant issues. As evaluation budgets will likely continue to have constraints, focusing evaluation questions to limit the scope of an evaluation is one reasonable approach to managing quality while maintaining budgets.

7.5 Meta-Evaluation Process

Finding 34: MFA’s current approach to meta-evaluations inhibits comparative analysis of the quality of TORs and evaluation reports over time, and does not provide a reliable basis to assess the effectiveness of any changes MFA makes to its evaluation capacity building and management processes.

MFA has been carrying out fairly regular meta-evaluations of its decentralised evaluations since 2007, which puts it in the same league as the United Kingdom and the European Commission, which also conduct regular meta-evaluations. Canada’s DFATD is currently looking at institutionalising such a process in 2015.

However, while MFA has carried out five meta-evaluations over the past eight years, it is not possible to compare their findings due to the lack of standardised meta-evaluation approaches, review tools and criteria, and the absence of reliable baseline information. This limits the potential utility of MFA’s meta-evaluation process, leaves the MFA without a reliable basis to assess the possible effects and impacts of any changes it makes (for example in its evaluation management and capacity development processes) on the quality and utility of evaluation TORs and reports, and limits the data to monitor performance in development co-operation. Overall, this represents a lost opportunity for MFA and its stakeholders.

MFA’s current approach to meta-evaluation inhibits comparative analysis and is considered a lost opportunity for MFA and its stakeholders.

The ratings of Finland's development co-operation based on OECD/DAC criteria were fairly similar in the 2012 and 2014 meta-evaluations.

7.6 Comparison of current and previous meta-evaluations

Finding 35: While differences in approach preclude direct comparison, the ratings of Finland's development co-operation based on OECD/DAC criteria were fairly similar in the 2012 and 2014 meta-evaluations; areas for improvement were much the same in both studies.

In the previous chapters of this report, some comparisons with previous meta-evaluations were noted where possible and relevant. However, it is not possible to directly compare findings of the current meta-evaluation with those conducted in previous years, as they were not conducted with a standardised approach (see section 7.5). The meta-evaluation team has some observations about this issue.

Although the 2007 and 2009 meta-analysis concluded that high quality **TORs** were a precondition for achieving high quality evaluations, the 2012 meta-evaluation was not able to find a correlation. In the current meta-evaluation, there was some evidence of a possible relationship between the quality of TORs and reports, though not necessarily causal.

The frequently large number of poorly prioritised **evaluation questions** in the TORs was highlighted in both the 2012 and 2014 meta-evaluation.

The 2012 meta-evaluation included all evaluation reports in its assessment of development co-operation, regardless of how they were rated for quality, while the 2014 meta-evaluation considered only those reports rated Good or higher. This makes comparisons between the two meta-evaluations even more delicate. Given this caveat, ratings on OECD DAC development effectiveness criteria in 2014 were roughly consistent with the 2012 meta-evaluation. **Relevance** continues to be strong, and the most highly rated of all the criteria. **Efficiency** continues to be an area of concern, with about two-thirds of projects in both 2012 and the current study receiving negative ratings. On a related note, the 2009, 2012 and 2014 meta-evaluations suggested that **risk management** in projects was inadequately addressed. **Sustainability** of project results was also rated poorly in the 2012 and 2014 meta-evaluations, with almost two-thirds of projects receiving negative ratings in both. **Overall effectiveness**, however, may be an exception, as a higher number of projects received positive ratings compared to the 2012 meta-evaluation. Both the 2012 and 2014 meta-evaluations found that **cross-cutting objectives** were not systematically integrated and addressed by projects.

7.7 Summary

The meta-evaluation noted the absence of a strong **evaluation culture** within the MFA which limits the use and quality of decentralised evaluations. The challenges identified include the lack of effective results-based management practices and high **staff turnover and rotation**. There is often a lack of clear guidance on how to put **evolving policies** into practice.

There seems to be a correlation, though not necessarily a causal relationship, between the **quality of reports** and the **quality of TORs**; higher quality TORs were more often followed by higher quality reports, and lower quality TORs were

more often followed by lower quality reports. The meta-evaluation team considered that **evaluation resources** were inadequate in 43% of the evaluations for which information was available. Evaluations with inadequate budgets tended to be rated lower in quality.

Finally, while MFA has carried out five **meta-evaluations** over the past eight years, it is not possible to compare their findings due to the lack of standardised meta-evaluation approaches. This limits the potential utility of the meta-evaluation process, and represents a lost opportunity for MFA and its stakeholders.

8 CONCLUSIONS

8.1 Introduction

This chapter summarises the main conclusions emerging from: i) the review of the quality of 57 MFA decentralised evaluations carried out between 2012 and 2014 and their accompanying Terms of Reference; ii) the assessment of Finland's development co-operation based on 34 MFA decentralised evaluation reports; and iii) the review of MFA's evaluation capacity building and management practices.

8.2 Quality of MFA Decentralised Evaluation Reports and TORs

Conclusion 1: Overall, the quality of most decentralised MFA evaluation reports and TORs was rated good or better. While there is some evidence of a correlation between well-written TORs and the quality of evaluation reports, it is not clear if this is a causal relationship, or if quality of reports may be explained by other variables as well.

The majority of reviewed **evaluation reports** were judged to be of acceptable quality overall (77%), providing MFA clients with coherent reports that included good descriptions of the evaluated object (including project components, budget, and key stakeholders), purpose and project objectives, evidence-based findings and useful recommendations. **Evaluation TORs** received similarly positive ratings, with 84% receiving ratings of good or very good. Most TORs reviewed were logically structured and well written.

The meta-evaluation found evidence of a positive relationship between the quality of TORs and reports, but no clear causal relationship as there are other factors that also affect the quality of reports. Data analysis suggests a correlation between higher quality TORs and higher quality evaluation reports.

Conclusion 2: The majority of reviewed evaluation reports were judged to be of acceptable quality overall. They generally provided MFA clients with evidence-based findings and useful recommendations, although these did not always have a clear target. The ten reports that received the highest overall quality ratings (very good or excellent) received high ratings for the evaluation methodology sections. Notable shortcomings in the evaluation reports overall included uneven coverage of HRBA and cross-cutting objectives and certain methodological limitations.

Most reports adequately indicated data sources and integrated the OECD/DAC evaluation criteria. The overall structure of most reports was considered adequate, executive summaries tended to cover key messages, and the annexes provided were generally considered useful. Notable weaknesses included: inad-

Most decentralised evaluation TORs and reports are considered of good quality. Data suggests some correlation between higher quality TORs and higher quality evaluation reports.

The majority of evaluation reports provided evidence-based findings and useful recommendations.

equate descriptions of the evaluation context (including the Finnish development policy context and/or the country/institutional context in relation to the evaluated intervention); incomplete or missing descriptions of the evaluation scope (temporal and/or geographic); methodological limitations (describing the rationale for methods and samples and/or methodological limitations); the frequent absence of lessons learned; the absence of targeted recommendations; and weak conclusions.

While coverage of the standard set of OECD/DAC criteria was strong (i.e., relevance, effectiveness, sustainability and efficiency), other areas of interest to MFA were often not addressed. For example, risk management was mentioned in 53% of the 57 reports; reports typically did not have a dedicated risk management section or provide in-depth analysis of risks.

Another element frequently missing was the cross-cutting development objectives of gender equality, reduction of inequality, and climate sustainability. Of these, climate sustainability was covered the least, explained in part by its more recent introduction to MFA policy in 2012, after the projects in the current sample were designed. Some dimensions of sustainability were also poorly covered; as the MFA introduced more specific requirements for sustainability in the 2012 Bilateral Manual, an improvement in coverage might be expected in future evaluations.

Conclusion 3: Most TORs reviewed were well written and articulated the requirements of the evaluation. Notable shortcomings included a lack of information on methodological aspects, an unreasonable number of evaluation questions, and the absence of evaluation questions on MFA priorities including human rights-based approaches and cross-cutting issues.

TORs generally specified the project objectives, rationale, and purpose, and made use of OECD/DAC and MFA evaluation criteria, and usually clarified the roles of various project stakeholders. The TORs usually included sufficient references to the phases of the evaluation and the expected results/deliverables.

Notable weaknesses in the TORs mirrored some of the shortcomings noted in the evaluation reports: incomplete descriptions and analysis of the programme context (particularly in relation to Finland's development policy and cross-cutting objectives); and limitations in the methodological sections of the TORs, including: lack of /or inappropriate proposed methodologies, limited guidance on methodology, lack of information on how data analysis should be conducted, lack of references to project evaluability and ethical considerations, insufficient attention to quality assurance, and in many cases, an unreasonable number of evaluation questions to be answered.

While MFA has underlined the importance of integrating cross-cutting objectives (formerly themes and issues) into evaluation TORs, methodology and evaluation reports, the quality assessment of TORs indicates that these objectives are only partially integrated in MFA decentralised evaluation processes and products. In the TORs reviewed, for example, evaluation questions frequently did not include cross-cutting objectives. In the evaluation reports reviewed, the evaluation methodologies and analysis paid more attention to gender and reduction of inequalities than to climate sustainability. In cases where climate

Most TORs covered the requirements of the evaluation. Gaps: methodology, number of evaluation questions, and questions on cross-cutting objectives.

Evaluated projects were rated most positively for relevance and effectiveness and significantly lower for efficiency and sustainability. Results achievement was most evident for completed projects.

change was considered applicable to the evaluation or to the programme, half of the reports adequately addressed this issue in the description of the evaluation approach. The meta-evaluation also considered the integration of human rights-based approaches, and found that it was often absent without justification in evaluation reports. Questions on human rights were missing in one third of the TORs reviewed.

8.3 Assessment of Finland's Development Co-operation

Conclusion 4: Overall, the quality of Finland's development co-operation was rated most positively in terms of the relevance and effectiveness of evaluated projects and least positively in relation to their efficiency and sustainability of results. When addressed, risk management was rated poor. Ratings were mixed on cross-cutting objectives and human rights-based approach.

Based on the information available through evaluation reports, the majority of the 34 projects in the 2012-2014 sample were achieving at least some of their objectives and making progress on expected results and project purposes. This was particularly evident for completed projects undergoing a final evaluation (as opposed to a mid-term evaluation). Many projects that were completed or near completion were not able to demonstrate significant contribution to overall objectives. However, in most cases it is not reasonable to expect long term overall objectives immediately on completion of a project.

As in past meta-evaluations, relevance emerged as a very strong element of Finland's development co-operation, with considerable evidence of project congruence with country, global, and/or MFA priorities.

Most projects faced challenges with efficiency, stemming either from the initial design or from weaknesses in the implementation process. Delays in implementation were most common, suggesting somewhat unrealistic expectations of what could be accomplished, especially given the difficult environments in which many projects operate, and compounded by the technical complexity of some projects. Unsurprisingly, low capacity of human resources was reported to be a recurring challenge, sometimes pertaining to difficulties in securing qualified project staff or in building the capacity of beneficiaries.

Sustainability, as in past meta-evaluations, was an ongoing challenge. While some aspects of projects were considered sustainable, overall sustainability (i.e. the likelihood that benefits would continue after project completion) was weak in most projects. Projects with more likelihood of sustainability were typically those in which activities or structures were being incorporated or institutionalised in a government unit or some other entity. Prospects for financial/economic sustainability were particularly low.

Where it could be assessed by the meta-evaluation team, project congruence with MFA policies on human rights-based approaches and the cross-cutting objectives of gender equality, reduction of inequality, and climate sustainability was high. However, this did not usually result in a high level of effectiveness in achieving those objectives. In many cases, for example, gender equality or reduction of inequality was included in the project design, but the focus on these was not reflected in project implementation.

Risk management was lacking in many projects. Risks were often not taken into account during the design phase, and risk assessments were not always complete or accurate. Where mitigating strategies were not in evidence, emerging risks contributed to delays in implementation, and in some cases became a barrier to projects achieving their intended project objectives.

8.4 Overall observations on MFA evaluations and meta-evaluations

Variables that affect quality of evaluations

Conclusion 5: In developing the TORs for meta-evaluations over time, MFA has repeatedly noted its interest in understanding the link between various variables such as: the quality of evaluation TOR and reports, the quality of evaluation reports and evaluation budgets, the quality of evaluation reports and the competencies of evaluation teams and so forth. As noted in the 2012 meta-evaluation, it is often difficult to answer such questions as there are many associated variables and assumptions (known, unknown or assumed).

This meta-evaluation identified several factors that could affect the quality of evaluation reports. For instance, the absence of a strong evaluation culture and the lack of effective results-based management practices in MFA may contribute to noted limitations in the quality of decentralised evaluation reports. Staff turnover and the need to adapt to new policies were also identified as institutional limitations to evaluation capacity building. This explains why the MFA has invested in capacity building for decentralised evaluations and has provided a broad range of activities to increase the evaluation skills of staff.

Another contributing factor might be the inadequate evaluation budgets (often too small given the scope of the evaluations) that could adversely affect the quality of MFA decentralised evaluation reports.

Mapping such variables, associated assumptions, and how they relate to one another and ultimately to the quality and use of evaluation results in decision making could be a valuable and exemplary exercise for EVA-11's consideration.

Comparison of meta-evaluations

Conclusion 6: The lack of institutionalised meta-evaluation tools and guidelines limits the potential effectiveness, efficiency, and utility of MFA's current meta-evaluation process. Being able to compare the findings of meta-evaluations over time would increase efficiency and reduce the time spent developing and testing new tools and guidelines. It would also allow MFA to make more meaningful comparisons of ratings over time, and examine the effects of its actions to address quality shortcomings (e.g. additional capacity building support, or changes in evaluation management practices).

Quality Assurance

Conclusion 7: While the majority of TORs and reports were considered of good or better quality, a closer review indicates that the bulk of these were considered good rather than very good or excellent. If MFA wishes to increase the overall quality of evaluation TORs and reports, greater attention would need to be paid

Mapping variables, assumptions, and how they relate to the quality and use of evaluation results could be a valuable exercise for EVA-11.

Being able to compare the findings of meta-evaluations over time would increase efficiency.

Greater attention to quality assurance is needed throughout the evaluation management cycle.

A framework agreement for decentralised evaluations could increase efficiency and provide some assurance regarding the quality of evaluators.

Few MFA Directors have the skills to ensure that evaluation TORs and reports meet quality standards and that recommendations are well utilized.

to quality assurance throughout the evaluation management cycle - starting from the point when the project is being designed (e.g. to ensure that cross-cutting objectives have been integrated, that systems are in place to collect data on results, that Finland's development partners are involved, that budget and scope of TORs are congruent, and so forth). Attention then needs to be continued during the development of the TORs, the allocation of financial and other resources to conduct and monitor the evaluation, evaluation implementation, and evaluation follow up.

Evaluation management capacity

Conclusion 8: Consultations with MFA officers on capacity building services and evaluation management revealed that officers' own limited knowledge and skills in evaluation have adversely affected their ability to select and manage evaluators.

While EVA-11 has established a framework agreement to support it in commissioning corporate evaluations, no similar mechanism exists for decentralised evaluations. The majority of MFA representatives interviewed indicated that a mechanism such as a framework agreement would increase the efficiency of the selection and contracting processes, provide some assurance regarding the quality of evaluators, and reduce the workload of evaluation managers.

Evaluation capacity building services

Conclusion 9: As MFA is planning an in-depth review of its evaluation capacity building services in the coming year, the 2014 meta-evaluation included a modest review of these services. The review identified a number of gaps that could be explored in this in-depth review.

Evaluation oversight and follow up

Conclusion 10: While some of the noted shortcomings in evaluation TORs and reports can be addressed through skills upgrading of MFA Directors, Officers and Consultants, MFA Directors also have an important oversight role in ensuring that evaluation TOR and reports meet acceptable quality standards and that evaluation findings and recommendations are well utilized.

The meta-evaluation revealed that to date, few MFA Directors have had the opportunity to develop sufficient knowledge and skills to provide this oversight role effectively, evidenced by the noted limitations with some evaluation TORs and reports. Moreover, interviewed persons within MFA suggest that this situation is exacerbated by the absence of a strong evaluation culture in the Ministry.

Attention to MFA policies and directives

Conclusion 11: The quality of MFA decentralised project implementation as well as MFA-commissioned evaluation reports is negatively affected by limited attention to some MFA policies and directives. While some of these are relatively new or have evolved over time (e.g. HRBA was officially introduced in the 2012 Development Policy Programme), others are well-known and have been identified as limitations in previous meta-evaluations (e.g. cross-cutting objectives related to gender).

The lack of consistent attention to these matters negatively affected MFA's development effectiveness ratings.

These limitations go beyond the mandate of EVA-11, and reflect shortcomings in MFA's decentralised project management practices overall in ensuring that new MFA policies and practices are addressed.

Limited attention to some MFA policies and directives reflect shortcomings in MFA's decentralised project management practices.

9 RECOMMENDATIONS

EVA-11 should consider developing a theory of change.

The following recommendations are based on the findings and analysis presented in the report.

Recommendation 1: To guide future support and actions intended to improve the quality and use of MFA evaluations, EVA-11 should consider developing a theory of change that outlines and links expected results and assumptions of conducting evaluations with evaluation use and evidence-based decision making.

This meta-evaluation (like similar processes undertaken by MFA peers) relied on a narrow set of documents/information (mainly TORs and evaluation reports).

The implicit assumption was that a review of these would shed light on the quality of MFA evaluations (and ultimately on the effectiveness of Finland's development co-operation) and help identify areas for improvement in how MFA carries out evaluations in the future. However, there are many variables that can affect the quality of evaluations, some of which are noted in this report (e.g. the budget, quality assurance processes, etc.) and others that were not examined in this meta-evaluation (e.g. the competence of the evaluation team, the quality of the inception report). In addition, there are several implicit/explicit assumptions about the expected results of MFA's actions to improve the quality of evaluations that are not systematically examined in its current meta-evaluation process.

EVA-11 should consider developing a theory of change that explicitly links the conduct of MFA evaluations (decentralised and perhaps others) to the use of the evidence generated by such evaluations. This exercise should in turn help EVA-11 clarify the various types of capacity building support, management practices and resources required to support a healthy evaluation culture and effective practices in MFA. It should also provide a framework for informing the scope and objectives of future MFA meta-evaluation processes.

Box 43. Practices in other agencies

- UN Women's Independent Evaluation Office has developed a Theory of Change (ToC) to strengthen the UN Women evaluation function as part of its 2014-2017 Evaluation Strategic Plan. The ToC takes into account UN Women's role in promoting gender responsive evaluation and aims to strengthen the capability of managers to demand and use evaluation, as well as the capability of evaluation specialists to supply high-quality evaluative evidence.

Recommendation 2: In order to enhance the long term utility of meta-evaluations, EVA-11 should consider institutionalising a regular meta-evaluation process and accompanying review tools and guidelines.

The lack of institutionalised meta-evaluation tools and guidelines limits the potential effectiveness, efficiency, and utility of MFA's current meta-evaluation process. Being able to compare the findings of meta-evaluations over time would increase efficiency and reduce the time spent developing and testing new tools and guidelines. It would also allow MFA to make more meaningful comparisons of ratings over time, and examine the effects of its actions to address quality shortcomings (e.g. additional capacity building support, or changes in evaluation management practices).

Box 44. Practices in other agencies

- DFID and Sida have had established meta-evaluation processes in their agencies for several years.
- In 2015, DFATD is in the process of establishing such a process for its decentralised evaluations.

Should MFA decide to institutionalise this meta-evaluation process in the near future, it should consider the following elements:

- Review the last three MFA meta-evaluation processes to identify strengths, areas for improvement, lessons learned, a strategy and priorities for future meta-evaluation processes. Ideally, this would result in a meta-evaluation process that would be used for a fixed period (say 3-5 years), as well as a schedule and process to review and revise the process periodically.
- Examine the potential advantages of carrying out meta-evaluations on an annual basis
- Finalise meta-evaluation tools and guidelines so that MFA is well-positioned to launch the next meta-evaluation
- Clarify the population (and samples) of decentralised evaluations to be subjected to quality assessments and reviews of Finland's development co-operation effectiveness in the future (e.g. whether the sample should include mid-term evaluations and/or evaluations commissioned by agencies other than MFA)
- Inform MFA officers and evaluators about its expectations and their obligations related to the quality of evaluation products for which they are responsible
- Establish a feedback form or process (e.g. an online survey) to gather views on completed evaluations by key users and management (e.g. was the evaluation good quality? did it provide information at the right level for their purposes?). This could be a meaningful and timelier source of information to incorporate into future meta-evaluations or ongoing quality assurance processes.
- Consider the calendar time allocated for meta-evaluations, given that the quality management process required to ensure consistency/reliability in ratings is time intensive (for example, the 2014 meta-evaluation took five months and was still constrained by time)

EVA-11 should institutionalise a regular meta-evaluation process.

MFA should increase attention to quality assurance throughout the evaluation cycle.

- Consider including other types of documents to review for the next meta-evaluation, including inception reports or other programme design documents
- Work with MFA's regional programmes and the unit(s) responsible for managing AHA (MFA's case management system) to identify a strategy to address the noted shortcomings with the system so that all relevant documents on evaluations are systematically stored in the system in the future (e.g. ITT, TORs, evaluation budgets, inception reports, evaluation reports and management responses to evaluations).

Recommendation 3: If MFA wishes to increase the overall quality of evaluation TORs and reports on decentralised evaluations, it should increase the attention paid and resources allocated to quality assurance throughout the evaluation management cycle.

The review of the current MFA evaluation cycle pointed to gaps in the evaluation quality assurance process, noting that most attention was paid to quality assurance in the earlier stages of the cycle (e.g. during the development of TORs and tendering processes). The implementation of the evaluation and associated reports are often not submitted to equally rigorous review processes. Consulted MFA representatives underlined that as clients for the evaluation, MFA officers are responsible for communicating their expectations for a useful, quality evaluation process and deliverables.

Box 45. Quality Assurance and Evaluability Practices in other agencies

- DFATD, DFID and Sida have defined quality assurance expectations and guidelines for inception reports and evaluation reports which are shared with consultants and used by agency representatives to provide feedback to evaluators.
- DFATD's Evaluation Group provides a one-hour briefing to consultants at the start of each assignment to clarify its requirements, and provides detailed feedback to consultants on deliverables which the consultants are required to address before deliverables are approved.
- DFATD guidelines for decentralised evaluations include a requirement to carry out evaluability assessments.

In order to address these shortcomings, we suggest that MFA consider the following measures to enhance its quality assurance practices.

- Ensure that staff responsible for reviewing the quality of and approving Terms of Reference, Inception Reports and Evaluation Reports have the right skills to do so
- Develop and disseminate necessary tools and guidelines to MFA staff and consultants to support and guide reviews of the quality of evaluation products
- Provide practical training to MFA staff (Directors and officers as required) in using quality assurance tools and guidelines
- Institutionalise evaluability assessments as part of the evaluation management cycle to increase the effective use of resources in evaluations

(i.e. that resources are not wasted in trying to evaluate in impossible situations)

- During the implementation of an evaluation, close follow-up of the work of consultants should be encouraged. Emerging findings and conclusions should be presented before a full draft report is submitted to ensure that deliverables meet MFA expectations.

Box 46. Evaluability

- The extent to which an activity or a program can be evaluated in a reliable and credible fashion.

Recommendation 4: To increase the potential efficiency of processes used to recruit and contract qualified evaluation experts, MFA should consider establishing a framework agreement(s) for decentralised evaluations. To ensure that it has ready access to sufficient numbers of qualified evaluators, MFA should consider ways to facilitate participation by individual evaluators as well as larger firm contractors.

Consultations with MFA representatives indicate that they invest considerable time in identifying, recruiting and contracting qualified evaluators to carry out evaluations. While EVA-11 has established a framework agreement to support it in commissioning corporate evaluations, no similar mechanism exists for decentralised evaluations. The majority of MFA representatives interviewed indicated that such a mechanism would increase the efficiency of the selection and contracting processes and might also provide some assurance regarding the quality of the evaluators. It would also help reduce the workload of evaluation managers.

Such a mechanism would be in keeping with the practices of other donors. At the time of writing, we understand that at least one regional department in MFA is taking steps to establish its own framework agreement. MFA should consider establishing a mechanism that would benefit all MFA regional programs, and ways to encourage qualified individual evaluators and larger firms to compete for the framework agreements.

Box 47. Practices in other donor agencies

- Since at least 1990, Canada's DFATD has had Evaluation Standing Offers (mechanisms similar to framework agreements) which it uses to identify qualified evaluators individuals and firms to support its decentralised and corporate evaluation processes.

Recommendation 5: Given the kinds of questions on evaluation capacity building services that emerged in this meta-evaluation, EVA-11 should undertake the planned review of MFA's evaluation capacity building support as soon as possible. The review should focus on areas of weaknesses identified in this meta-evaluation.

The cursory review of MFA evaluation capacity building support in Chapter 7 suggests that EVA-11 provides the right mix of evaluation capacity development

MFA should consider establishing a framework agreement(s) for decentralised evaluations.

EVA-11 should undertake the planned review of MFA's evaluation capacity building support.

MFA should clarify senior management's responsibilities for evaluation quality assurance, follow-up and use of results.

services, but identified a number of questions that could be explored in greater depth by EVA-11 in 2015 as part of the planned in-depth study. Examples of questions that might be considered in this study include:

- To what extent do MFA staff have the skills to carry out their responsibilities for evaluation planning, management, and follow up?
- Should EVA-11 carry out regular needs assessments of the evaluation capacity of its staff (and/or others)? If so, what should be the scope, focus, and frequency of such needs assessments? See Box 48 for examples of the kinds of guidance that might be helpful.
- Should MFA increase the access of embassy-based staff to evaluation capacity building support? If so, how?

Box 48. Evaluation Capacity Building Needs Assessments

- **Scope:** Only MFA staff? If so, what levels (Directors, officers)? Helsinki only or embassy staff? Consultants?
- **Focus:** Writing TORs? Managing evaluations? Reviewing the quality of reports? Addressing cross-cutting objectives? More advanced course content?
- **Frequency:** Annual? Bi-annual?

Recommendation 6: To increase the utility of decentralised evaluation reports and processes in the future, MFA should clarify and institutionalise senior management's roles and responsibilities for quality assurance throughout the evaluation management cycle, including follow-up and use of evaluation results.

The main concern linked to the evaluation management cycle was the lack of follow-up on evaluation recommendations. This issue was particularly important with final decentralised evaluations for which there are currently no clearly defined processes for how recommendations should be used. Responsibilities for sharing associated lessons learned are also unclear.

EVA-11 and evaluation units in other bilateral and multilateral agencies do follow up on the implementation of evaluation recommendations emerging from corporate or centralised evaluations (those directly managed by evaluation units). This is typically based on the agency's Management Response to each evaluation and the progress on implementation is reported to the Senior Management and the Board. The implementation of such a system for decentralised evaluations is more challenging and requires:

- Management responses on each evaluation that are saved in a data base
- Action plans that are developed as part of the Management Response
- An entity that is responsible for monitoring implementation
- Senior management who value the use of performance information and see this as one measure of how feedback on performance (via evaluative evidence) is being used to make improvements in programming or policy in the organisation.

These different components may/may not be present for all of MFA decentralised evaluations.

To address noted limitations with the use of decentralised evaluations, MFA should consider establishing a way to keep track of the follow-up on recommendations and management responses and how recommendations are addressed or used.

Moreover, since decentralised evaluation reports should be useful to senior managers and regional managers in planning future interventions, MFA should consider taking steps to clarify their responsibilities for ensuring and demonstrating how evaluation findings and recommendations are used.

Box 49. The use of evaluation results

- This is also an issue in other organisations, such as the European Commission. In a recent report on the update of evaluation recommendations, the EC recommended that:
 - 1) more incentive and promotion on learning and evaluation should come from political leadership and senior management;
 - 2) to review the evaluation process so as to enhance ownership and uptake (e.g. evaluation unit should assume a more proactive role in knowledge translation and brokerage in order to ensure that evidence is readily available and accessible, explore new ways to enhance participation of users, etc.);
 - 3) formal and informal processes for uptake should be better exploited (e.g. better aligning evaluations with policy development as well as programming cycle);
 - 4) strengthening the focus on outcomes (project purposes) in both evaluations and the management response system.

Source: Development and Co-operation EuropeAid (2014). *Study on the uptake of learning from EuropeAid's strategic evaluations into development policy and practice: Final report*. 61 pages.

Recommendation 7: To address noted shortcomings in TORs and evaluation reports related to coverage of human rights-based approaches, cross-cutting objectives, sustainability, and risk management, MFA should try to identify the causes of these shortcomings and address them.

Chapter 7 highlighted several significant shortcomings related to the limited information in many of the evaluation reports related to human rights-based approaches (HRBA), several of the cross-cutting objectives, and risk management. While some of these shortcomings may be because MFA policies and directives have evolved over time (e.g. HRBA was officially introduced in the 2012 Development Policy Programme), others are well-known and have been identified in several meta-evaluations in the past (e.g. cross-cutting objectives related to gender).

The meta-evaluation team lacked the information to carry out an in-depth assessment of the reasons for this situation and is unable to answer questions such as those listed in Box 50.

Box 50. Questions about limited coverage in MFA evaluations

Are noted problems due to:

- Ways in which the TORs are written and reviewed?
- Limited clarity among MFA staff about what is required?
- The competence and experience of the evaluation team?
- The adequacy of quality assurance processes for TORs, inception and evaluation reports?
- The lack of integration of these components in the overall programme design?

MFA should address limited evaluation coverage of HRBA, cross-cutting objectives, sustainability, and risk management.

MFA might want to review these questions as part of its capacity building review or as a separate exercise. Examples of possible approaches to address noted limitations with coverage of HRBA and sustainability are discussed below.

HRBA

The meta-evaluation team noted that many MFA-funded projects have strong potential to contribute to human rights and the opportunity to develop a stronger HRBA through implementation. The following preliminary suggestions are provided for MFA's consideration, with the caveat that the meta-evaluation team has not carried a thorough review of all MFA policies and guidelines on HRBA. MFA may wish to:

- Provide assistance and/or a quality check in the design stage to support programme teams in applying a clear HRBA framework; this may include conducting an initial analysis to identify rights issues, incorporating rights language in the framework (for example, making clear who the rights holders and duty bearers are) and including indicators in the log-frame to measure progress
- Analyse or have quality check carried out by a knowledgeable resource person about the rights issues for a specific programme topic
- Have a more explicit invocation of the relevant international human rights treaties (e.g., CEDAW) and hold programmes accountable to these
- Partner with or use support of NGOs who specialise in HRBA for assistance during implementation.

Sustainability

MFA's bilateral manual provides some interesting guidance on the dimensions of sustainability that should be addressed in MFA projects (and implicitly in the evaluations of projects). Reviewed project evaluations paid mixed attention to the different kinds of sustainability. When it was actually addressed in evaluation reports, it tended to focus on financial/technical sustainability and/or institutional sustainability; relatively modest attention was paid to other types of sustainability such as social, environmental or technical sustainability. Based on information in reviewed evaluations, it appears that sustainability also has mixed coverage in project design and implementation. MFA should identify the kinds of support, guidelines and checks and balances required to operationalize its approach to sustainability throughout the project cycle.

REFERENCES

Development and Co-operation EuropeAid (2014). *Study on the uptake of learning from EuropeAid's strategic evaluations into development policy and practice: Final report*. 61 pages.

EVA-11 (2010). *Evaluation Training Needs Assessment* (internal memo), 3 pages.

Ministry for Foreign Affairs of Finland (2012). *Checklist Climate Proofing*, 3 pages.

Ministry for Foreign Affairs of Finland (2012). *Development Policy Programme*, Government Decision in Principle 2012, 48 pages.

Ministry for Foreign Affairs of Finland (2004). *Development Policy Programme*, Government Resolution 5.2. 2004, 40 pages.

Ministry for Foreign Affairs of Finland (2007). *Development Policy Programme*, Government Decision in Principle 2007, 40 pages.

Ministry for Foreign Affairs of Finland (2007). *Evaluation guidelines*, Between Past and Future, 85 pages.

Ministry for Foreign Affairs of Finland (2013). *Evaluation Manual*, 104 pages.

Ministry for Foreign Affairs of Finland (2012). *Manual for bilateral programmes*, 107 pages.

Ministry for Foreign Affairs of Finland (2008). *Meta-analysis of Development Evaluations in 2007-2008*, 99 pages.

Ministry for Foreign Affairs of Finland (2012). *Meta-evaluation of decentralised evaluations in 2010-2011*, 130 pages.

OECD (2002). *Glossary of Key Terms in Evaluation and Results Based Management*, 37 pages.

Wholey, J. S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J.S. Wholey, H. J. Hatry & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15-39). San Francisco: Jossey-Bass.

THE EVALUATION TEAM

Katrina Rojas

Ms. Rojas is a Senior Evaluator with over 14 years of experience in evaluation methodologies and approaches. Since 2013, she has been responsible for implementing the review process and meta-evaluation for UNICEF's Global Evaluation Reports Oversight System (GEROS). She has been involved in other meta-evaluations for IUCN, the Government of Chile, and for DFATD. She also led the Development Effectiveness Review of the Inter-American Development Bank, which includes a meta-synthesis methodology developed under the auspices of DAC-EVALNET. Since 2010, Ms Rojas has led or advised in the implementation of the Multilateral Organisation Performance Assessment Network's (MOPAN) Common Approach to assessing multilateral organisations. As such, she has managed stakeholder engagement and complex data gathering processes for more than 20 organisational assessments of UN funds and programmes, specialised agencies, humanitarian organisations, and development banks.

Gerry Cooney

Ms Cooney is a credentialed evaluator with over 30 years of experience in leading and managing evaluations of large and complex international development projects, programmes, and organisations in a variety of sectors and themes, often involving large, diverse, and geographically separate teams of evaluators and technical advisors. Over the past three decades, she has worked in over 50 countries around the world on evaluation, evaluation capacity building, results monitoring, and performance measurement assignments. This includes extensive experience in evaluating education, health, information technology, intellectual property, debt reduction, business development, capacity development and infrastructure programmes as well as child protection, gender equality, and human rights programmes.

Mariane Arsenault

Ms. Arsenault has over 8 years of experience in program monitoring and evaluation and organisational assessment. She is very well versed in qualitative and quantitative research methodologies. She has conducted and lead assessments for multilateral and bilateral organisations to assess their interventions. Her portfolio of work includes assignments conducted for several multilateral institutions such as the World Bank, UN agencies, the Caribbean Development Bank among others, as well as bilateral organisation such as the Canadian International Development Agency.

Päivi Pelkonen

Ms. Pelkonen is an expert in education and intercultural learning with over 20 years of experience both in Finland and abroad. She has been involved in various teaching and evaluation-related assignments in the Balkans Region, notably a feasibility study on an Environmental education development programme in the Western Balkans for the Finnish Ministry for Foreign Affairs in 2008 and a multi-country evaluation of implemented and ongoing MFA projects in Podgorica, Belgrade, and Sarajevo in 2007. Ms. Pelkonen is currently completing a PhD degree in Intercultural Learning and International Co-operation.

-
- Mélissa Rodrigue** Ms. Rodrigue is a research associate and evaluation analyst. Her experience in the field of international consulting has included projects in monitoring and evaluation, evaluation report assessment, as well as research in social media and organisational assessment. Ms. Rodrigue holds a professional Master's degree in International Studies from the University of Montreal, where her research focused on international treaties and the adoption of a human rights-based approach to environmental law.
- Kelly Babcock** Ms. Babcock is an evaluation consultant with experience in conducting organisational assessments and project, program, and country level portfolio evaluations both within and outside Canada. She has experience in performance evaluation of both NGOs and governmental bodies, and she has supported NGOs in results-based performance monitoring and reporting. Ms. Babcock is experienced with group facilitation in multicultural settings, and has used these settings to apply her training in participatory monitoring and evaluation. She has also participated in innovative work on applying a form of social value accounting to community economic development organisations, and contributed to the development of the Treasury Board Secretariat Value-for-Money Tool.
- Mary Picard** Dr. Picard has over 19 years of development experience on five continents. She has in-depth experience and expertise in all aspects of monitoring and evaluation and is also recognised as a facilitator and technical advisor in organisational learning and knowledge management. Ms. Picard is a skilled conceptual writer and practitioner in gender, rights-based programming and cross-cutting program quality issues.
- Heather Buchanan** Ms. Buchanan has over 25 years of public sector management experience. Ms. Buchanan specialises in issues of organisation and programme effectiveness in domestic and international contexts. She has extensive experience in the application of performance management principles, theories, methods and tools - including evaluation, performance measurement, organisational and human resource development projects. Ms. Buchanan provides rigorous research, comprehensive analysis and innovative results for performance challenges. She is an analytical and creative consulting professional with articulate communication and dynamic facilitation skills. Ms. Buchanan has developed courses and taught evaluation in a graduate programme at Carleton University, in the World Bank's International Program for Development Evaluation Training (IPDET) and in the UN Evaluation Certificate programme at the United Nations Staff College, as well as specialised evaluation training for UNICEF and UNIFEM.

ANNEX 1 TERMS OF REFERENCE

META-EVALUATION OF PROJECT AND PROGRAMME EVALUATIONS IN 2012-2014

1 BACKGROUND TO THE EVALUATION

Development Evaluation Unit (EVA-11) commissions regularly meta-evaluations in order to summarise the quality of the decentralised evaluations and assess the findings of Finland's development co-operation. This is the Terms of Reference (TORs) for the meta-evaluation of project and programme evaluations (decentralised evaluations) from 2012 to 8/2014. This evaluation will include meta-evaluation of the quality of decentralised evaluations, and a meta-analysis based on selected decentralised evaluations of the effectiveness of Finland's development co-operation.

Evaluations are divided into two functional entities, centralised and decentralised evaluations, within the structure of the Ministry for Foreign Affairs (MFA). The centralised evaluations include comprehensive, strategically important sector, thematic, country programme etc. evaluations conducted by EVA-11.

In addition, the EVA-11 provides help-desk services within the MFA. It arranges trainings to MFA and embassy staff and other stakeholders (e.g. individual consultants, consulting companies, CSOs, other ministries, academia etc.) involved in development evaluations. EVA-11 consults MFA departments, units and embassies, and takes into account the proposed subjects for evaluations. The EVA-11 also compiles a file of planned and finalised decentralised evaluations. In 2013 a new evaluation manual was published.

Decentralised evaluations are the responsibility of the MFA departments and units that are in charge of the development co-operation programmes in specific countries, regions or international institutions. These decentralised evaluations include appraisals, mid-term, final and ex-post evaluations as an integral part of project or programme and programme cycle management. The embassies are involved and facilitate evaluations at country level. Annually some 30-40 decentralised evaluations are carried out.

There were two important findings of the 2012 meta-evaluation that will guide the specific focus of this meta-evaluation. In the meta-evaluation carried out in 2007 and 2009 it was claimed that high quality TORs correlate with high quality reporting. However, such correlation was not found in the 2012 meta-evaluation and other factors than TORs seemed to be more relevant for the quality of an evaluation report. These were the overall institutional context of the MFA, including staffs' skills, time and turnover; the quality and use of policy and practical guidance and tools for effectively addressing project design and implementation; the approval process of decentralised evaluation reporting; the quality of the evaluation team and its technical and reporting skills.

The meta-evaluation 2012 recommended that factors that contribute to or impede high quality reporting should be studied. Meta-evaluations should be complemented with analyses of MFA's management practices of evaluation oversight and quality assurance.

The 2012 meta-evaluation considered the quality of the development co-operation poor, showing little effect towards sustained outcomes. For example, the risk analysis was inadequately addressed although it is crucial for development co-operation to be successful. Identifying and managing risks should be

targeted at every level of the project and high priority in project design. The MFA is starting to develop approaches for systematic risk management and in this meta-evaluation risk management is one of the focus areas.

Furthermore, the 2014 government report on impact and coherence of Finland's development policy describes Finland's new approaches when working in challenging conditions. It emphasises the importance of further developing the effectiveness, openness and risk management of development co-operation.

2 PURPOSE AND OBJECTIVE OF THE EVALUATION

The purpose of the meta-evaluation is twofold: Firstly, this meta-evaluation helps the MFA to improve the quality of evaluations, the evaluation management practices and the evaluation capacity development services. In order to do that the consultant will analyse the decentralised evaluations and their ToRs as well as the MFA's evaluation management practices and capacity development services.

Secondly, the evaluation is expected to bring forward issues and lessons learned which will help the MFA to improve the development co-operation. Meta-analysis based on selected decentralised evaluations can bring together otherwise scattered results of what has been achieved. Meta-evaluation is also seen as a tool for accountability and improved transparency towards partner countries, general public, parliamentarians, academia, media and development professionals outside the MFA.

The objective of the evaluation is firstly to give innovative and out-of-the-box but practical and concrete recommendations to enhance the quality of evaluations by conducting a meta-evaluation of the quality of decentralised evaluation reports and ToRs (2012-2014) against the OECD-DAC and EU quality principles, standards and criteria. This may also indicate the effectiveness and impact of the evaluation capacity development services in the MFA as well as bring up some challenges in the evaluation management practices of the MFA hindering the quality of evaluations. The evaluation will also summarise what sectors, countries, when and where evaluations have been conducted, the budget and the main reasons why the evaluation was conducted.

Secondly, the objective is to give innovative and out-of-the-box but practical and concrete recommendations to improve the results of Finland's development co-operation by conducting a meta-analysis of selected decentralised evaluations. In this analysis, a special attention will be paid to the risk management as described in the Manual for Bilateral Programmes (2012) and to the integration of human rights-based approach and cross-cutting objectives. The meta-analysis will assess if the MFA is doing right things in the right way and analyse the impact of the development co-operation to the extent possible.

The results of this meta-evaluation will be compared to the 2012 Meta-Evaluation of Decentralised Evaluation in order to find trends, patterns and changes.

3 SCOPE

The first part of the meta-evaluation will assess and collate information contained in the decentralised evaluation reports and their TORs from 2012 to 8/2014 (completed reports up to the signing of the contract). The sample of documents contains mid-term evaluations, final and ex-poste evaluations and related ToRs. A preliminary categorisation shows that the decentralised evaluations cover mostly environment, agriculture and rural development sectors (e.g. water, forest and energy) (50 %). Other sectors present are for example human rights (20 %) as well as business development, ICT, technology and innovation (20 %). The total number of decentralised evaluation reports is 60-70. Then the consultant will perform an overall synthesis based on documents. A tentative list of evaluation reports for this meta-evaluation can be found in Annex II.

The first phase of the evaluation is expected to make recommendations on the capacity development materials and guidelines provided by EVA-11. The recommendations will be compared with other Nordic + donor countries. The consultant will identify 1-2 benchmarking countries in order to formulate innovative but practical recommendations on how the evaluation capacity building service should be improved.

This meta-evaluation will not include an assessment of appraisals. This is different from the previous meta-evaluations, and it needs to be considered when comparisons are made.

The second phase is a summative meta-analysis of the effectiveness of Finland's development co-operation verified against the OECD-DAC and EU evaluation criteria and how Finnish development policy's goals have been achieved. It will be based only on those decentralised evaluation reports that are assessed as reliable and of high quality in the first part of the evaluation. Consultant will propose the criteria and list of evaluations for this analysis and the proposed list will be presented in the phase one draft report.

4 EVALUATION QUESTIONS

The main evaluation questions for the first phase (meta-evaluation) are:

1. In what extent decentralised evaluation reports (2012-8/2014) and their TORs are based on OECD/DAC and EU evaluation principles, standards and criteria?

Sub-questions:

- 1.1 Are the reports well-written, can they be easily understood and are they user-friendly?
- 1.2 What are the main challenges and strengths of the decentralised evaluation reports and ToRs applying OECD/DAC and EU evaluation principles, standards and criteria?
- 1.3 What are the main reasons hindering the quality of decentralised evaluations and reports?
- 1.4 What are the recommendations for evaluation capacity development services?

The main evaluation questions for the second phase (meta-analysis) are:

2. Based on the selected evaluations how effective has Finland's development co-operation been applying the OECD/DAC and EU evaluation criteria?

Sub-questions:

- 2.1. Has the MFA done right things in a right way and did it change what it was supposed to change? What have been the unintended results?
- 2.2. To what extent has the risk management improved in comparison with previous meta-evaluation findings?
- 2.3. How has the human rights-based approach been applied as implied in the Finnish development policy and guidelines?
- 2.4 To what extent have the cross-cutting objectives of development policy been implemented as implied in Finland's development policy and guidelines?

The OECD DAC criterion *impact* has to be included in the analysis; even though it is the understanding that in many cases the material does not provide the grounds for the analysis.

The evaluation will define a limited number of additional sub-questions (up to a maximum of 6 in each part) that will be prepared as a part of the inception report.

In addition, in both parts of the evaluation, the results shall be compared to the 2012 meta-evaluation in order to identify possible patterns and trends of change.

5 GENERAL APPROACH AND METHODOLOGY

The approach and working modality with data collection and analysis methodology as well as tentative timetable will be presented in the mini tender, and further elaborated in the inception report.

The main method used in the meta-evaluation will be document review. The methods of data analysis will be mixed multiple methods (both quantitative and qualitative). An evaluation matrix will be constructed and included in the inception report which will attribute the criteria to the main evaluation questions opened up into specific research questions.

The main sources of information will be the evaluation reports and their ToRs, information related to evaluation capacity development services, Development Policy Programme documents, guidelines, earlier centralised evaluations and Government Reports to the Parliament as well as administrative in-house norms.

The consultant is encouraged to raise issues that are important to the evaluation but are not mentioned in this TORs. Similarly, the consultant might exclude issues what are in the TORs but may not be feasible and those remarks will be presented by latest in the inception report.

6 EVALUATION PROCESS AND DELIVERABLES

The evaluation consists of the following phases and will produce the respective deliverables. A new phase is initiated only when all the deliverables of the previous phase have been approved by EVA-11. The reports will be delivered in Word-format (Microsoft Word 2010) with all the tables and pictures also separately in their original formats.

I. Start-up meeting

The purpose of the start-up meeting is to discuss the entire evaluation including evaluation approach and methodology, practical issues related to the evaluation, reporting and administrative matters.

The start-up meeting will be organized by EVA-11 after the signing of the contract. Start-up meeting will be a video conference.

Deliverables: PowerPoint Presentation of evaluation approach and methodology including tentative evaluation questions and sub-questions and methods for data collection analyses. Power point presentation will be sent to EVA-11 three (3) working days before the meeting.

II. Inception phase

This phase includes preparation of the main evaluation questions and sub-questions, evaluation matrix and the work plan. The general evaluation questions will also be opened into specific research questions and respective indicators. The approach, methodology and sources of verification will be explained in detail, including the methods and tools of analyses, scoring or rating systems and alike.

The division of tasks between the team members will be finalised in the inception report. The inception report will also suggest an outline of the final report. The structure of the report will follow the established overall structure of the evaluation reports of the MFA. Inception report will be kept concise and will not exceed 25 pages, annexes excluded.

The consultant will organise the inception meeting in Helsinki (EVA-11 can offer the meeting room).

Deliverables: Inception report and inception meeting agenda. These will be sent to EVA-11 five working days before meeting. The evaluation team is also expected to write the minutes of the meeting and send them to EVA-11 by email two (2) days after the meeting was held.

III. Phase one (meta-evaluation) and phase two (meta-analysis)

In the phase one, the meta-evaluation assess the quality of decentralised evaluation reports and ToRs and other relevant documents, including documents related to the capacity building services and MFA's evaluation management practices.

The phase one draft report will be submitted to EVA-11. This draft report should also propose a list of decentralised evaluations that are selected to the phase two and is subject to the approval of EVA-11 prior to conducting phase two (meta-analysis). The report will be kept concise and clear.

Deliverable: Draft report of the first phase (meta-evaluation)

The second phase is a meta-analysis of the effectiveness of Finland's development co-operation verified against the OECD-DAC and EU evaluation criteria and how Finnish development policy's goals have been achieved. It will be based only on those decentralised evaluation reports that are assessed as reliable and of high quality in the first part of the evaluation.

Deliverable: Draft report of the second phase (meta-analysis).

IV. Reporting

In order to compile the findings, the consultant will prepare a final draft report of phase one and two. The report will be kept clear, concise and consistent (max 40 pages + annexes). The report will contain inter alia the evaluation findings, conclusions and recommendations. They should be logical and based on evidence.

Public presentations and meetings with relevant MFA staff will be organised when the final draft report is ready. The final draft report will be subjected to a round of comments. The presentation in Helsinki may also be organized through a webcast, webinar or video conference.

Deliverables: Final draft report and a power point presentation for public presentation.

The final report will be finalised based on the discussion and comments raised in public presentation, meetings with relevant MFA staff and commenting round. The final report must include abstract and executive summary in Finnish, Swedish and English and summary matrix in Finnish and English. The consultant is responsible for the translations.

Deliverables: The final report; Individual assessment of each evaluation report; Note about quality control and Interim evidence documents.

The MFA requires access to the evaluation team's interim evidence documents, e.g. completed matrices, although it is not expected that these should be of publishable quality. All confidential information will be handled properly.

The Consultant will submit individual assessments of each report and a methodological note explaining how the quality control was addressed during the evaluation and how the capitalization of lessons learned has been addressed.

It should be noted that the draft reports of first phase and second phase, the final draft report or final report may be subjected to an external peer review of internationally recognised experts. In that case the remarks addressed by the peer reviewers will be anonymously made available to the contracted Consultant.

ANNEX 2 PEOPLE INTERVIEWED

Below is the list of people consulted on capacity building services and evaluation management cycle. Titles of positions reflect the situation that prevailed at the time of the interviews in 2014.

Name	Position	Country/Sector
Helena Airaksinen	Director	Africa Department
Eeva Alarcon	Deputy Head of Mission	Nairobi, Kenya
Roy Eriksson	Director	Latin America
Pauliina Hellman	Regional Advisor	Africa
Vesa Kaarakka	Sector Advisor	Forestry
Jussi Karakoski	Sector Advisor	Education
Gunilla Kullberg	Regional Advisor	Latin America
Riitta Oksanen	Senior Advisor	Unit for Development Evaluation
Arto Valjas	Regional Advisor	Africa

ANNEX 3 DOCUMENTS CONSULTED

Evaluation reports and documents

1. Ministry for Foreign Affairs of Finland (2015) Evaluation on Finland's Development Policy and Co-operation, 20 pages.
2. Ministry for Foreign Affairs of Finland (2015) Powerpoint Presentation of the Evaluation on Finland's Development Policy and Co-operation, Programs for a Results-based Management Point of view, 2003-2013, 24 pages.
3. Ministry for Foreign Affairs of Finland (2015) Evaluation guidelines of the Evaluation on Finland's Development Policy and Co-operation, Programs for a Results-based Management Point of view, 2003-2013, 24 pages.
4. Ministry for Foreign Affairs of Finland (2015) Evaluation manual of the Evaluation on Finland's Development Policy and Co-operation, Programs for a Results-based Management Point of view, 2003-2013, 24 pages.
5. Ministry for Foreign Affairs of Finland (2012) Comments, meta-evaluation 2012 Quality assurance, 1 page.
6. Ministry for Foreign Affairs of Finland (2012) Management response implementation report, 14 pages.
7. Ministry for Foreign Affairs of Finland (2012) Inception-report, meta-analysis of decentralized development evaluations 2010-2011, 74 pages.
8. Ministry for Foreign Affairs of Finland (2012) Management decision of implementation of the meta-evaluation, 4 pages.

MFA Manual

9. Ministry for Foreign Affairs of Finland (2013) *Evaluation guidelines request for comments*, 2 pages.
10. Ministry for Foreign Affairs of Finland (2014) *Info evaluation manual*, 2 pages.
11. Ministry for Foreign Affairs of Finland (2013) *Guidelines for Programme Design, Monitoring and Evaluation*, 69 pages.
12. Ministry for Foreign Affairs of Finland (2008) *Meta-analysis of Development Evaluations in 2007-2008 Management Response*, 2 pages.

Terms of Reference and contracts

13. Ministry for Foreign Affairs of Finland (2010) *Plan point*, 12 October 2010, 2 pages.
14. Ministry for Foreign Affairs of Finland (2012) *Plan point*, 12 October 2012, 1 page.
15. Ministry for Foreign Affairs of Finland (2013) *Plan point*, 13 May 2013, 1 page.
16. Ministry for Foreign Affairs of Finland (2014) *Plan point*, 8 January 2014, 1 page.
17. Ministry for Foreign Affairs of Finland (2014) *Plan point*, 4 February 2014, 1 page.

18. Ministry for Foreign Affairs of Finland (2014) *Plan point*, 5 February 2014, 1 page.
19. Ministry for Foreign Affairs of Finland (2013) *Terms of Reference*, 13 December 2013, 3 pages.
20. Ministry for Foreign Affairs of Finland (2014) *Terms of reference for mid-term review*, 5 pages.
21. Ministry for Foreign Affairs of Finland (2012) *Impact Evaluation training*, held in Helsinki on 27-28 September 2012, 2 pages.
22. Ministry for Foreign Affairs of Finland (2012) *Terms of Reference for the Updating of the Guideline on Development Evaluation*, 2 pages.
23. Ministry for Foreign Affairs of Finland (2012) *Terms of Reference for the Technical Assistance to the Watershed Monitoring and Evaluation (WME) Component of the Tana Beles Integrated Water Resources Development Project (TBIWRDP) in Ethiopia Mid-Term Review (MTR) mission* - January 2012, 7 pages.
24. Ministry for Foreign Affairs of Finland (2014) *Terms of Reference for the independent impact evaluation of MSIA's programme activities* - 2014, 4 pages.
25. Ministry for Foreign Affairs of Finland (2012) *Terms of Reference for the Environment and Security Initiative (ENVSEC) – Draft version* - 2012, 5 pages.
26. Ministry for Foreign Affairs of Finland (2012) *Terms of Reference for the Environment and Security Initiative (ENVSEC) – Final version* - 2012, 5 pages.
27. Ministry for Foreign Affairs of Finland (2011) *Terms of Reference for a mid-term evaluation*, 2 pages.
28. Ministry for Foreign Affairs of Finland (2011) *Mid-Term review of the Water and Sanitation Project for Small Towns (WSPST)*, 6 pages.
29. Ministry for Foreign Affairs of Finland (2011) *Terms of Reference for Appraisal*, 10 pages.
30. Ministry for Foreign Affairs of Finland (2011) *Terms of Reference for Final/Ex-post evaluation*, 10 pages.
31. Ministry for Foreign Affairs of Finland (2011) *Terms of Reference for Mid-term evaluation*, 10 pages.
32. Ministry for Foreign Affairs of Finland (2011) *Terms of Reference – programme planning*, 7 pages.

Feedback on training, assessment of learning and certificates

33. Ministry for Foreign Affairs of Finland (2012) Feedback on training, assessment of learning, Training in evaluation of development co-operation held in Helsinki on 17 April 2012, 20 pages.
34. Ministry for Foreign Affairs of Finland (2012) Feedback form on training, assessment of learning, Training in evaluation of development co-operation held in Helsinki on 21-22 May 2012, 2 pages.
35. Ministry for Foreign Affairs of Finland (2012) Feedback on training, assessment of learning, Training in evaluation of development co-operation held in Helsinki on 25 May 2012, 20 pages.
36. Ministry for Foreign Affairs of Finland (2012) Feedback on training, assessment of learning, Training in evaluation of development co-operation held in Helsinki on 27 May 2012, 24 pages.
37. Ministry for Foreign Affairs of Finland (2013) Feedback on training, assessment of learning, Training in evaluation of development co-operation held in Helsinki on 28 May 2013, 30 pages.
38. Ministry for Foreign Affairs of Finland (2014) List of the trainings in evaluation of development co-operation between 2011 and 2014, 21 pages.
39. Ministry for Foreign Affairs of Finland (2013) Certificate, Training in evaluation of development

co-operation held in Helsinki on 28 May 2013, 1 page.

40. Ministry for Foreign Affairs of Finland (2011) Feedback collection, 1 page.
41. Ministry for Foreign Affairs of Finland (2011) Feedback collection form, 2 pages.
42. Ministry for Foreign Affairs of Finland (2013) Feedback collection form, 1 page.
43. Ministry for Foreign Affairs of Finland (2014) Feedback collection form, 2 pages.
44. Ministry for Foreign Affairs of Finland (2013) Evaluation training changes feedback needs, 5 December 2013, 1 page.
45. Ministry for Foreign Affairs of Finland (2010) Evaluation training concept, 1 page.
46. Ministry for Foreign Affairs of Finland (2010) Evaluation training in MFA memo, 4 pages.
47. Ministry for Foreign Affairs of Finland (2010) Evaluation training presentation in KEPO, 4 pages.
48. Ministry for Foreign Affairs of Finland (2010) Training needs memo, 3 pages.
49. Ministry for Foreign Affairs of Finland (2010) Evaluation training in MFA memo, 3 pages.
50. Ministry for Foreign Affairs of Finland (2010) Survey for evaluation training 2010-2011, 3 pages.

Material for trainings

51. Ministry for Foreign Affairs of Finland (2012) *Powerpoint presentation for the training in evaluation of development co-operation* held in Helsinki on 21-22 May 2012, 1 page.
52. Ministry for Foreign Affairs of Finland (2012) *Welcome note for the training in evaluation of development co-operation* held in Helsinki on 21-22 May 2012, 1 page.
53. Ministry for Foreign Affairs of Finland (2012) *Draft Programme for the training in evaluation of development co-operation* held in Helsinki on 21-22 May 2012, 2 pages.
54. Ministry for Foreign Affairs of Finland (2012) *Evaluation Criteria for the training in evaluation of development co-operation* held in Helsinki on 21-22 May 2012, 3 pages.
55. Ministry for Foreign Affairs of Finland (2013) *Welcome note for the training in evaluation of development co-operation* held in Helsinki on 27-28 May 2013, 1 page.
56. Ministry for Foreign Affairs of Finland (2013) *Logframe for the training in evaluation of development co-operation* held in Helsinki on 27-28 May 2013, 1 page.
57. Ministry for Foreign Affairs of Finland (2013) *Powerpoint presentation for the training in evaluation of development co-operation* held in Helsinki on 27-28 May 2013, 1 page.
58. Ministry for Foreign Affairs of Finland (2013) *Programme for the training in evaluation of development co-operation* held in Helsinki on 27-28 May 2013, 1 page.
59. Ministry for Foreign Affairs of Finland (2014) *Invitation letter for the training in evaluation of development co-operation* held in Helsinki on 26-27 May 2014, 1 page.
60. Ministry for Foreign Affairs of Finland (2014) *Welcome note for the training in evaluation of development co-operation* held in Helsinki on 26-27 May 2014, 1 page.
61. Ministry for Foreign Affairs of Finland (2014) *Powerpoint presentation for the training in evaluation of development co-operation* held in Helsinki on 26-27 May 2014, 1 page.
62. Ministry for Foreign Affairs of Finland (2014) *Programme for the training in evaluation of development co-operation* held in Helsinki on 26-27 May 2014, 1 page.

63. Ministry for Foreign Affairs of Finland (2012) *Agenda for RealWorld Evaluation and Holistic Approach*, 1 page.
64. Ministry for Foreign Affairs of Finland (2012) *Contract between the Ministry for Foreign Affairs and Real World Evaluation on Services for Training in impact evaluation*, 1 page.
65. Ministry for Foreign Affairs of Finland (2012) *Draft Programme for the training in evaluation of development co-operation* held in Helsinki on 21-22 May 2012, 2 pages.
66. Ministry for Foreign Affairs of Finland (2012) *Draft Programme for the training in evaluation of development co-operation* held in Helsinki on 21 December 2010, 1 page.
67. Ministry for Foreign Affairs of Finland (2010) *Draft Programme for the training in evaluation of development co-operation* held in Helsinki on 20-21 June 2011, 1 page.
68. Ministry for Foreign Affairs of Finland (2012) *Guidelines evaluation internal inspection*, 4 pages.
69. Ministry for Foreign Affairs of Finland (2010) *Role of KEPO – Draft version*, 2 pages.
70. Ministry for Foreign Affairs of Finland (2010) *Checklist cross-cutting objectives*, 1 page.
71. Ministry for Foreign Affairs of Finland (2011) *Evaluation norms*, 6 pages.
72. Ministry for Foreign Affairs of Finland (2012) *Cross-cutting objectives in the development of policy programme*, 3 pages.
73. Ministry for Foreign Affairs of Finland (2010) *Guidelines on Child Rights*, 2 pages.
74. Ministry for Foreign Affairs of Finland (2008) *Guidelines on the formulation participation*, 2 pages.
75. Ministry for Foreign Affairs of Finland (2008) *Guidelines for people living with HIV and AIDS*, 3 pages.
76. Ministry for Foreign Affairs of Finland (2008) *Guidelines for the rights and inclusion of people with disabilities in development co-operation*, 3 pages.
77. Ministry for Foreign Affairs of Finland (2010) *Guidelines for the quality assurance*, 7 pages.
78. Ministry for Foreign Affairs of Finland (2010) *Guidelines for the rights of indigenous peoples*, 3 pages.
79. Ministry for Foreign Affairs of Finland (2010) *Checklist for design, implementation and review, Human rights based approach to development*, 2 pages.
80. Ministry for Foreign Affairs of Finland (2010) *Human rights based approach principles*, 2 pages.
81. Ministry for Foreign Affairs of Finland (2013) *Human rights based approach*, 8 pages.
82. Ministry for Foreign Affairs of Finland (2012) *Gender Equality Mainstreaming*, 3 pages.
83. Ministry for Foreign Affairs of Finland (2012) *Tools reducing qualities*, 3 pages.
84. Ministry for Foreign Affairs of Finland (2012) *Powerpoint presentation “Good Planning – Key to Evaluability and results” 2012 Development Evaluation Day*, 12 pages.
85. Ministry for Foreign Affairs of Finland (2012) *Memo Evaluation report group 1, 2012 Development Evaluation Day*, 1 page.

List of participants in the trainings

86. Ministry for Foreign Affairs of Finland (2011) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 21 June 2011, 1 page.

87. Ministry for Foreign Affairs of Finland (2012) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 17 April 2012, 1 page.
88. Ministry for Foreign Affairs of Finland (2012) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 21-22 May 2012, 1 page.
89. Ministry for Foreign Affairs of Finland (2013) *List of participants for the training in evaluation of development co-operation* held in Helsinki on 27-28 May 2013, 1 page.
90. Ministry for Foreign Affairs of Finland (2013) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 23-24 September 2013, 1 page.
91. Ministry for Foreign Affairs of Finland (2013) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 10-11 October 2013, 1 page.
92. Ministry for Foreign Affairs of Finland (2013) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 13 December 2013, 3 pages.
93. Ministry for Foreign Affairs of Finland (2014) *List of participants in the Training in evaluation of development co-operation* held in Helsinki on 7 January 2014, 2 pages.

Exercises and PowerPoint presentation

94. Ministry for Foreign Affairs of Finland (2011) *BKK Training Group 1 for the training on management and evaluation of development co-operation*, held on 26-27 May 2011, 1 page.
95. Ministry for Foreign Affairs of Finland (2011) *BKK Training Group 2 for the training on management and evaluation of development co-operation*, held on 26-27 May 2011, 1 page.
96. Ministry for Foreign Affairs of Finland (2014) *Exercises for the Water Supply and Sanitation Project for Small Towns (WSPST) Iyhyesti*, 1 page.
97. Ministry for Foreign Affairs of Finland (2014) *Logical Framework for the Water Supply and Sanitation Project for Small Towns (WSPST) Iyhyesti*, 1 page.
98. Ministry for Foreign Affairs of Finland (2014) *Logical Framework Matrix, Annexe A of the project document: Environmental Civil Society Centre, Zambia*, 5 pages.
99. Ministry for Foreign Affairs of Finland (2014) *Logical Framework for the Water Supply and Sanitation Project for Small Towns (WSPST) Iyhyesti*, 1 page.
100. Ministry for Foreign Affairs of Finland (2012) *Quality Assessment Grid*, 10 pages.
101. Ministry for Foreign Affairs of Finland (2010) *Evaluation Process*, 2 pages.
102. Ministry for Foreign Affairs of Finland (2012) *Results based Management Matrix*, 15 May 2012, 1 page.
103. Ministry for Foreign Affairs of Finland (2012) *Logical Framework for the support to private plantation forestry and value chains in Tanzania*, 2 pages.
104. Ministry for Foreign Affairs of Finland (2012) *Example of a risk management matrix*, 4 pages.
105. Ministry for Foreign Affairs of Finland (2012) *Clarity of analysis in the evaluation report*, 4 pages.
106. Ministry for Foreign Affairs of Finland (2011) *Powerpoint Presentation for the training in evaluation of development co-operation* held in Helsinki on 20-21 June 2011, 1 page.
107. Ministry for Foreign Affairs of Finland (2012) *Powerpoint Presentation for the training in evaluation of development co-operation* held in Helsinki on 16 February 2012, 1 page.
108. Ministry for Foreign Affairs of Finland (2012) *Powerpoint Presentation for the training in evalua-*

- tion of development co-operation* held in Helsinki on 16-17 April 2012, 1 page.
109. Ministry for Foreign Affairs of Finland (2013) *Powerpoint Presentation for the training in evaluation of development co-operation* held in Helsinki in 2013, 1 page.
 110. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation for the training in evaluation of development co-operation* held in Helsinki on February 2014, 1 page.
 111. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation for cross-cutting themes in evaluation of Finnish development co-operation*, 11 pages.
 112. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation for human rights and gender equality in development evaluation – What capacity is needed?*, 2 pages.
 113. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation for principles, standards and criteria*, 1 page.
 114. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation for Evaluation Criteria*, 1 page.
 115. Ministry for Foreign Affairs of Finland (2011) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 11 November 2011, 16 pages.
 116. Ministry for Foreign Affairs of Finland (2011) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 1 April 2011, 10 pages.
 117. Ministry for Foreign Affairs of Finland (2011) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 2 December 2011, 10 pages.
 118. Ministry for Foreign Affairs of Finland (2010) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 2 September 2010, 15 pages.
 119. Ministry for Foreign Affairs of Finland (2010) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 19 October 2010, 17 pages.
 120. Ministry for Foreign Affairs of Finland (2010) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 10 November 2010, 19 pages.
 121. Ministry for Foreign Affairs of Finland (2010) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 25 November 2010, 17 pages.
 122. Ministry for Foreign Affairs of Finland (2010) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 26 November 2010, 12 pages.
 123. Ministry for Foreign Affairs of Finland (2012) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 26 October 2012, 36 pages.
 124. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation for the training in Evaluation of Development Co-operation*, held in Helsinki on 2 February 2014, 24 pages.
 125. Ministry for Foreign Affairs of Finland (2014) *Powerpoint Presentation on Principles and best practices of PCM and LFA*, held in Helsinki on 2 February 2014, 23 pages.
 126. Ministry for Foreign Affairs of Finland (2012) *Powerpoint Presentation of the evaluation function in MFA*, 23 pages.
 127. Ministry for Foreign Affairs of Finland (2010) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 5 November 2010, 1 page.
 128. Ministry for Foreign Affairs of Finland (2010) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 26 November 2010, 1 page.

-
129. Ministry for Foreign Affairs of Finland (2011) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 4 March 2011, 1 page.
130. Ministry for Foreign Affairs of Finland (2011) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 1 April 2011, 1 page.
131. Ministry for Foreign Affairs of Finland (2011) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 11 November 2011, 1 page.
132. Ministry for Foreign Affairs of Finland (2011) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 2 December 2011, 1 page.
133. Ministry for Foreign Affairs of Finland (2012) *List of documents for the training in evaluation of development co-operation* held in Helsinki on 16 November 2012, 1 page.

ANNEX 4 METHODOLOGY

1. Introduction

1.1 Overview

The meta-evaluation was carried out between October 2014 and March 2015. It was managed by EVA-11 with the support of an evaluation Reference Group made up of representatives from MFA's regional and policy departments. (EVA-11 is responsible for managing all corporate evaluations. This study focuses on decentralised evaluations managed by the MFA departments and units that are in charge of development co-operation programmes in specific countries, regions, or international institutions.)

As noted in the Terms of Reference (TORs) in Appendix 1, the meta-evaluation had two purposes: i) to help the MFA improve the quality of decentralised evaluations, associated evaluation management practices, and evaluation capacity development services, and ii) to identify issues and lessons learned that could help the MFA improve development co-operation.

The objectives were to assess the quality of evaluation reports, the quality of evaluation TORs, Finland's Development Co-operation, and MFA evaluation management and capacity development services.

During the week of 9 March 2015, the draft evaluation report was shared and discussed with MFA senior management, presented to the public in Helsinki as well as during a webinar. Feedback from stakeholders informed the final version of the report.

The following sections describe:

- the approach, criteria and sample for the meta-evaluation (section 2)
- the types of analysis, sources of data, and rating schemes used in assessing each of the objectives (sections 3-5)
- the evaluation process (section 6) and
- limitations (section 7).

2. Evaluation Approach

2.1 Framework

The meta-evaluation was guided by a framework (Appendix III) that included the key questions as presented by the Terms of Reference.

2.2 Population

MFA provided 64 reports on decentralised evaluations of projects, programmes and other initiatives that were carried out between January 2012 and August 2014 along with their associated Terms of Reference, and in some cases Instructions to Tenderers (ITT). During the inception phase, these were screened by Universalis in collaboration with EVA-11 to eliminate reports that did not fit into the scope of the meta-evaluation (i.e. removing appraisals and evaluations of initiatives funded exclusively through the multi-lateral channel). This resulted in a population of 57 reports that became the basis for two assessments: the assessment of quality of evaluation reports and TORs, and the subsequent assessment of Finland's

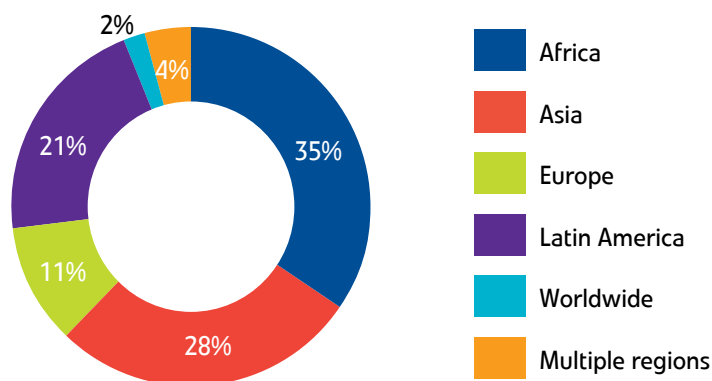
development co-operation. The list of evaluations reviewed is shown in Appendix IV and the samples for each assessment are described below.

Population for Quality Assessment

In the quality assessment, 57 reports were assessed along with their associated Terms of Reference, and in some cases ITT. These included 33 mid-term evaluations, 20 final evaluations, two ex-post evaluations and two reviews.

As shown in the chart below, the distribution of decentralised evaluations in this period varied by region. Most focused on African initiatives, followed closely by Asian initiatives. The great majority of evaluations reviewed were national (30 evaluations) or sub-national (22 evaluations) in scope. A smaller number of evaluations had a regional scope (13) or multi-country scope (9). In terms of partner countries covered, 10 of the 57 evaluations focused on Nicaragua, followed by Nepal (7). In Africa, the countries with greatest coverage were Kenya and Tanzania (with 3 and 4 evaluations respectively). In Central and Eastern Europe, 3 evaluations were conducted in Kosovo.

Evaluation Reports Reviewed per Region



Sector data for each of the decentralised evaluations was provided by EVA-11 and based on the classification of the project or programme according to the OECD/DAC sector classification system. The sectors covered by the evaluations were diverse.

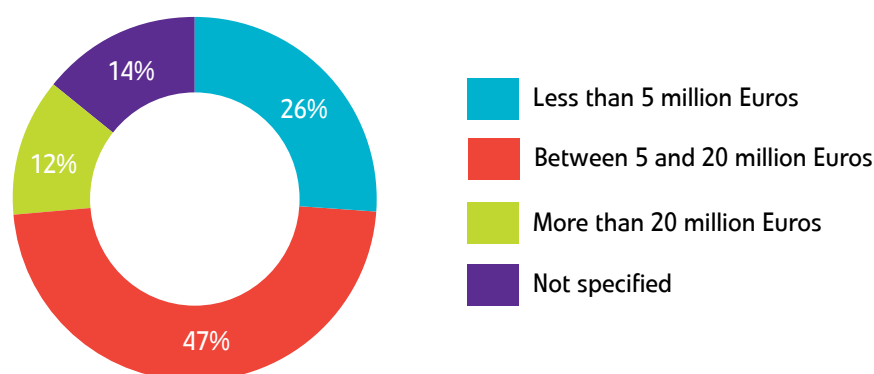
Reports Reviewed by Sector



The chart below summarises data on the size of budgets in the evaluations reviewed. Data on the project/programme budget was not provided in nine of the 57 reports. In some reports, the total budget of the project since inception was not presented, but financial data were provided on the current phases of the project. (USD figures were converted into Euros using a December 2014 exchange rate: 1 Euro = 1.239

USD on 11 December 2014). The average size of projects was approximately € 20 million. (Two outliers were removed to calculate the average budget of programmes: Mid-Term Evaluation of the Neighbourhood Investment Facility under the European Neighbourhood and Partnership Instrument; and Final Report and Joint Evaluation of Budget Support to Tanzania. The budgets were € 20 billion and € 5 billion respectively.)

Size of Projects/Programmes Reviewed (in €)



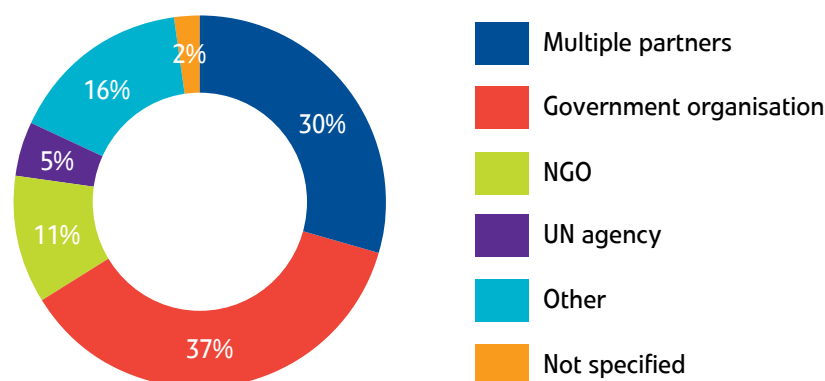
Finland's financial contribution to these projects (over the life of the project or programme) varied, ranging from € 150,000 to slightly over € 130,000,000.

Most of the evaluations had an evaluation budget between € 30,000 and € 120,000. A smaller proportion had budgets under € 30,000 (5%) and over € 120,000 (11%). It was not possible to locate the total budgets for 28% of the evaluations.

The overall duration of the projects or programmes evaluated ranged from two to over five years. A majority of evaluations were for projects that had a duration of two to five years (81%) in their current phase, while 9% were for a period of more than five years (9%) and 2% for less than two years. In five cases, the duration was not specified in the evaluation reports or the TORs.

In the 57 evaluations reviewed, 37% of projects were implemented by government organisations, 30% by multiple partners, and 33% by NGOs, UN agencies, and others. Finnish consulting groups provided services in 40% of the evaluations reviewed; in the other 60%, that information was not specified.

Types of Implementing Partners



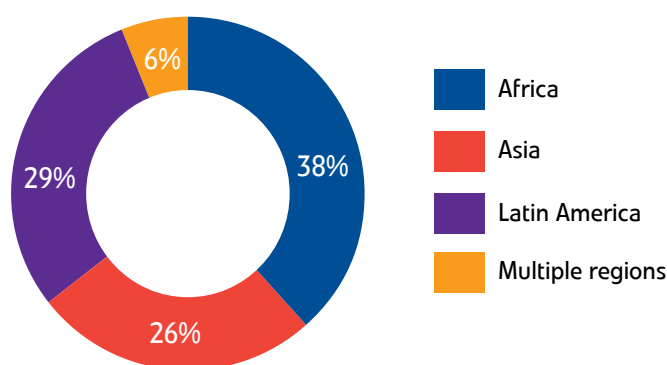
Sample for Assessment of Finland's Development Co-operation

In the assessment of Finland's development co-operation, the population of 57 decentralised evaluations was first reduced to a purposeful sample of those evaluation reports that had been rated good or better

in the quality assessment (44 evaluations) - in order to focus on reports that provided sufficient and trustworthy information to assess Finland's development co-operation. At the request of the MFA, the purposeful sample was then further refined to include only those evaluations that had been commissioned by the MFA. This resulted in a final sample of 34 evaluations (18 mid-term evaluations, 13 final evaluations, one ex-post evaluation and two reviews).

The distribution of the 34 evaluations varied by region: Most focused on African initiatives (13), followed closely by Latin American initiatives (10). This distribution is in keeping with the population in Asia and Africa. Latin America is overrepresented in the purposeful sample. Europe was not represented as there were no evaluations for that region that had been commissioned by Finland. The majority of evaluations reviewed were national (18 evaluations) or sub-national (13 evaluations) in scope. A smaller number of evaluations had a regional scope (8) or multi-country scope (6). The sectors covered in the sample were similar to those in the population of 57 reports reviewed for quality

Sample by Region

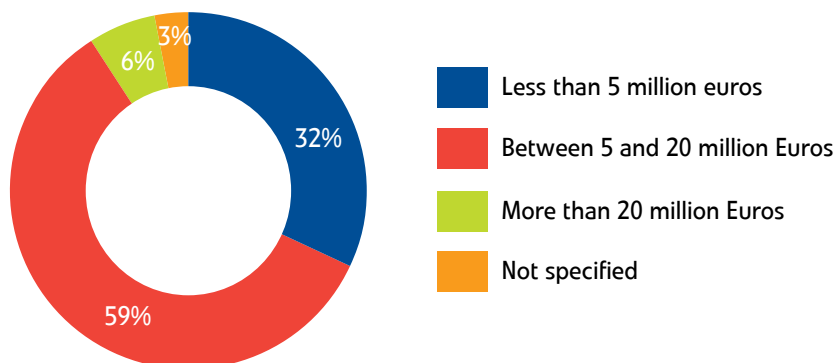


Sample by Sector

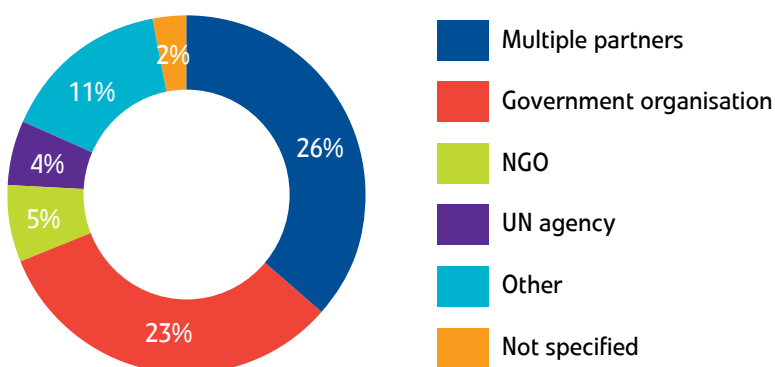


The size of projects and programmes in the sample varied, but most were in the range of €5 million to €20 million, similar to the population. Budget information was not available for one project. Most implementing partners represented in the sample were a combination of multiple partners, followed by government organisations; this was slightly different from the population where the dominant type of partner was government organisations.

Sample by project value



Sample by implementing partner type



2.3 Rating scales

The rating scales used in the assessment of quality and the assessment of development co-operation are shown in the sections below.

It should be noted that not all evaluation reports addressed every criterion in the review tool, or did not address them all clearly. In such cases the reviewers rated the criterion as ‘not addressed/not clear’. In a very few cases, this rating was also used to indicate criteria that were considered not applicable: for example, in rating a mid-term evaluation, it would not be reasonable to expect impact level results.

3. Assessment of Quality of Reports and TORs

For the quality assessment, 57 reports were reviewed against criteria drawn from the standards and other guidelines shown below.

Criteria, Standards, Guidelines Applied
OECD DAC/EU evaluation criteria
OECD DAC/EU standards for reporting
Finland’s Development Policy Programme, 2012
MFA Evaluation Manual, 2013
MFA Evaluation Guidelines, 2007
UNEG Evaluation Standards ²
Guidelines from discussions with the Reference Group

² The UNEG evaluation standards were applied to very few criteria - mostly those that dealt with how clearly and logically the TORs and evaluation reports were structured - as these were not addressed by other sources such as the OECD DAC and MFA.

Two review tools were developed in an Excel worksheet. The TORs quality review tool contained 43 statements to be rated and the report quality review tool contained 52 statements. The five-point rating scale is shown below.

Rating Scale	Definitions
Excellent	Criteria associated with the standard exceed expectations
Very good	Criteria associated with the standard are fully satisfied
Good	Criteria associated with the standard are mostly satisfied with no major gaps
Poor	Criteria associated with the standard are barely satisfied
Unacceptable	Criteria associated with the standard are totally absent
Not addressed/not clear	This element was either not addressed or not clear in the TORs or evaluation report

When appropriate, comments from a qualitative assessment were added by each reviewer in order to complement and explain the ratings. Background questions on the project/programme and the evaluation were also asked.

The sections of the TORs review tool and Evaluation Report review tool are shown below. The two review tools are presented in Annexes 7 and 8.

TORs Quality Review Tool	Evaluation Report Quality Review Tool
<ol style="list-style-type: none"> 1. Background information including programme context, programme description, results of previous evaluations, and budget 2. Description of the rationale, purpose, and objectives 3. Description of evaluation scope 4. Evaluation questions 5. Proposed methodology 6. Evaluation process and management structure 7. Description of deliverables 8. Description of resources required 9. Structure and clarity of writing 	<ol style="list-style-type: none"> 1. Description of the object, context, and intervention logic 2. Description of the purpose, objectives, and scope 3. Explanation of the evaluation methodology 4. Presentation of the findings, conclusions, lessons, and recommendations 5. Clarity, logic, and evaluation report structure

4. Assessment of Finland's Development Co-operation

Of the 57 evaluation reports that were reviewed against quality standards, 34 were selected for a review of the development effectiveness of Finland's Development Co-operation, which entailed looking at the content of the reports to make judgements about how evaluated projects/programmes/interventions were performing against a set of standards and criteria.

A review tool was developed in an Excel worksheet, based on the documents shown in the box below.

Criteria, Standards, Guidelines Applied
MFA Evaluation Manual, 2013
OECD/DAC Glossary of key terms in evaluation and results-based management
Finland's Development Policy Programme, 2012
MFA Manual for Bilateral Programme, 2012
Guidelines from discussions with the Reference Group

The tool included 38 statements to be rated. Two distinct rating scales were used.

The results terminology used in the development co-operation review tool was based on OECD/DAC terminology (e.g. outputs, outcomes, impact). In the final draft report, these terms were changed to be consistent with the results-based management terms commonly used in Finland and described in the Bilateral Manual (i.e. overall objectives, project purposes, expected results).

The main criteria of the review tool provided a basis for judgment about how well sampled projects and programmes met the OECD/DAC evaluation criteria and expectations for cross-cutting objectives. These main criteria were rated on a five-point scale that was adapted for each OECD/DAC criterion and cross-cutting objective. The rating scale for effectiveness is shown below as an example. Reviewers had additional guidance in the tool on how to assign those ratings. Reviewers were also able to select ‘not addressed/not clear’ when the evaluation report did not provide adequate information for a rating on the criteria.

Rating Scale	Definitions
Excellent	Programme exceeds objectives, or is likely to exceed the planned objectives/results
Very good	Programme fully met objectives, or is likely to fully meet the planned objectives/results
Good	Programme mostly met objectives
Poor	Programme barely met objectives
Unacceptable	Programme did not meet objectives
Not addressed/not clear	Effectiveness was not addressed in the evaluation report

The sub-criteria of the review tool were added to inform the overall rating on the main criteria, allow for more consistent ratings among raters (minimise subjectivity,) and provide a more complete and in-depth analysis. Sub-criteria were meant to add explanatory value and provide richer description. These sub-criteria were rated on a three-point rating scale: yes, partially, no. Reviewers were also able to select ‘not addressed/not clear’ when the evaluation report did not provide adequate information for a rating. When appropriate, comments from a qualitative assessment were added by the reviewers to complement and explain the ratings. Background questions on the project/programme and the evaluation were also asked. The ratings of sub-criteria were used to inform the rating of the main criteria.

The sections of the review tool are shown below. The review tool is presented in Annex 11.

Project Review Tool
1. Effectiveness: achievement of outputs, outcomes, successes, shortcomings, and facilitating factors
2. Impact: reported impacts (intended and unintended), and contribution towards reducing poverty
3. Relevance: priorities of target groups, national development policies, contributing factors, and link to Finland’s priorities for development co-operation
4. Efficiency: transforming inputs into outputs, achieving objectives on schedule, management, and contributing factors
5. Sustainability: social, financial and economic, environmental, technical, institutional, and contributing factors
6. Human Rights-Based Approach: congruency, successes, shortcomings, and sustainability
7. Cross-cutting objectives: gender equality, reduction of inequality, climate sustainability, and other emerging cross-cutting objectives
8. Risk Management: types of risks managed well and not managed well

Most of the programmes evaluated between 2012 and August 2014 were designed before the introduction of Finland’s Development Policy Programme 2012. While current approaches to human rights, gender equality, and climate sustainability were not yet fully developed at that time, they were included in

the meta-evaluation review tool because the concepts had existed for many years, albeit with various different labels.

For example, promotion of human rights has been a long standing value of the MFA and was described in previous policy. Therefore, while many projects may not have had a comprehensive approach such as HRBA, there was often data on human rights in the evaluation reports and the meta-evaluation was an opportunity to assess the extent to which HRBA was included in current projects.

5. Assessment of MFA Evaluation Capacity Building Services and Evaluation Management Cycle

The assessment of MFA evaluation management practices and evaluation capacity development services was based on consultations with nine MFA stakeholders (see box below) and a document review (see References). The documents reviewed included mostly internal memos, training materials and participant feedback on training workshops.

Categories and numbers of MFA Stakeholders consulted
EVA-11 (1)
Regional advisors (2)
Sector advisors (2)
Directors (2)
Embassy staff (1)
EVA-11 staff (1)
Total (9)

From the outset, this component was intended to be ‘light-touch’ assessment that, together with meta-evaluation findings, could inform EVA-11’s planned in-depth analysis of its evaluation management practices and capacity development services.

The meta-evaluation team developed a review framework based on evaluation questions in the Terms of Reference and refined in the Inception Report, and then developed a consultation protocol to collect data (see sections of protocol below).

Evaluation management practices and capacity development services
<p>1. Evaluation capacity building support</p> <p>Types and frequency of support provided</p> <p>Relevance and appropriateness of support provided</p> <p>Adequacy of resources provided</p> <p>Appropriateness of audiences targeted</p> <p>Institutional limitations</p> <p>Recommendations</p>
<p>2. Evaluation management cycle</p> <p>Overall cycle – clear and complete</p> <p>Strengths and areas for improvement</p> <p>Roles and responsibilities of MFA stakeholders</p> <p>Engaging MFA developing country partners</p> <p>Recommendations</p>

Subsequently, the team analysed data from consultations and documents reviewed, identified shortcomings and recommendations, and assembled examples from other countries to illustrate alternative or new options to could help the MFA improve its practices.

6. Evaluation Process

Inception and workplanning: The inception phase included a review of key MFA documents and OECD-DAC and EU evaluation guidelines and tools; and consultations in Helsinki with EVA-11, the Meta-Evaluation Reference Group, and stakeholders from regional and policy departments in the Ministry. The Inception Phase culminated with the Inception report, approved in November 2014.

Development and testing of tools: Review tools (instruments) were developed to assess the quality of TORs and evaluation reports and Finland's development co-operation. The tools were tested by the assessment team and then revised based on the results of testing and feedback from the Meta-Evaluation Reference Group. Once the review of reports began, additional adjustments were made to help streamline and clarify key concepts for the reviewers.

Data collection/review process: The reviews were conducted by a team of four reviewers (see Appendix V) who recorded ratings and qualitative comments in an Excel worksheet. Reviews were subject to a quality assurance process led by the Senior Reviewers/Co-Team Leaders (see quality control processes below).

Data analysis and reporting: In a qualitative analysis, all of the meta-evaluation teams' qualitative comments on evaluation reports and TORs (i.e. comments that the reviewers had made to justify or explain ratings) were analysed to identify common trends, themes, and patterns that were then classified according to the main review themes. The meta-evaluation team used these trends to explain the overall ratings.

The meta-evaluation team also analysed the frequency of certain factors noted in the reviewer comments. These are noted in the report when the frequency of such factors was significant or noteworthy (i.e. noted in 50-60% of reviews).

Due to differences in the methodologies and rating scales used in the 2012 and in this 2014 meta-evaluation (see limitations below), there were difficulties in direct comparison of criteria. In Section 6, the team compared findings across these studies selectively and cautiously and shared other observations that emerged from a cross-sectional review of study findings.

Quality control processes: The quality of the meta-evaluation process was ensured by three main components/procedures.

- **Quality of review tools and approaches:** The three review tools were critiqued by EVA-11 and the evaluation reference group to ensure that appropriate standards and criteria would be used to assess the evaluation TORs and reports. The tools were tested by applying them to evaluation TORs and reports and revised at least three times during the process to improve clarity and consistency in the reviewer approach.
- **Checking Inter-Rater reliability:** In any meta-evaluation or meta-analysis that requires ratings, ensuring consistency in the ratings and judgments made by individual reviewers is a challenge. The meta-evaluation team took several measures to help ensure inter-rater reliability. All team members participated in a full day training to review and apply the instrument. Regular exchanges and briefings were held in order to identify special cases. Two inter-reliability "checks" were conducted in which the reviewers and the Co-Team Leaders reviewed the same report, shared the results, and subsequently developed consensus ratings. In addition, half of the reviews of TORs and evaluation reports were subject to a senior review by the Co-Team Leaders or another external reviewer. All development co-operation reviews were subject to senior peer reviews. These senior reviews provided the opportunity to discuss and challenge ratings and help to ensure consistency in providing judgments and ratings.

- **Reviewing draft meta-evaluation products against standards:** The draft reports and review instruments were critiqued by an external reviewer, Ms. Heather Buchanan, using the same “grid” that the meta-evaluation team used to judge the quality of evaluation reports. Her feedback has been incorporated in this draft report.

7. Limitations

The methodology for the meta-evaluation had certain limitations.

- The **subjective nature of the review process** is an inherent limitation: The primary review method was document review and analysis, which was used by individual reviewers to judge the quality of the evaluation reports and TORs. The reliance on solely qualitative methods has both strengths and inherent limitations. The strengths are the in-depth analysis of nuances and circumstantial information. The limitation is the lack of hard data commonly expected from evaluation work and the acknowledged limits of subjective reviews. To the extent possible, and in the confines of the time available, the meta-evaluation team tried to offset the subjective nature of the process through quality assurance mechanisms described above.
- The scope of the meta-evaluation did not include a review of evaluation **Inception Reports** or any other project documents which could have helped to explain or justify certain choices, approaches or foci in the evaluation methodologies and reports (e.g. those that differed from what was originally stipulated in the TORs, expected criteria or traditional practice). The scope also did not allow for **consultations with the MFA officers who commissioned evaluations or the evaluators who conducted the studies**, which could have helped validate judgements and/ or address information gaps.
- The 2012 and 2014 **meta-evaluations used different methodologies and different rating scales**. While the TORs for the 2014 assignment requested that the results of this meta-evaluation be compared to the 2012 Meta-Evaluation to identify trends, patterns and changes, it was agreed with MFA during the Inception Phase that the previous meta-evaluations should not be used as baseline for the 2014 meta-evaluation because of differences in methodology. Comparisons made in this document are thus to be treated with caution.
- In reviewing the effectiveness of Finland’s development co-operation, **the meta-evaluation reviewers relied on information contained in the evaluation reports**. Thus, if those reports either understated or overstated the effectiveness of Finland’s development co-operation, this bias was mirrored in the meta-evaluation.
- **Coverage of development effectiveness criteria in evaluation reports** was sometimes an issue. The areas most frequently not addressed were risk management, HRBA, and cross-cutting objectives. While this meant it was not possible to draw clear conclusions in these areas, the data collected could serve as a baseline for future meta-evaluations. Due to the **small number of reports in the development co-operation sample** (34 reports), it was not possible to analyse data by sub-unit or to develop conclusions on how MFA is doing in specific sectors, countries or regions.
- The sample of good quality reports used for the assessment of Finland’s Development Co-operation included **only those that were commissioned by the MFA**. Five of the reports excluded received an overall quality rating of Very good (4 reports) or Excellent (1 report).
- Including both **mid-term and final evaluation reports** in the review of development effectiveness may not be appropriate and this approach should be reassessed for future meta-evaluations. Final evaluations are more likely than mid-term evaluations to demonstrate tangible results in terms of overall objectives and sustainability.

ANNEX 5 META-EVALUATION MATRIX

Main Questions	Sub-Questions	Proposed Indicators	Sources of Data	Method
1. Phase 1: Meta-Evaluation				
1. To what extent are decentralised evaluation reports (2012-8/2014) and their TORs based on OECD-DAC and EU evaluation principles, standards and criteria?	<p>1.1. Are the reports well written, can they be easily understood and are they user-friendly?</p> <p>1.2. What are the main challenges and strengths of the decentralised evaluation reports and TORs applying OECD-DAC and EU evaluation principles, standards and criteria?</p> <p>1.3. What are the main reasons hindering the quality of decentralised evaluation TORs and reports?</p> <p>1.4. Is there any difference in the quality of the evaluation reports and TORs published after the 2013 MFA guidance? In what areas?</p> <p>1.5. Are there any differences in the trends identified in the 2012 meta-evaluation (taking into account differences in methodology)?</p>	<p>% of evaluation reports that are rated as good or higher overall based on whether they meet OECD-DAC and EU evaluation principles, standards and criteria</p> <p>Types and frequencies of identified positive and negative trends in the quality of MFA decentralised evaluations</p> <p>Comparison of results and trends of 2012-14 results with those identified in previous 2012 MFA meta-evaluation if/where relevant</p>	<p>Documents</p> <p>Population of MFA evaluation reports 2012-2014</p> <p>Previous MFA meta-evaluations and MFA management responses</p> <p>MFA evaluation manual, policies and guidelines</p> <p>OECD-DAC and EU evaluation principles, standards and criteria</p> <p>The evaluation policies, guidelines, practices and meta-evaluations of other selected donors</p>	<p>Application of Report Review Tool</p> <p>Application of TOR and Report Review Tools</p> <p>Review of MFA and other donors' documents</p> <p>Consultations with MFA stakeholders</p>

Main Questions	Sub-Questions	Proposed Indicators	Sources of Data	Method
2. Phase 2: Meta-Analysis				
2. Based on the selected evaluations, how effective has Finland's development co-operation been in applying the OECD-DAC and EU evaluation criteria?	<p>2.1. Has the MFA done right things in a right way and did it change what it was supposed to change? What have been the unintended results?</p> <p>2.2. To what extent has risk management improved in comparison with previous meta-evaluation findings?</p> <p>2.3. How has the human rights-based approach been applied as implied in the Finnish development policy and guidelines?</p> <p>2.4 To what extent have the crosscutting objectives of development policy been implemented as implied in Finland's development policy and guidelines?</p> <p>2.5 What types of capacity development services are provided by EVA-11 and how are they viewed by consulted stakeholders?</p> <p>2.6 How have other Nordic+ countries addressed gaps in quality of evaluation through capacity development materials?</p> <p>2.7 What is the MFA cycle for managing evaluations, and what do consulted stakeholders identify as its strengths and areas for improvement?</p> <p>2.8 What are the recommendations for evaluation capacity development services and management practices?</p>	<p>% of selected decentralised projects that fully realised, partially realised or did not at all realise planned results (effectiveness)</p> <p>% of selected decentralised projects that are considered highly relevant, somewhat relevant, or not at all relevant</p> <p>% of selected decentralised projects whose results are highly likely, somewhat likely or not all likely to be sustained (sustainability)</p> <p>% of selected decentralised projects that are rated as highly efficient, partially efficient or not at all efficient (efficiency)</p> <p>% of selected decentralised projects that had intended or unintended effects and impacts (impacts)</p> <p>Types of variations of identified positive and negative trends in the performance and lessons learned of MFA decentralised projects with particular attention to risk management, human rights based approaches and cross-cutting objective guidelines</p> <p>Types of strengths, areas for improvement in MFA evaluation management practices and capacity building services</p> <p>Comparison of results and trends of 2012-14 results with those identified in previous 2012 MFA meta-evaluation if/where relevant</p>	<p>Policy documents</p> <p>Sub-sample of high quality MFA evaluation reports 2012-2014</p> <p>Previous MFA meta-evaluations and MFA management responses</p> <p>MFA evaluation manual, policies and guidelines</p> <p>Documents from Nordic + countries available online</p> <p>Consultations with Nordic + countries (if required)</p>	<p>Content review of documents</p> <p>Application of Review Tool for Phase 2 to the sub-sample</p> <p>Consultations with MFA</p> <p>Consultation with MFA</p>

ANNEX 6 LIST OF REPORTS REVIEWED: SAMPLE AND SUB-SAMPLE

	Title of the evaluation report	OECD/DAC Sector	Year	Region	Type of Evaluation	Evaluation Company	Included in the Assessment of Dev. Coop.
1	Mid-Term Review of the SADC Renewable Energy Support Programme	Energy generation and supply	2012	Africa	Mid-Term Evaluation	Camco	no (1)
2	Mid-term Evaluation of the Basel Convention Regional Centre Support Programme	Water and sanitation	2013	Africa	Mid-Term Evaluation	Orgut & Gaia	yes
3	Building Trade Capacity through Customs Modernization in the East and Southern Africa Region. Mid-Term Evaluation Report.	Business and other services	2014	Africa	Mid-Term Evaluation	EPRD	yes
4	Mid-Term Review: Regional Biodiversity Programme for the Amazon Region of the Andean Countries (BioCAN)	Multi Sector/ Cross-cutting	2012	Latin America	Mid-Term Evaluation	Indufor	yes
5	Draft Final Report Mid-Term Evaluation of African Peacebuilding Coordination Programme Phase II	Government and civil society	2013	Africa	Mid-Term Evaluation	COWI & Media Progress	yes
6	Acquisition of a Fisheries and Marine Research Vessel for the Republic of Namibia - Final Review	Fishing	2014	Africa	Final Evaluation	Niras	no (2)
7	Mid-Term Evaluation of the Wider Europe Initiative	Government and civil society	2012	Europe	Mid-Term Evaluation	FCG Finnish Consulting Group	no (3)
8	Technical Assistance for Leasehold Forestry and Livestock Development Programme	Forestry	2014	Asia	Final Evaluation	Niras	yes
9	Tana-Beles Monitoring and Evaluation Project: Mid-Term Review 2012	Not specified	2012	Africa	Mid-Term Evaluation	Not specified	no (4)
10	UNDP Nepal: Evaluation of Strengthening the Capacity of the National Human Rights Commission (SCNHR) Project	Government and civil society	2014	Asia	Final Evaluation	Langan & H. Phuyal	no (5)
11	Evaluation of the Finnpartnership Programme	Business and other services	2012	World-wide	Final Evaluation	KPMG, Ramboll	no (6)
12	Alianza en Energía y Ambiente con Centroamérica - Sistematización de Lecciones Aprendidas por Tipo de Tecnología y Evaluación de Resultados e Impactos de Proyectos Piloto y Estudios Finalizados	Energy generation and supply	2013	Latin America	Ex-post Evaluation	CCAD	no (7)

	Title of the evaluation report	OECD/DAC Sector	Year	Region	Type of Evaluation	Evaluation Company	Included in the Assessment of Dev. Coop.
13	Mid-term Evaluation of the Southern Africa Innovation Support Programme (SAIS) funded by the Ministry for Foreign Affairs of Finland	Industry	2014	Africa	Mid-Term Evaluation	EPRD	yes
14	Framework agreement for the support to the concessional credit scheme final report of the rural electrification project in Ghana	Energy generation and supply	2012	Africa	Final Evaluation	Niras	no (8)
15	Midterm Evaluation of Equal Before the Law: Access to Justice in Central Asia	Government and civil society	2013	Asia	Mid-Term Evaluation	Channel Research	yes
16	Independent Final Evaluation - From the Crisis towards Decent and Safe Jobs in Southern Caucasus and Central Asia	Other social infrastructure and services	2014	Europe, Asia	Final Evaluation	A. Kuzmin	yes
17	Kenya-Finland MDGs Programme Phase II	Government and civil society	2012	Africa	Mid-Term Evaluation	FCG Finnish Consulting Group	no (9)
18	The Final Review of the Zinjiang Fruit and Vegetable Cold Store, China - Final Report	Agriculture	2013	Asia	Final Evaluation	Niras	no (10)
19	Mid-term Review of the Environment Management Support Programme (EMSP) 2010-2014	Multi Sector/ Cross-cutting	2012	Asia	Mid-Term Evaluation	Hatfield Consultants	yes
20	The Final Review of the Kuitun and Shihezi Areas Farming Development Project in Xinjiang - Draft Report	Agriculture	2013	Asia	Final Evaluation	Niras	no (11)
21	Strengthening National Geographic services in Lao PDR (SNGS), Lao PDR and Ministry for Foreign affairs of Finland, Mid-Term Review, Final Report	Government and civil society	2012	Asia	Mid-Term Evaluation	Econet	no (12)
22	Mid-Term Evaluation of "Miti Mingi Maisha Bora" - Support to Forest Sector Reform in Kenya - Final Report	Forestry	2013	Africa	Mid-Term Evaluation	Impact	yes
23	Mid-Term Reviews of the Energy and Environment Partnership Programme with Southern and East Africa and with Mekong	Energy generation and supply	2012	Africa, Asia	Mid-Term Evaluation	Not specified	yes
24	Assessment of the Past Support to the Harmonised Rural Window of the WSTF (2010-2013) and Appraisal of the Co-Funding Proposal (2014-2016)	Water and sanitation	2013	Africa	Final Evaluation	Niras	yes

25	Mekong Water Dialogues, Phase II/Mid-Term Review	Water and sanitation	2012	Asia	Mid-Term Evaluation	Econet & Finnmap	yes
26	Midterm Review of the "Diaspora Engagement in Economic Development of Kosovo" (DEED) Project	Business and other services	2013	Europe	Mid-Term Evaluation	FCG Finnish Consulting Group	no (13)
27	Mid-Term Review of Establishment of a Regional Flood Information System (RFIS) in the Hindu-Kush Himalaya (RFIS; aka HKH-HYCOS)	Multi Sector/ Cross-cutting	2012	Asia	Mid-Term Evaluation	Impact	yes
28	Midterm Evaluation of Project CGP/KOS/005/ FIN "Support to Implementation of Forest Policy and Strategy in Kosovo"	Not specified	2013	Europe	Mid-Term Evaluation	FAO	no (14)
29	Mid-term evaluation of Forest Resource Assessment in Nepal 2009-2014 and Evaluation of Improving Research Capacity of Forest Resource Information Technology in Nepal 2010-2012	Forestry	2012	Asia	Mid-Term Evaluation	Niras	yes
30	Support to Addressing Gender Based Violence in Kosovo through Strengthening the Implementation of the Kosovo Law, National Action Plan and Strategy Against Domestic Violence 2010-2013	Government and civil society	2013	Europe	Mid-Term Evaluation	FCG Finnish Consulting Group	no (15)
31	Supporting Nepal's Constitution Building and Implementation: Realizing the Promise of Social Inclusion and Accountability: A Mid-Term Review	Government and civil society	2012	Asia	Mid-Term Evaluation	International IDEA	no (16)
32	Mid-term Review Report - Support to Minority Communities During and After the Decentralisation Process in Kosovo, Implemented by ECMI Kosovo	Government and civil society	2013	Europe	Mid-Term Evaluation	FCG Finnish Consulting Group	no (17)
33	Mid-term Evaluation of the School Sector Reform Program	Education	2012	Asia	Mid-Term Evaluation	Individual consultants	no (18)
34	International Law Project (ILP) Impact Evaluation, Report taking into consideration comments by ILP and UNDP	Government and civil society	2013	Asia	Final Evaluation	Appraisal Consulting	yes
35	Institutional Support to Civil Society Organizations in Nicaragua	Government and civil society	2012	Latin America	Final Evaluation	Impact	yes

	Title of the evaluation report	OECD/DAC Sector	Year	Region	Type of Evaluation	Evaluation Company	Included in the Assessment of Dev. Coop.
36	ForInfo: Report of the Mid Term Review	Forestry	2013	Asia	Mid-Term Evaluation	Cor Veer	no (19)
37	Life in Democracy Programme in Nicaragua	Government and civil society	2012	Latin America	Final Evaluation	Impact	yes
38	Midterm Evaluation to Support to Rural Development in Zambezia Province, Second Phase (PRODEZA II): Final Report	Multi Sector/ Cross-cutting	2013	Africa	Mid-Term Evaluation	Niras	yes
39	Final Evaluation of Municipal ICT Programme in Nicaragua	Government and civil society	2012	Latin America	Final Evaluation	FCG Finnish Consulting Group	yes
40	Midterm evaluation of Co-operation in Science, Technology, and Innovation between Finland and Mozambique (STIFIMO) Project	Communications	2013	Africa	Mid-Term Evaluation	EPRD	yes
41	PRORURAL Inuyente (2010-2014) Evaluación de Medio Término (EMT): Enfoque por Resultado y Sectorial	Agriculture	2012	Latin America	Mid-Term Evaluation	Individual consultants	no (20)
42	Mid-Term Review of the Rural Village Water Resources Management Project (RVWRMP II) in Nepal 2010-2015 Final Report	Water and sanitation	2013	Asia	Mid-Term Evaluation	Ramboll	yes
43	Midterm final Evaluation of Finland's Bilateral Programmes in Health Sector in Nicaragua	Health	2013	Latin America	Final Evaluation	FCG Finnish Consulting Group	yes
44	Final Review of the Restoration/ Rehabilitation of Waterways Project in the Philippines Phase I	Water and sanitation	2012	Asia	Final Evaluation	FCG Finnish Consulting Group	yes
45	Evaluacion of FONSALUD and Institutional Plans of the Ministry of Health Nicaragua	Health	2013	Latin America	Final Evaluation	FCG Finnish Consulting Group	yes
46	Midterm Evaluation of BioFISA	Communications	2012	Africa	Mid-Term Evaluation	FCG Finnish Consulting Group	yes
47	Final Evaluation of Nicaragua-Finland Agrobiotechnology Programme NIFAPRO	Agriculture	2013	Latin America	Final Evaluation	Impact	yes

48	Review of Finnish Support to The Institute of African Leadership For Sustainable Development (Uongozi Institute) Final Report	Multi Sector/ Cross-cutting	2012	Africa	Review	Impact	yes
49	Evaluación de la Cooperación entre Nicaragua y Finlandia en el Sector Rural - Informe Final de Evaluación de: Programa Integral de Desarrollo Rural (PIDR) y Programa de Desarrollo Ganadero (PRODEGA)	Agriculture	2013	Latin America	Ex-post Evaluation	Impact	yes
50	External Review of Core Support under Joint Financial Agreement to Zambia National Farmers Union	Agriculture	2012	Africa	Mid-Term Evaluation	Indevelop AB	no (21)
51	Enhancing Small Enterprise Growth of Nicaragua through the Development of Existing Value Chains PROPEMCE	Trade policy	2013	Latin America	Final Evaluation	Impact	yes
52	Zanzibar Sustainable Management of Land and Environment, Phase II (SMOLE II) Mid-term Review Report	Multi Sector/ Cross-cutting	2012	Africa	Mid-Term Evaluation	Land Equity International	yes
53	Evaluation of "Promotion of Sexual and Reproductive Health with Emphasis on Adolescent and Youth in Nicaragua" (VOZJOVEN) Programme	Health	2013	Latin America	Final Evaluation	FCG Finnish Consulting Group	yes
54	Mid-Term Evaluation of the Neighbourhood Investment Facility under the European Neighbourhood and Partnership Instrument (ENPI) 2007-2013:Final Report	Business and other services	2013	Europe	Mid-Term Evaluation	DRN, ECDPM, ECORYS, PARTICIP	no (22)
55	Joint Evaluation of Budget Support to Tanzania	Commodity Aid and General Programme Assistance	2013	Africa	Final Evaluation	ADE, ITAD, COWI	no (23)
56	Midterm Review of the Information Society and ICT Sector Development Project (TANZICT)	Communications	2013	Africa	Review	FCG Finnish Consulting Group	yes
57	Midterm Evaluation of the Programme for Luapula Agricultural and Rural Development (PLARD) II	Agriculture	2013	Africa	Mid-Term Evaluation	Impact	yes

ANNEX 7 QUALITY ASSESSMENT: REVIEW TOOL FOR EVALUATION REPORTS

MFA of Finland Meta Evaluation 2012-2014

Information of the Evaluation Report
Title of the evaluation report
Universalialia Sequence number
Date of review by Universalialia (dd/mm/yyyy)
Name of Universalialia reviewer (2)
Year of the evaluation report
Country or region
Type of report (Based on report content, not the title)
TORs are available
TORs are annexed to the report
ITT is available
What is the total budget of programme/project?
What sector does the evaluation focus on?
Is the purpose of the evaluation clearly defined?
What is the duration of the programme/project? [Current phase]
When was the programme/project designed? Specify the year. [Current phase]
What type of organisation implements the programme? Who is the partner?
Are support services provided to the partner by a consultant company?
What is the geographic scope of the project?

The following section illustrates standards that should be found in Evaluation Reports. Where possible, sources for these standards are identified on the right hand side. Each standard is presented in a yellow cell and the components of these standards are numbered below. The components described under the headline standards are formulated as questions and are designed to illustrate what each component should ideally include. **Sections shaded in blue are for information gathering purposes only, they should not affect overall ratings.**

Rating Scale	Definitions
Excellent	Criteria associated with the standard exceed expectations
Very good	Criteria associated with the standard are fully satisfied
Good	Criteria associated with the standard are mostly satisfied with no major gaps
Poor	Criteria associated with the standards are barely satisfied
Unacceptable	Criteria associated with the standard are totally absent
Not addressed/not clear	This element was either not addressed or not clear in the report

Review of Evaluation Reports

Questions	Rating	Comments	Sources of questions
Please provide a rating for headline standards and their numbered components. Note that the overall judgement on the headline standard is based on “expert judgment” and is not a mathematical sum of the ratings provided on the components. The sub-questions are provided in order to ensure that the reviewer considered all the relevant dimensions.		Please provide detailed comments against each headline standard and numbered component	
1. The object of the evaluation* and the context of the development intervention are adequately described in the evaluation report. (*Text below refers to evaluations, but is also applicable to reviews)			From MFA 2013 Evaluation Manual/ OECD/DAC
1.1 The context of the development intervention (programme, project, etc) is adequately described.			From MFA 2013 Evaluation Manual/ OECD/DAC
Is the development context, including socio-economic, political and cultural factors described?			
Does the evaluation report describe the institutional context and stakeholders involvement?			
Does the evaluation report describe Finland’s development policy, relevant sector, thematic and geographic priorities, including cross-cutting objectives/themes/ issues?			From MFA 2013 Evaluation Manual
Does the evaluation report discuss the influence of the context on programme performance?			From MFA Manual, report quality checklist, p.70
1.2 There is a clear description of the intervention to be evaluated (object of the evaluation).			From MFA 2013 Evaluation Manual/ OECD/DAC
Are the components of the intervention described? Are the costs or level of investment described?			
Are key stakeholders clearly identified? These include implementing agency(ies), development partners, beneficiaries, etc.			
Is the implementation status described? This includes the phase of implementation and significant changes that have happened to plans, strategies, performance frameworks, etc that have occurred - including the implications of these changes			
1.3 The evaluation report adequately describes and assesses the intervention logic or theory.			From MFA 2013 Evaluation Manual/ OECD/DAC
Are underlying assumptions and factors affecting the success of the intervention described?			
Are the results chain or other logic models clearly described and explained?			
Is the validity of the logic examined and shortcomings to logic or theory identified?			

Review of Evaluation Reports

Questions	Rating	Comments	Sources of questions
2. The evaluation report clearly describes the objective, purpose and scope of the evaluation.			From MFA 2013 Evaluation Manual/ OECD/DAC
2.1 There is a clear description of the purpose and objectives of the evaluation.			From MFA 2013 Evaluation Manual/ OECD/DAC
Do the objectives explain what the evaluation is seeking to achieve?			
Does the evaluation explain why the evaluation is being undertaken at this particular point in time?			
2.2 There is a clear description of the scope of the evaluation.			From MFA 2013 Evaluation Manual/ OECD/DAC
Does the scope describe and justify what the evaluation will and will not cover?			
Is there a description of the geographic and temporal scope?			
3. The evaluation report adequately describes and explains the evaluation methodology and its application.			From MFA Centralised Evaluation Tool
3.1 The validity and reliability of information sources are adequately addressed.			From MFA 2013 Evaluation Manual/ OECD/DAC
Does the evaluation report describe the sources of information used, including documents, respondents, administrative data, literature, etc. in sufficient detail so that the adequacy of the information can be assessed?			
Is the data collection approach in line with the purpose of the evaluation?			
Does the evaluation cross-validate the information sources and critically assess the validity of available data?			
3.2 The evaluation report clearly explains why certain methods/tools were selected.			From MFA 2013 Evaluation Manual/ OECD/DAC
Does the evaluation report describe why specific case studies or samples were selected?			
Does the evaluation report describe the data collection tools that were used, including surveys, questionnaires, field observations, etc.			
Are there well-designed indicators selected in order to provide evidence about the project/programme?			
3.3 The evaluation report adequately addresses limitations.			From MFA 2013 Evaluation Manual/ OECD/DAC
For instance, does the report address limitations regarding the representativeness of the sample identified?			
Does the report acknowledge constraints encountered and how these have affected the evaluation?			
Are methodological limitations acceptable for the task in hand?			

Review of Evaluation Reports			
Questions	Rating	Comments	Sources of questions
3.4 The evaluation report clearly outlines all the OECD/DAC criteria that are justified given the purpose of the evaluation.			From MFA 2013 Evaluation Manual/ OECD/DAC
Are the evaluation questions clearly listed in the report?			
If some criteria/questions are left out, is there an explanation for this?			
3.5 Does the evaluation approach integrate cross-cutting objectives/themes/issues?			From MFA 2013 Evaluation Manual
Is gender equality addressed?			From MFA 2013 Evaluation Manual
Are human rights addressed?			From MFA 2013 Evaluation Manual
Is the reduction of inequalities addressed?			From MFA 2013 Evaluation Manual
Is climate sustainability addressed?			
3.6 Is risk management addressed in the evaluation framework?			
4. The evaluation reports presents findings, conclusions, recommendations and lessons separately with a clear and logical distinction between them.			From MFA 2013 Evaluation Manual/ OECD/DAC
4.1 Findings are clearly presented and based on the objective use of the reported evidence.			
Do the findings address all of the evaluation's stated criteria and questions? Are explanations provided where this is not possible?			
Do findings logically flow and are they justified by data, analysis, and interpretation through carefully described logical reasoning?			
Are interpretation and extrapolations justified and supported by sound arguments?			From MFA Centralised Evaluation Tool
4.2 Conclusions are clearly substantiated by findings and analysis.			From MFA Centralised Evaluation Tool
Are conclusions substantiated by findings and analysis?			
Are conclusions free of personal or partisan considerations?			
Do the conclusions represent actual insights into important issues that add value to the findings?			From UNEG evaluation standards
Do the conclusions provide insights on criteria or main issues (foci) the evaluation was supposed to explore?			From UNEG evaluation standards
4.3 Recommendations are well-grounded in the evidence and conclusions reported.			From MFA Centralised Evaluation Tool
Do recommendations flow logically from the conclusions?			

Review of Evaluation Reports

Questions	Rating	Comments	Sources of questions
Are recommendations clearly stated and prioritised?			
Does each recommendation identify the target group for action?			
Are recommendation realistic and appropriate in the context of the evaluation?			
4.4 Are lessons learned correctly identified?			From UNEG Evaluation standards
Are lessons learned generalised to indicate what wider relevance they may have?			
4.5 The evaluation report as a whole holds together in a clear and coherent way.			From MFA 2013 Evaluation Manual
Does the evaluation report presents findings, conclusions, recommendations and lessons separately and with a clear logical distinction between them?			
4.6 The report adequately addresses all the OECD/DAC criteria.			
Is the relevance of the intervention adequately addressed?			
Is the efficiency of the intervention adequately addressed?			
Is the effectiveness of the intervention adequately addressed?			
Is the impact of the intervention adequately addressed?			
Is the sustainability of the intervention adequately addressed?			
4.7 The report adequately addresses the cross-cutting objectives/themes/issues.			
Is gender equality adequately addressed?			
Are human rights adequately addressed?			
Is the reduction of inequalities adequately addressed?			
Is climate sustainability adequately addressed?			
4.8 Was risk management adequately addressed?			
Types of risks to the project or programme may include political risks, organizational risks, stakeholder risks, etc.			
5. The report is well structured, logical and clear.			UNEG evaluation standards
5.1 The evaluation report contains a clear, concise and representative executive summary of the report.			From MFA Centralised Evaluation Tool
Does the executive summary summarise the main findings, conclusions, recommendations in a summary table?			
Does the executive summary present overall lessons learned?			
Can the executive summary stand alone?			

Review of Evaluation Reports

Questions	Rating	Comments	Sources of questions
5.2 The evaluation report is well structured and adapted to its various audiences.			From MFA Centralised Evaluation Tool
Are tables and graphs used to enhance the readability of the report?			
Are specialised concepts clearly defined? Is there a list of acronyms?			
Are the sections of the report presented in a logical order?			
Is there balance between the lengths of sections dealing with the most important issues and the length of sections of minor importance?			
5.3 Annexes contain all the appropriate elements.			From MFA Centralised Evaluation Tool
Do annexes include ToRs; List of interviewees and site visits; List of documentary evidence; Details on methodology; Data collection instruments; Information about the evaluators; Copy of the evaluation matrix; Copy of the Results chain; etc.?			
Is the length annexes well balanced?			
6. Please provide an overall rating of the evaluation report .			
6.1 Would you recommend including this report in Phase 2 and why?			
6.2 Were the resources allocated sufficient to carry out the evaluation?			
6.3 Are there any points that should be pursued in terms of evaluation capacity building?			
6.4 Are there points that should be pursued in terms of how MFA manages the evaluation cycle? (insights that emerged through reading these TORs/evaluation report)			

ANNEX 8 QUALITY ASSESSMENT: REVIEW TOOL FOR TORS

MFA of Finland Meta Evaluation 2012-2014

Background Information on TOR
Title of the evaluation in TOR
Universalialia Sequence number
Date of review by Universalialia (dd/mm/yyyy)
of Universalialia reviewer
Year on TOR or ITT
Country or region
Type of report requested by TOR
ITT is available
Budget for the evaluation is available in the TOR or ITT
What is the total budget of the evaluation?
What sector does the evaluation focus on?
Is the purpose of the evaluation clearly defined?
What is the duration of the programme/project?
When was the programme/project designed? Specify the year if available
What type of organisation implements the programme? Who is the partner?
Are support services provided to the partner by a consultant company? (e.g. was a Finnish consulting firm hired to provide TA?)
What is the geographic scope of the project?

The following section illustrates standards that should be found in Terms of Reference (TOR). Where possible, sources for these standards are identified on the right hand side. Each standard is presented in a yellow cell and the components of these standards are numbered below. The components described under the headline standards are formulated as questions and are designed to illustrate what each component should ideally include. **Sections shaded in blue are for information gathering purposes only, they should not affect overall ratings.**

Rating Scale	Definitions
Excellent	Criteria associated with the standard exceed expectations
Very good	Criteria associated with the standard are fully satisfied
Good	Criteria associated with the standard are mostly satisfied with no major gaps
Poor	Criteria associated with the standards are barely satisfied
Unacceptable	Criteria associated with the standard are totally absent
Not addressed/not clear	This element is either not addressed or not clear in the TOR

Review of Terms of Reference

Standards	Rating	Comments	Sources of questions
Please provide a rating for headline standards and their numbered components. Note that the overall judgement on the headline standard is based on “expert judgment” and is not a mathematical sum of the ratings provided on the components. The sub-questions are provided in order to ensure that the reviewer considered all the relevant dimensions.		Please provide detailed comments against each headline standard and numbered component	
1. There is sufficient background information to the evaluation* provided in the TOR or ITT. (*Text below refers to evaluations, but is also applicable to reviews)			From MFA 2013 Evaluation Manual
1.1 The programme context (policy, country, regional, global, thematic context) is sufficiently described.			From MFA 2013 Evaluation Manual
Is the broader context of the programme described, including development objectives of partner countries?			
Is Finland’s development policy and its linkages to the programme being evaluated described?			
1.2 There is a clear description of the programme to be evaluated.			From MFA 2013 Evaluation Manual
Are programme objectives, implementing strategies, resources for implementation and intervention logic covered?			
Are the issues related to the promotion of human rights, gender equality, reduction of inequalities and promotion of climate change included in the background information provided?			
Are stakeholders involved and their roles described, including beneficiaries and institutions?			
Is there a description of the history of the programme, including how results and targeted outcomes (project purposes) have changed over time?			From Independent Evaluation Group Guide - “Writing Terms of Reference for an Evaluation: A How-to Guide”
1.3 The results of previous evaluations are presented.			From MFA 2013 Evaluation Manual
This section should describe what is already known from previous evaluations and what this evaluation will add.			
2. The rationale, purpose and objectives of the evaluation are clearly described in the TOR or ITT.			From MFA 2013 Evaluation Manual/ OECD/DAC
2.1 The rationale and purpose of the evaluation are stated clearly in the TOR.			
Do the TOR describe why the evaluation was undertaken at this particular time and for whom?			
Is the use of results of the evaluation adequately described? (e.g. learning and accountability purposes)			

Review of Terms of Reference

Standards	Rating	Comments	Sources of questions
Is it clear from the ToR what type of a study this is (review, mid-term, final, ex-post)?			From UNEG Standards
2.2 The specific objectives of the evaluation are clearly stated.			From MFA 2013 Evaluation Manual
Are the issues, analysis and recommendations the evaluation will focus on sufficiently described?			
3. There is an appropriate and sufficiently detailed description of the scope of the evaluation in the TOR or ITT.			From MFA 2013 Evaluation Manual/ OECD/DAC
3.1 The scope of the evaluation clearly describes the time span the evaluation covers, stakeholder groups involved, and geographical areas covered.			From MFA 2013 Evaluation Manual
Do the TOR clearly define what is included/excluded from the evaluation and the reasons why?			
Does the timeline offer an appropriate scope for achieving the evaluation objectives?			
Is the proposed range of stakeholders groups and sampling targets an appropriate scope for achieving the evaluation objectives?			
4. The evaluation objectives are translated into relevant and specific evaluation questions.			From MFA 2013 Evaluation Manual/ OECD/DAC
4.1 The TOR apply the agreed OECD/DAC and MFA criteria for evaluation development assistance.			
For instance, are relevance, efficiency, effectiveness, impact and sustainability covered?			
If a particular criterion is not applied and/or any additional criteria added, is there an explanation in the TOR as to why?			
4.2 Relevance of the object of the evaluation is adequately addressed.			MFA and OECD/DAC
Are questions on the extent to which the objectives of the programme are consistent with beneficiaries' requirements, country priorities, global priorities and partners' and Finland's policies included?			
4.3 Impact of the object of the evaluation is adequately addressed.			MFA and OECD/DAC
Given the stage of implementation, are questions on how the programme has succeeded in contributing to its wider, overall objectives, i.e. impact for final beneficiaries included? The evaluation of impact covers intended and unintended short and long term, positive and negative impacts.			
4.4 Effectiveness of the object of the evaluation is adequately addressed.			MFA and OECD/DAC
Do questions seek to describe if the results have furthered the achievement of the programme purpose?			

Review of Terms of Reference

Standards	Rating	Comments	Sources of questions
4.5 Efficiency of the object of the evaluation is adequately addressed.			MFA and OECD/DAC
Do questions define how well the various activities have transformed the available resources into the intended results in terms of quantity, quality and timeliness?			
4.6 Sustainability of the object of the evaluation is adequately addressed.			MFA and OECD/DAC
Do questions refer to the likely continuation of achievements when external support ends?			
4.7 Cross-cutting objectives/themes/issues relating to human rights (HRBAP) are adequately addressed in the TOR evaluation questions. If not, an explanation is provided.			
4.8 Cross-cutting objectives/themes/issues relating to gender equality are adequately addressed in the TOR evaluation questions. If not, an explanation is provided.			
4.9 Cross-cutting objectives/themes/issues relating to reduction of inequalities are adequately addressed in the TOR evaluation questions. If not, an explanation is provided.			
4.10 Cross-cutting objectives/themes/issues relating to climate sustainability are adequately addressed in the TOR evaluation questions. If not, an explanation is provided.			
4.11 Together, do the evaluation questions provide evidence of clear focus and context of the evaluation?			
Does the prioritisation of the questions seem reasonable given the scope?			From discussion with reference group
Is the number of questions reasonable given the scope?			
Are the questions adequate for the tasks at hand? (quality of questions)			
Is there an explanation of why certain criteria are left out?			
5. The proposed methodology is appropriate and capable of addressing the evaluation questions.			From MFA 2013 Evaluation Manual/ OECD/DAC
5.1 General guidelines for the methodology are included in the TOR for data collection and analysis.			From MFA 2013 Evaluation Manual/ OECD/DAC
Is there general information on the methods to be used (e.g. qualitative or quantitative)?			
Is there information on the sources of data to be analysed?			

Review of Terms of Reference

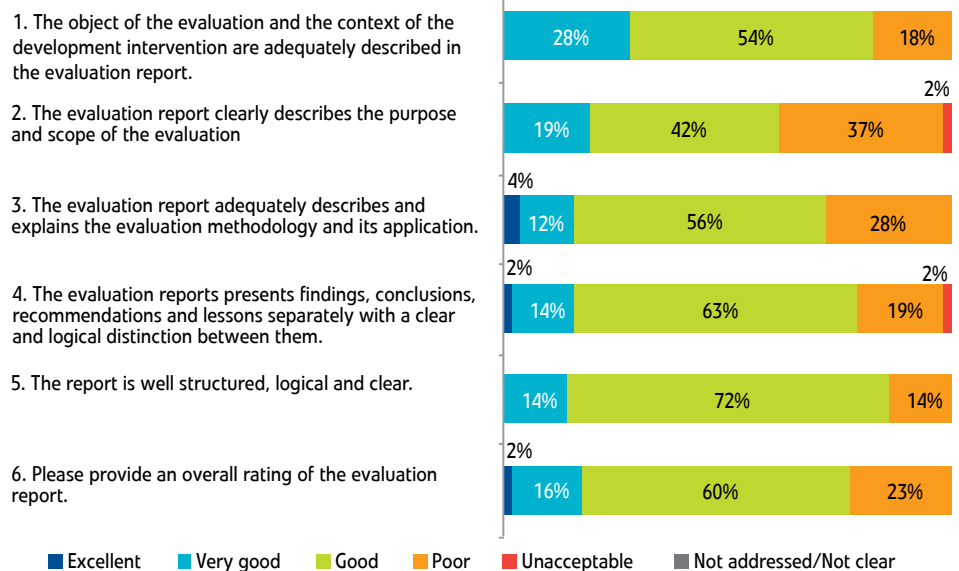
Standards	Rating	Comments	Sources of questions
Is there information on how data analysis will be conducted, ensuring for example that data will be disaggregated by gender, age group or other relevant group?			
5.2 The TOR contain reference to evaluability of the project/programme.			From OECD/DAC
Is there an evaluability assessment that covers aspects such as logical framework, availability of baseline data, indicators, targets, outputs (results) and outcome (project purpose) data available through M&E?			
Is there an assessment of data reliability or data availability?			
5.3 Ethical considerations are addressed in the ITT/TOR?			
6. The evaluation process and management structure are adequately described in the TOR or ITT.			From MFA 2013 Evaluation Manual/ OECD/DAC
6.1 The TOR adequately describe the evaluation phases, their sequencing and approximate duration.			From MFA 2013 Evaluation Manual
Do the TOR describe phases such as the kick-off meeting, inception and desk study phase, inception meeting, interviews and field missions, reporting, presentation of the evaluation results in the field, reporting and presentation of the evaluation results?			
Are key milestones in the process described?			
6.2 The evaluation governance and management structures described in the TOR are designed to adequately fit the evaluation's context, purpose, scope and objectives.			From OECD/DAC
Does the governance structure include a description of the role to be played by MFA during the evaluation and the tasks of the evaluation team?			
Are responsibilities of each stakeholder clearly described?			
7. The deliverables to be submitted during the evaluation are clearly described in the TOR or ITT.			From MFA 2013 Evaluation Manual
7.1 The reports and outputs (results) to be submitted in each phase of the evaluation are clearly specified.			From MFA 2013 Evaluation Manual
Do the TOR specify types of deliverables required such as: inception report, presentation on the field findings, draft final report, final report, presentation of the evaluation findings.			
7.2 There is a specific request in the ITT or TOR to propose and implement a quality assurance system for the evaluation and its deliverables.			From MFA 2013 Evaluation Manual

Review of Terms of Reference

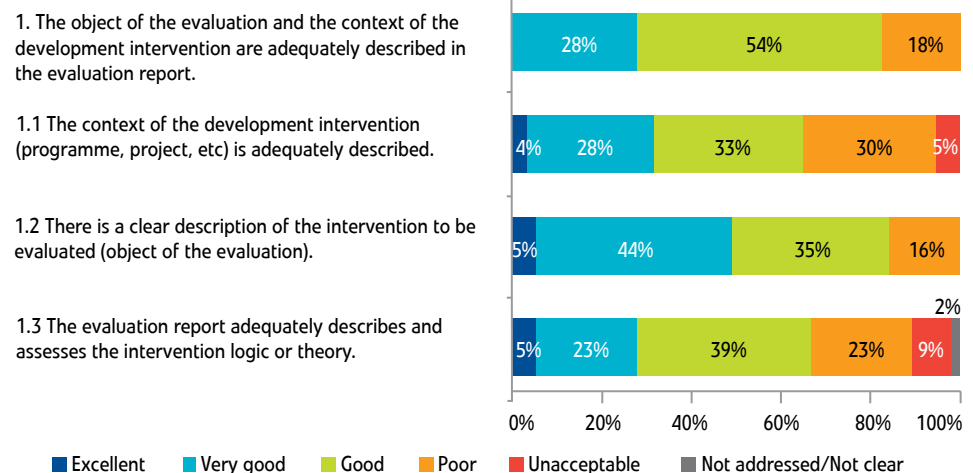
Standards	Rating	Comments	Sources of questions
8. The resources required for this evaluation are sufficiently described in the TOR or ITT.			From MFA 2013 Evaluation Manual
8.1 The expertise required to conduct the mandate is sufficiently described by the TOR or ITT (overall composition of the evaluation team).			From MFA 2013 Evaluation Manual
Is the desired expertise/knowledge of the team described in the TOR (or ITT)?			
Does the team composition require both international and national experts? Do the ToR request a team leader to be nominated ?			
Does the evaluating team have to possess a mix of evaluative skills and thematic knowledge?			
Is the gender balance of the team considered?			
Are the the following field specified: programme evaluations and planning in the relevant sector, relevant sectors in developing countries, integrating cross-cutting objectives, HRBAP etc.?			
Is the expertise required in terms of cross-cutting objectives/themes/issues described?			
9. The TOR are logically structured and clearly written.			From UNEG Standards
9.1 The TOR sections hold together in a logically consistent way that will allow for a coherent evaluation Report.			From UNEG Standards
9.2 The style of the TOR is adequate (brief, to the point, logically structured and easy to understand).			From UNEG Standards
10. Please provide an overall rating of the TOR.			
10.1 Do the TOR give focus and direction to enable a good evaluation response?			
10.2 Is there enough information for evaluators to prepare a proposal?			
10.3 Are the resources allocated sufficient to carry out the evaluation (alignment between scope and resources)?			

ANNEX 9 QUALITY ASSESSMENT: SUMMARY RATINGS OF EVALUATION REPORTS

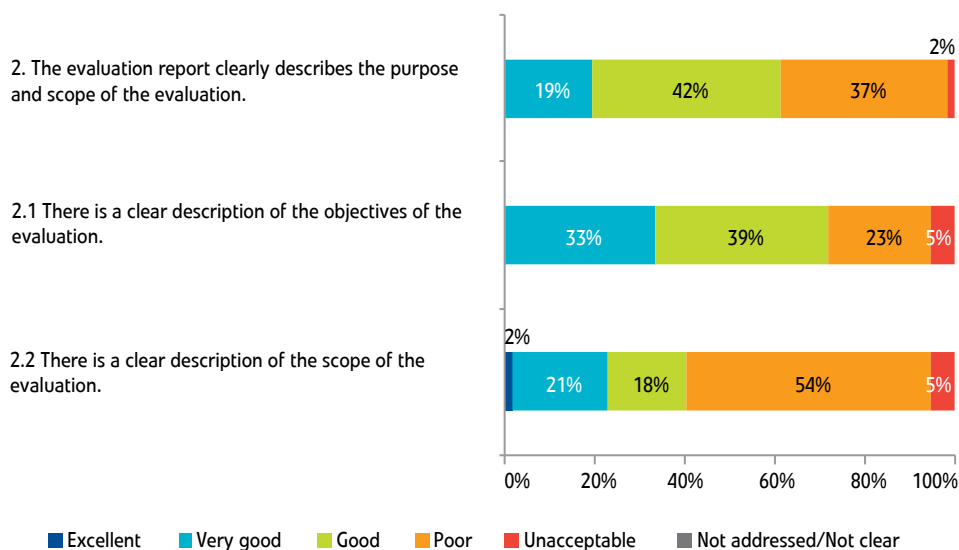
Summary – Main Questions



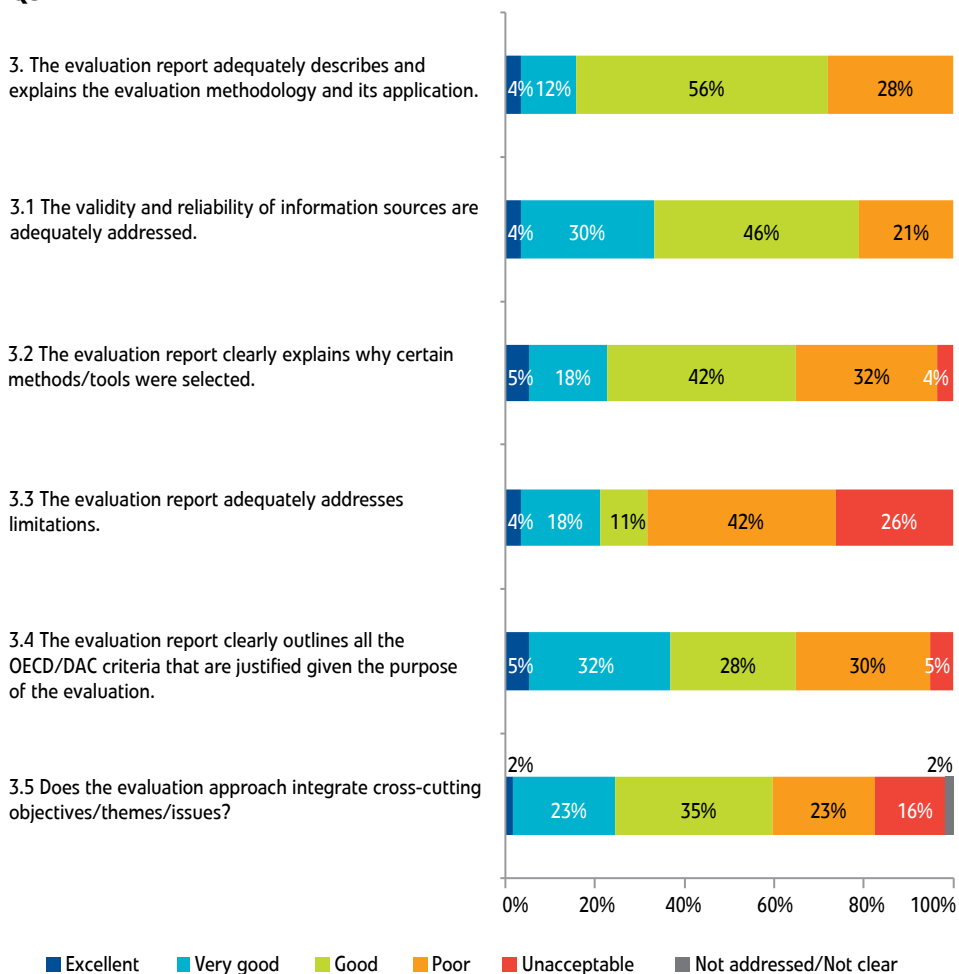
Q1 –



Q2 –



Q3 –



Q4-4.5

4. The evaluation reports presents findings, conclusions, recommendations and lessons separately with a clear and logical distinction between them.

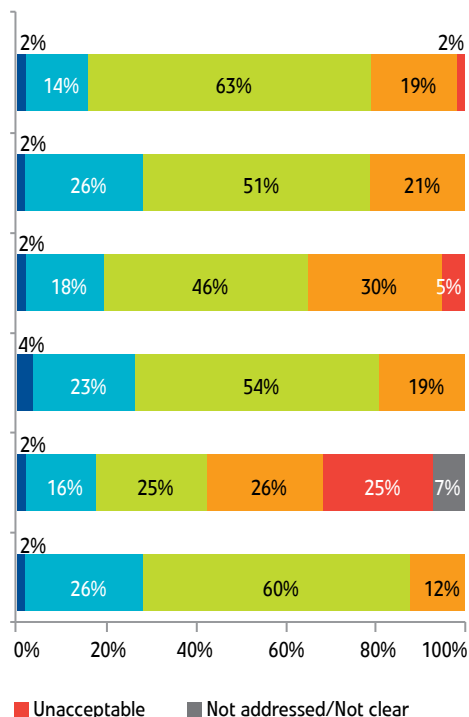
4.1 Findings are clearly presented and based on the objective use of the reported evidence.

4.2 Conclusions are clearly substantiated by findings and analysis.

4.3 Recommendations are well-grounded in the evidence and conclusions reported.

4.4 Are lessons learned correctly identified?

4.5 The evaluation report as a whole holds together in a clear and coherent way.



Q4.6

4.6 The report adequately addresses all the OECD/DAC criteria.

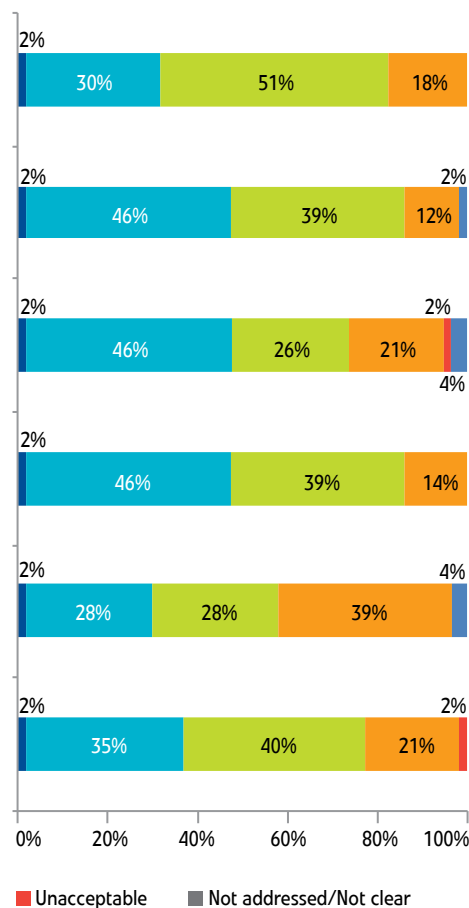
Is the relevance of the intervention adequately addressed?

Is the efficiency of the intervention adequately addressed?

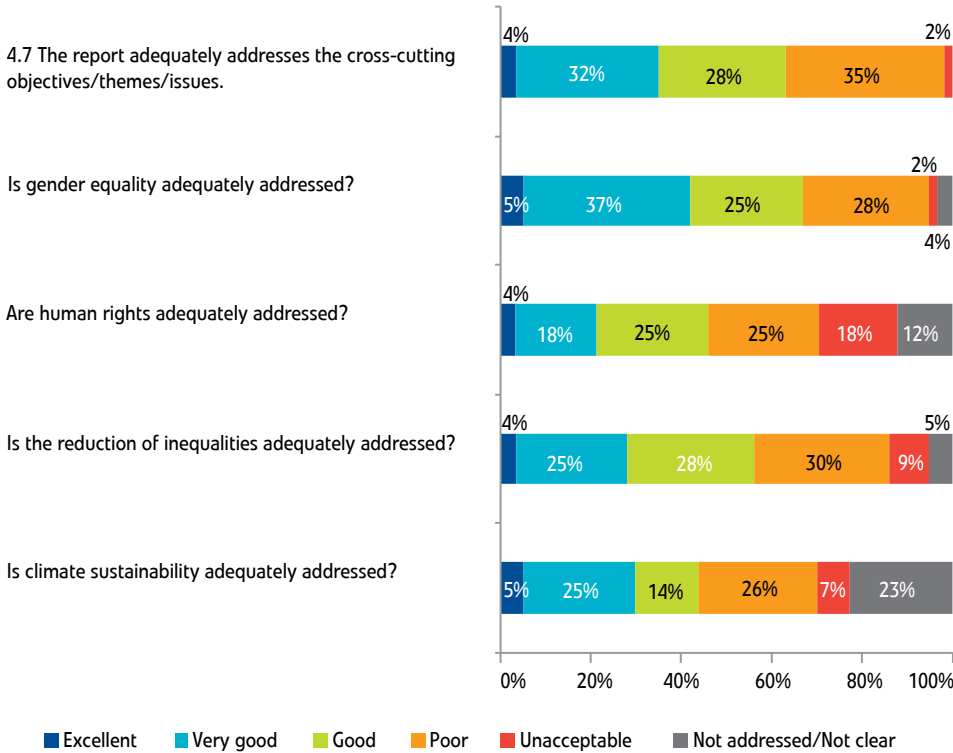
Is the effectiveness of the intervention adequately addressed?

Is the impact of the intervention adequately addressed?

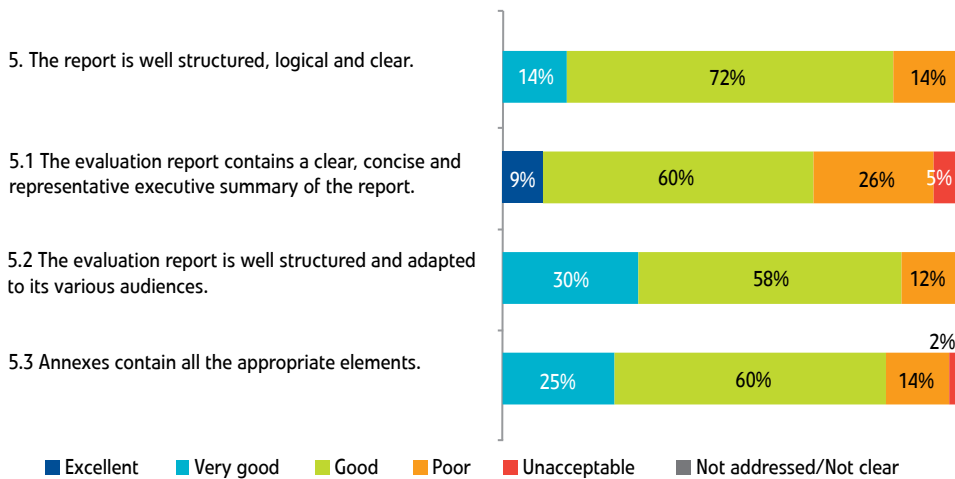
Is the sustainability of the intervention adequately addressed?



Q4.7

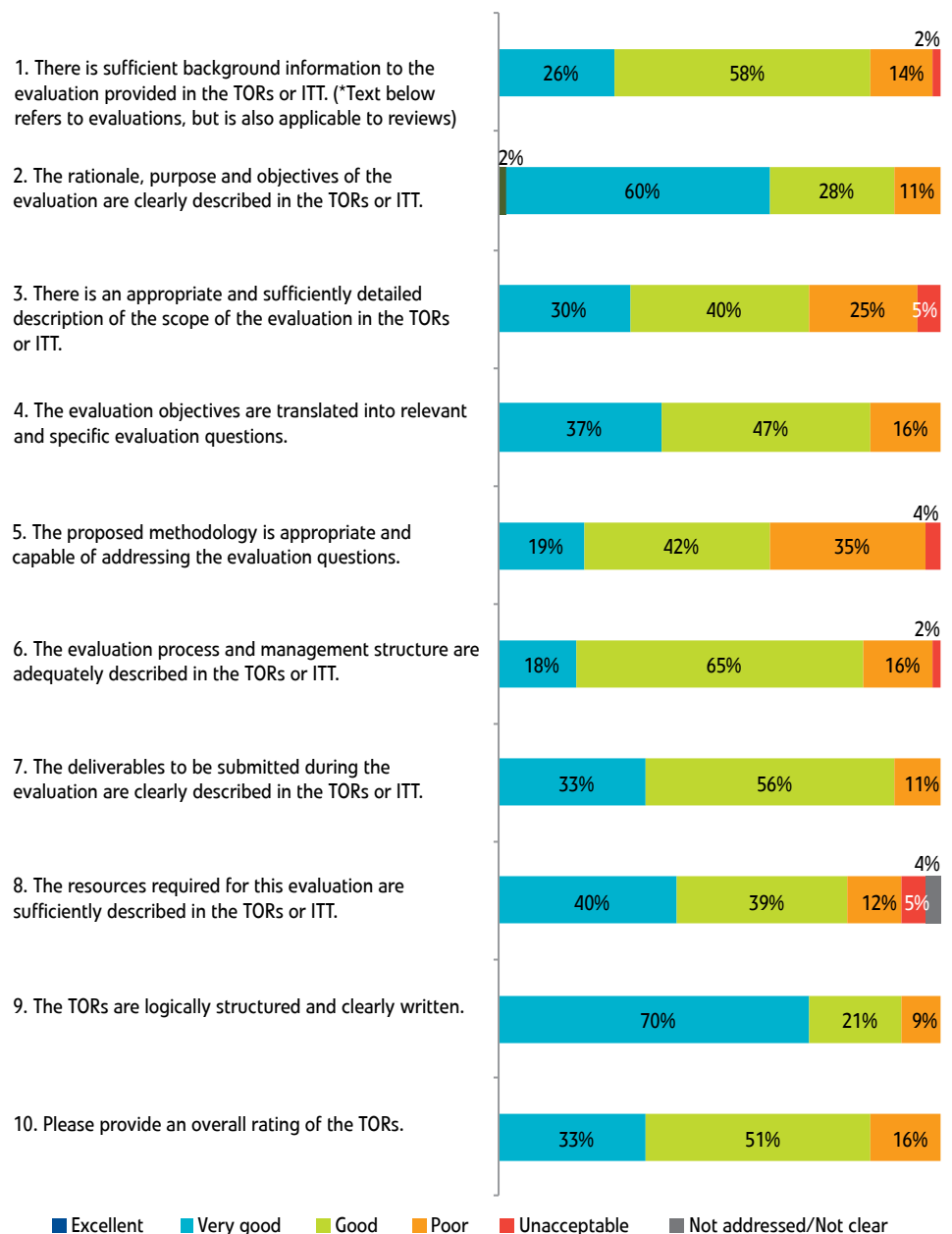


Q5



ANNEX 10 QUALITY ASSESSMENT: SUMMARY RATINGS OF TOR

Summary – Main Questions

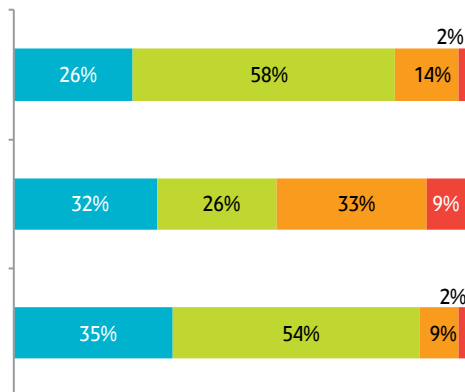


Q1

1. There is sufficient background information to the evaluation provided in the TORs or ITT. (*Text below refers to evaluations, but is also applicable to reviews)

1.1 The programme context (policy, country, regional, global, thematic context) is sufficiently described.

1.2 There is a clear description of the programme to be evaluated.



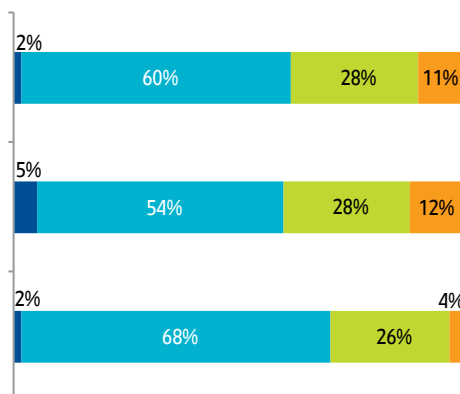
■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q2

2. The rationale, purpose and objectives of the evaluation are clearly described in the TORs or ITT.

2.1 The rationale and purpose of the evaluation are stated clearly in the TORs.

2.2 The specific objectives of the evaluation are clearly stated.

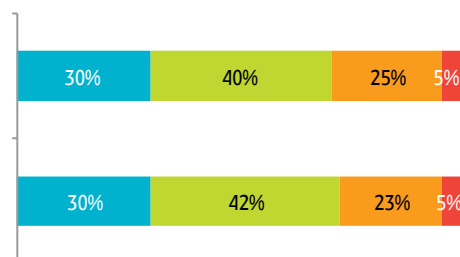


■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q3

3. There is an appropriate and sufficiently detailed description of the scope of the evaluation in the TORs or ITT.

3.1 The scope of the evaluation clearly describes the time span the evaluation covers, stakeholder groups involved, and geographical areas covered.



■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q4

4. The evaluation objectives are translated into relevant and specific evaluation questions.



4.1 The TORs apply the agreed OECD/DAC and MFA criteria for evaluation development assistance.



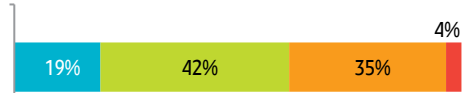
4.11 Together, do the evaluation questions provide evidence of clear focus and context of the evaluation?



■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q5

5. The proposed methodology is appropriate and capable of addressing the evaluation questions.



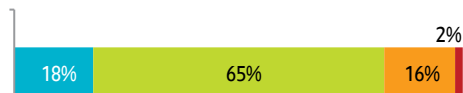
5.1 General guidelines for the methodology are included in the TORs for data collection and analysis.



■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q6

6. The evaluation process and management structure are adequately described in the TORs or ITT.



6.1 The TORs adequately describe the evaluation phases, their sequencing and approximate duration.



6.2 The evaluation governance and management structures described in the TORs are designed to adequately fit the evaluation's context, purpose, scope and objectives.



■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q7

7. The deliverables to be submitted during the evaluation are clearly described in the TORs or ITT.



7.1 The deliverables to be submitted in each phase of the evaluation are clearly specified.



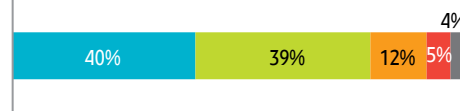
■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q8

8. The resources required for this evaluation are sufficiently described in the TORs or ITT.



8.1 The expertise required to conduct the mandate is sufficiently described by the TORs or ITT (overall composition of the evaluation team).



■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

Q9

9. The TORs are logically structured and clearly written.



9.1 The TORs sections hold together in a logically consistent way that will allow for a coherent evaluation Report.



9.2 The style of the TORs is adequate (brief, to the point, logically structured and easy to understand).



■ Excellent ■ Very good ■ Good ■ Poor ■ Unacceptable ■ Not addressed/Not clear

ANNEX 11 DEVELOPMENT CO-OPERATION ASSESSMENT: REVIEW TOOL

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
Effectiveness 1. The evaluation report indicates that the project or programme supported by the MFA achieved their stated objectives (or are likely to do so in the near future). Effectiveness: The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.		Excellent = Programme exceeds objectives, or is likely to exceed the planned objectives/results Very good = Programme fully met objectives, or is likely to fully meet the planned objectives/results Good = Mostly met objectives Poor = Barely met objectives Unacceptable = No objective is met Not addressed/Not clear = Effectiveness was not addressed in the report (deficiency of the evaluation)		Sum up information from the report and write a high level comment. Take into account whether this is a mid-term evaluation and whether the progress towards objectives is congruent with the implementation period (actual implementation time vs. planned). Comments need to be grounded in programme objectives, i.e. to what extent were those realised?	OECD/DAC: Glossary of Key Terms in Evaluation and Results Based Management
1.1 Are the outputs achieved as planned in the implementation period?				Base rating on what is reported in the evaluation report (preferred). When necessary, reviewer's judgment is acceptable.	
1.2 Are the outcomes achieved as planned in the implementation period?				Base rating on what is reported in the evaluation report (preferred). When necessary, reviewer's judgment is acceptable.	
1.3 What are the main reported successes of the project or programme?				Instead of noting specifics from the report, be more general in the types of successes you report.	
1.4 What are the main reported shortcomings of the project or programme?				Instead of noting specifics from the report, be more general in the types of shortcomings you report. Reported shortcomings should be within the programme's control, e.g. was programme design one of the shortcomings? See if you can identify weaknesses in the design vs. project implementation.	

1.5 What reported factors facilitated the successful implementation of the project or programme?				Note the main factors, or contextual issues	MFA Manual/ Policy
1.6 What reported factors hindered the successful implementation of the project or programme?				Note the main implementing issues, or contextual issues outside programme control.	MFA Manual/ Policy
Impact					
2. The evaluation report indicates that projects and programmes supported by the MFA have had impact level results when appropriate (even if not an "impact evaluation" per se). Impacts: Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.			<p>Excellent = Programme exceeds planned impact level results or is likely to exceed impact results</p> <p>Very good = Evaluation indicates that programme fully achieved impact results, or is likely to fully achieve the planned impact level results</p> <p>Good = Mostly achieved impact level results</p> <p>Poor = Barely achieved impact level results</p> <p>Unacceptable = No impact level result has been achieved</p> <p>Not addressed/Not clear = This is a mid-term evaluation</p>	Sum up information from the report and write a high level comment. Rating required for mid-term evaluations: N/A. If a section labeled "impacts" actually contains information on outputs or outcomes, these examples should not be included here.	OECD/DAC: Glossary of Key Terms in Evaluation and Results Based Management
2.1 What are the main reported impacts of the project or programme (intended)?				Not required for MTE. Look for "types" of impacts, but will need to see it is possible to establish some kind of typology once we see patterns from the first set of reviews.	MFA Evaluation Manual 2013 page 27
2.2 What are the main reported impacts of the project or programme (unintended)?				Not required for MTE. Look for "types" of impacts	MFA Evaluation Manual 2013 page 27
2.3 Do the reported results make a contribution towards reducing poverty?				Not required for MTE. Please note what the contributions are.	Development Programme 2012, page 27

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
<p>Relevance</p> <p>3. The evaluation reports indicates that the project or programme supported by MFA is relevant.</p> <p>Relevance: The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies.</p>		<p>Excellent = Programme exceeds expectations for relevance</p> <p>Very good = Evaluation indicates that programme fully achieved expectations for relevance</p> <p>Good = Mostly achieved expectations for relevance</p> <p>Poor = Barely achieved expectations for relevance</p> <p>Unacceptable = No expectation for relevance has been achieved</p> <p>Not addressed/Not clear = Relevance is not addressed in the report (deficiency of the evaluation)</p>		<p>Sum up information from the report and write a high level comment.</p>	OECD/DAC: Glossary of Key Terms in Evaluation and Results Based Management
3.1 The evaluation report indicates that the project or programme supported by the MFA is suited to the needs and/or priorities of the target groups.				Could be ultimate target groups, beneficiaries or intermediary groups.	MFA Evaluation Manual 2013 page 27
3.2 The evaluation report indicates that the project or programme supported by the MFA aligns with national development goals and policies of partner countries.					MFA Evaluation Manual 2013 page 27
3.3 The evaluation report indicates that the programme is aligned with global trends and best practices.				Global practices will vary by sector (e.g., education) or type of intervention (e.g., capacity building). Base your rating on specific judgments and evidence provided by the evaluator.	
3.4 What are the factors that affected the relevance of the program in positive ways?					
3.5 What are the factors that affected the relevance of the program in negative ways?					

<p>3.6 Do the intended results reported in the evaluation report link to Finland's priorities for development cooperation for the period under review (Based on Development Policy Programme 2012, p. 27)? And more specifically on:</p>					<p>MFA Development Policy Programme 2012 (list on page 27)</p>
<p>3.6.1 A democratic and accountable society that promotes human rights</p>				<p>Interpret information provided and use the words below to summarize your thoughts:</p> <p>This relates to such issues as supporting partner countries' capacity for monitoring and implementing human rights, independent judiciary, freedom of expression, association and assembly, free and fair elections, an accountable government, freedom for civil society, access to decision-making in society, transparency of how public funds are spent, corruption, peace mediation activities, reforming the security sector, disarming combatants, etc.</p>	<p>MFA Development Policy Programme 2012, pages 28-31</p>
<p>3.6.2 An inclusive green economy that promotes employment</p>				<p>Interpret information provided and use the words below to summarize your thoughts:</p> <p>This relates to generating human development with the sustainable use of natural resources, and creating equal possibilities for all for decent work. This may include such things as bringing micro-entrepreneurs into the formal economy through mobilizing small and medium-sized enterprises, integration of countries into international trade in a balanced way, preventing tax evasion, illicit capital flight, improved sharing of information, accounting standards, corporate social responsibility, and innovation.</p>	<p>MFA Development Policy Programme 2012 pages 32-35</p>

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
Relevance					
3.6.3 Sustainable management of natural resources and environmental protection				<p>Interpret information provided and use the words below to summarize your thoughts:</p> <p>Refers to issues such as energy efficiency and renewable energy, sustainable urban development, food security, agricultural productivity, drinking water and sanitation, management of water resources, forests and community forestry.</p>	MFA Development Policy Programme 2012, pages 36-39
3.6.4 Human development				<p>Interpret information provided and use the words below to summarize your thoughts:</p> <p>Refers to issues such as literacy, quality education (whether basic, vocational, or higher education), access to education, including people with special needs or from linguistic minorities, high quality research and innovation linked with producing skilled labour for entrepreneurship, and health issues such as maternal and child health, community-based preventive health work, and communicable and non-communicable diseases.</p>	MFA Development Policy Programme 2012, pages 40-41

Efficiency			
<p>4. The evaluation report indicates that the programme was generally efficient.</p> <p>Efficiency: A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.</p>		<p>Excellent = Programme exceeds expectations for efficiency</p> <p>Very good = Evaluation indicates that programme fully achieved expectations for efficiency</p> <p>Good = Mostly achieved expectations for efficiency</p> <p>Poor = Barely achieved expectations for efficiency</p> <p>Unacceptable = No expectation for efficiency has been achieved</p> <p>Not addressed/Not clear = Efficiency was not addressed in the report (deficiency of the evaluation)</p>	<p>Sum up information from the report and write a high level comment.</p>
<p>4.1 The evaluation report indicates that the programme is efficient in transforming inputs into outputs.</p>			<p>OECD/DAC http://www.oecd.org/development/evaluation/dacriteriaforevaluatingdevelopmentassistance.htm</p>
<p>4.2 The evaluation report indicates that objectives were achieved on schedule.</p>			<p>OECD/DAC http://www.oecd.org/development/evaluation/dacriteriaforevaluatingdevelopmentassistance.htm</p>
<p>4.3 The evaluation report indicates that systems and procedures (management) for project or programme implementation and follow up are efficient.</p>			<p>MFA Evaluation Manual 2013 page 28</p>

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
Efficiency					
4.4 What are the factors that affect the efficiency of the program in positive ways?					
4.5 What are the factors that affect the efficiency of the program in negative ways?					
Sustainability					
5. The evaluation report indicates that benefits of the project or programme are likely to continue after project or programme completion. Sustainability: The continuation of benefits from a development intervention after major development assistance has been completed.		<p>Excellent = Evaluation indicates that programme is highly likely to sustain benefits after programme completion</p> <p>Very good = Evaluation indicates that programme is likely to sustain programme benefits after programme completion</p> <p>Good = There is some likelihood that programme benefits will be sustained after programme completion</p> <p>Poor = There is little likelihood that programme benefits will be sustained after programme completion</p> <p>Unacceptable = There is no likelihood that programme benefits will be sustained after programme completion.</p> <p>Not addressed/Not clear = Sustainability was not addressed in the report (deficiency of the evaluation)</p>		<p>Sum up information from the report and write a high level comment.</p>	OECD/DAC: Glossary of Key Terms in Evaluation and Results Based Management
5.1 The programme is sustainable from a social perspective.				<p>Social sustainability refers to sensitivity to cultural issues, including the dynamics of relationships amongst stakeholders, and broad participation of beneficiaries through an inclusive approach. Rights to assess to information and freedom of opinion, and ensuring that a programme should do no harm to easily marginalised groups.</p>	Bilateral Manual page 20

5.2 The programme is sustainable from a financial/economic perspective.			<p>Financial sustainability refers to programme benefits being sustained after external financing ends, such as prospects for funding activities through incomes from improved services, beneficiary contribution, tax or other domestic revenues over longer term.</p> <p>Economic sustainability refers to the distribution of benefits and costs among programme stakeholders, as well as the flows of resources between them, analyzed and depicted separately for each stakeholder, especially final beneficiaries (qualitative analysis is normally sufficient).</p>	Bilateral Manual page 21
5.3 The programme is sustainable from an environmental perspective.			<p>Refers to issues such as ecological sustainability, sustainable use of natural resources, prevention of environmental pollution, biodiversity, desertification, soil impoverishment, sustainable use and control of chemicals and waste, social protection and adaptation for those who are most vulnerable to environmental disasters and climate change, and/or consideration of environmental regulations such as Environmental Impact Assessment and international conventions such as the United Nations Framework Convention on Climate Change.</p>	Bilateral Manual page 21
5.4 The programme is sustainable from a technical perspective.			<p>Refers to appropriate technology selections and building up capacity for sustainable operation and maintenance of technologies. Includes considerations such as whether the technology is socially acceptable, affordable, generates employment, and replicable.</p>	Bilateral Manual page 23

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
Sustainability					
5.5 The programme is sustainable from an institutional perspective.				Refers to whether the capacity of the implementing organisation is sufficient to manage the programme and deliver its results. Considerations include external factors affecting the organisation (e.g. political climate in country), relationships with other organisations, capacity for planning, budgeting, etc. management practices (e.g. organisational structure) and financial management systems.	Bilateral Manual page 23/24
5.6 What factors are identified as enhancing sustainability of results?					MFA Evaluation Manual 2013 page 29
5.7 What factors are identified as inhibiting sustainability of results?					MFA Evaluation Manual 2013 page 29
Human Rights-Based Approach					
6. The evaluation report indicates that the project or programme supported by the MFA effectively addresses the cross-cutting objective of HRBA.		<p>Excellent = Programme exceeds planned results in the area of HRBA or is likely to exceed planned results</p> <p>Very good = Evaluation indicates that programme fully achieved results on HRBA, or is likely to fully achieve the planned results</p> <p>Good = Mostly achieved results in the area of HRBA</p> <p>Poor = Barely achieved results in the area of HRBA</p> <p>Unacceptable = No result has been achieved in the area of HRBA</p> <p>Not addressed/Not clear = HRBA was not addressed in the report (deficiency of the evaluation)</p>		<p>Sum up information from the report and write a high level comment.</p> <p>Value-based development policy promotes the core human rights principles such as universality, self-determination, non-discrimination and equality. The human rights-based approach to development includes civil and political rights and freedoms as well as economic, social and cultural rights. Finland emphasises the rights of women, children, ethnic, linguistic and religious minorities and indigenous people, the rights of persons with disability, people living with HIV and AIDS, and the rights of sexual and gender minorities.</p>	<p>MFA Evaluation Manual 2013 pages 27-29</p> <p>2012 Development Policy Programme</p>

6.1 The programme is congruent with the MFA policy on HRBA.					MFA Evaluation Manual 2013 page 28
6.2 What are the reported successes related to HRBA?					
6.3 What are the reported shortcomings related to HRBA?					
6.4 Is there evidence in the report that achievements in terms of HRBA will be sustained after the programme is completed?					MFA Evaluation Manual 2013 page 29
Cross-Cutting Objectives					
Gender Equality					
7. The evaluation report indicates that the project or programme supported by the MFA effectively contributes to the cross-cutting objective of gender equality (planned results in this area are realized).		<p>Excellent = Programme exceeds planned results in the area of gender equality or is likely to exceed planned results</p> <p>Very good = Evaluation indicates that programme fully achieved results on gender equality, or is likely to fully achieve the planned results</p> <p>Good = Mostly achieved results in the area of gender equality</p> <p>Poor = Barely achieved results in the area of gender equality</p> <p>Unacceptable = No result has been achieved in the area of gender equality</p> <p>Not addressed/Not clear = Gender equality was not addressed in the report (deficiency of the evaluation)</p>		<p>Sum up information from the report and write a high level comment.</p> <p>Gender equality includes strengthening the role of women, through economic development and the promotion of well-being. Also includes support to the participation of women in decision-making and rejects any form of discrimination giving rise to gender inequality.</p>	OECD/DAC Manual for bilateral programmes 2012 Development Policy Programme 2012
7.1 The programme is congruent with the MFA policy on gender equality.					
7.2 What are the reported successes related to gender equality (including any level of results: impact, outcomes or outputs)?					MFA Evaluation Manual 2013 page 28
7.3 What are the reported shortcomings related to gender equality?					

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
Cross-Cutting Objectives					
7.4 Is there evidence in the report that achievements in terms of gender equality will be sustained after the programme is completed?					MFA Evaluation Manual 2013 page 29
Reduction of Inequality					
8. The evaluation report indicates that the project or programme supported by the MFA effectively addresses the cross-cutting objective of reduction of inequality.		<p>Excellent = Programme exceeds planned results in the area of reduction of inequality or is likely to exceed planned results</p> <p>Very good = Evaluation indicates that programme fully achieved results on reduction of inequality or is likely to fully achieve the planned results</p> <p>Good = Mostly achieved results in the area of reduction of inequality</p> <p>Poor = Barely achieved results in the area of reduction of inequality</p> <p>Unacceptable = No result has been achieved in the area of reduction of inequality</p> <p>Not addressed/Not clear = Reduction of inequality was not addressed in the report (deficiency of the evaluation)</p>		<p>Sum up information from the report and write a high level comment.</p> <p>Reduction of inequalities includes support to social policies that increase equal opportunities for social, economic, and political participation as well as access to basic services and a social protection floor. Good nutrition, health, education, decent work and basic social protection as well as the realisation of the basic labour rights have a key role.</p>	OECD/DAC Manual for bilateral programmes 2012 Development Policy Programme
8.1 The programme is congruent with the MFA policy on reduction of inequalities.					
8.2 What are the reported successes related to reduction of inequality?					MFA Evaluation Manual 2013 page 28
8.3 What are the reported shortcomings related to reduction of inequality?					
8.4 Is there evidence in the report that achievements in terms of reduction of inequalities will be sustained after the programme is completed?					MFA Evaluation Manual 2013 page 29

Climate Sustainability			
9. The evaluation report indicates that the project or programme supported by the MFA effectively addresses the cross-cutting objective of climate sustainability.	<p>Excellent = Programme exceeds planned results in the area of climate sustainability or is likely to exceed planned results</p> <p>Very good = Evaluation indicates that programme fully achieved results on climate sustainability or is likely to fully achieve the planned results</p> <p>Good = Mostly achieved results in the area of climate sustainability</p> <p>Poor = Barely achieved results in the area of climate sustainability</p> <p>Unacceptable = No result has been achieved in the area of climate sustainability</p> <p>Not addressed/not clear = Climate sustainability was not addressed in the report (deficiency of the evaluation)</p>	<p>Sum up information from the report and write a high level comment.</p> <p>Finland promotes low carbon development and the capacity of its partner countries to adapt to climate change, and furthers integration of these goals into partner countries' own development planning. Finland also supports long-term measures that reduce the vulnerability of people and communities to natural disasters.</p>	OECD/DAC Manual for bilateral programmes 2012 Development Policy Programme
9.1 The programme is congruent with the MFA policy on climate sustainability.			
9.2 What are the reported successes related to climate sustainability?			MFA Evaluation Manual 2013 page 28
9.3 What are the reported shortcomings related to climate sustainability?			
9.4 Is there evidence in the report that achievements in terms of climate sustainability will be sustained after the programme is completed?			MFA Evaluation Manual 2013 page 29

OECD/DAC Criteria	Rating	Definition of Rating Scale	Comments	Notes to Reviewers	Source
Other Cross-cutting Objective or Emerging Theme					
<p>10 The evaluation report indicates that the project or programme supported by the MFA effectively addresses another cross-cutting objectives or emerging themes.</p>		<p>Excellent = Programme exceeds planned results in the area of another cross-cutting objective/theme or is likely to exceed planned results</p> <p>Very good = Evaluation indicates that programme fully achieved results in another cross-cutting objective/theme, or is likely to fully achieve the planned results</p> <p>Good = Mostly achieved results in the area of another cross-cutting objective/theme</p> <p>Poor = Barely achieved results in the area of another cross-cutting objective/theme</p> <p>Unacceptable = No result has been achieved in the area of another cross-cutting objective/theme, although one is present in the programme</p> <p>Not addressed/not clear = There is no other cross-cutting objective or emerging theme indicated</p>		<p>Sum up information from the report and write a high level comment.</p> <p>MFA is interested in whether there are cross-cutting objectives or emerging themes from projects OTHER than those listed in their manual. Please briefly describe it under Comments. Those may include: capacity building, innovation, etc.</p> <p>Leave blank if there is no emerging theme</p>	MFA, verbal request
<p>10.1 What was the other cross-cutting objective(s) or emerging theme(s)?</p>				<p>How would you label this objective or theme in a few words? Leave blank if there is no emerging theme</p>	MFA, verbal request
<p>10.2 What are the reported successes related to the other cross-cutting objective(s) or emerging theme(s)?</p>				<p>Leave blank if there is no emerging theme</p>	
<p>10.3 What are the reported shortcomings related to the other cross-cutting objective(s) or emerging theme(s)?</p>				<p>Leave blank if there is no emerging theme</p>	
<p>10.4 Is there evidence in the report that achievements in terms of the other cross-cutting objective(s) or emerging theme(s) will be sustained after the programme is completed?</p>				<p>Leave blank if there is no emerging theme</p>	MFA Evaluation Manual 2013 page 29

Risk Management				MFA Evaluation Manual 2013 page 28
11. The evaluation report indicates that risks related to the project or programme have been appropriately managed.	<p>Excellent = Programme exceeds expectations for risk management</p> <p>Very good = Evaluation indicates that programme fully achieved expectations for risk management</p> <p>Good = Mostly achieved expectations for risk management</p> <p>Poor = Barely achieved expectations for risk management</p> <p>Unacceptable = No expectations for risk management have been achieved</p> <p>Not addressed/not clear = Risk management is not addressed in the report (deficiency of the evaluation)</p>	<p>Sum up information from the report and write a high level comment.</p> <p>You may need to infer what risks are. Risk related to sustainability should be described in more detail, e.g. what is the underlying risk that may cause sustainability issues?</p> <p>Types of risks according to the bilateral manual include:</p> <ul style="list-style-type: none"> - Developmental risks - Financial of fiduciary risks - Reputational risks - Risks of action and avoiding actions particularly in fragile states 		
11.1 What type(s) of risks were reported to have been well managed in the evaluation report?				
11.2 What type(s) of risks were reported to have NOT been well managed in the evaluation report?				
Additional Comment				
12. Include any additional comments or examples of best practices on how the programme contributed to Finland's development cooperation.				

ANNEX 12 DEVELOPMENT CO-OPERATION ASSESSMENT: SUMMARY OF RATINGS

Summary – Main Questions

QUESTION 1 The evaluation report indicates that the project or programme supported by the MFA achieved their expected results and project purposes (or are likely to do so in the near future).

QUESTION 2 The evaluation report indicates that programmes have achieved their overall objectives.

QUESTION 3 The evaluation reports indicates that the programme is relevant.

QUESTION 4 The evaluation report indicates that the programme was efficient.

QUESTION 5 The evaluation report indicates that benefits of the programme are likely to continue after programme completion.

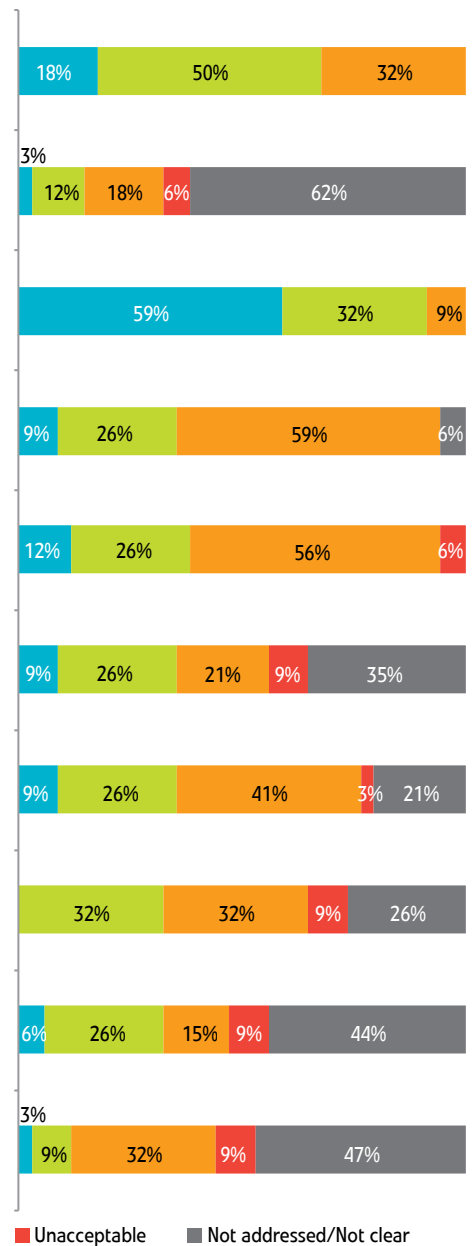
QUESTION 6 The evaluation report indicates that the programme effectively addresses HRBA.

QUESTION 7 The evaluation report indicates that the programme effectively contributes to the cross-cutting objective of gender equality.

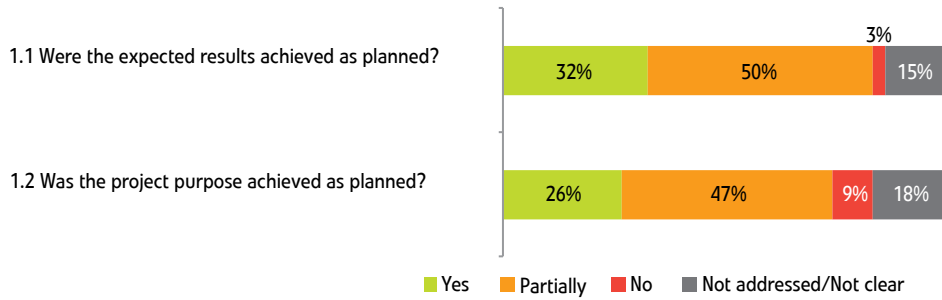
QUESTION 8 The evaluation report indicates that the programme effectively addresses the cross-cutting objective of reduction of inequality.

QUESTION 9 The evaluation report indicates that the programme effectively addresses the cross-cutting objective of climate sustainability.

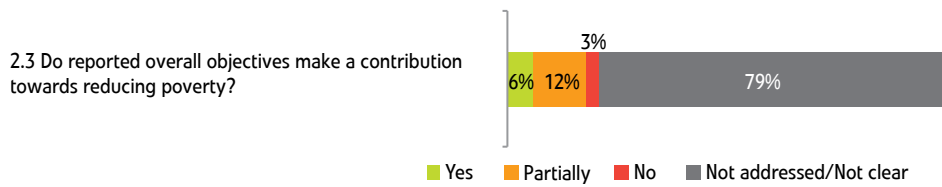
QUESTION 11 The evaluation report indicates that risks related to the programme have been appropriately managed.



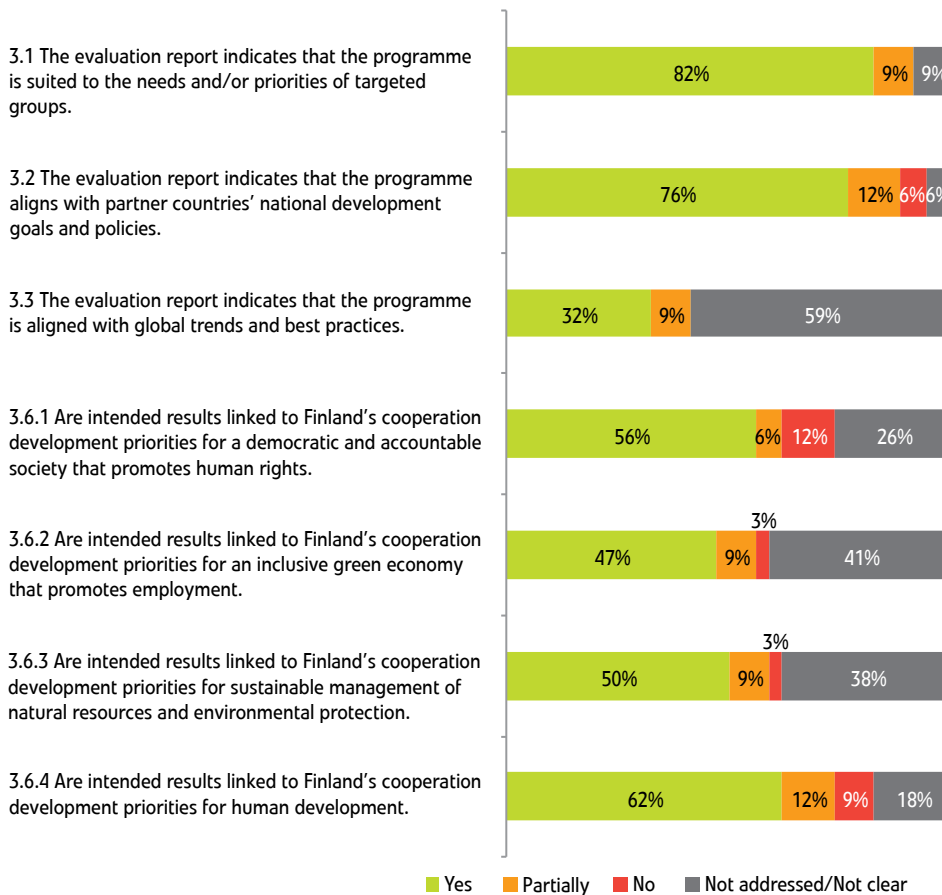
Q1



Q2

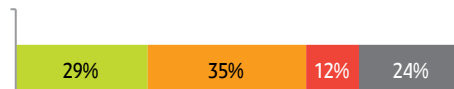


Q3



Q4

4.1 The evaluation report indicates that the programme is efficient in transforming inputs into expected results.



4.2 The evaluation report indicates that expected results were achieved on schedule.



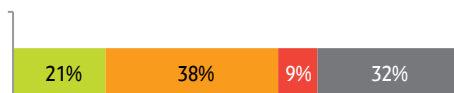
4.3 The evaluation report indicates that systems and procedures for programme implementation and follow up are efficient.



■ Yes ■ Partially ■ No ■ Not addressed/Not clear

Q5

5.1 The programme is sustainable from a social perspective.



5.2 The programme is sustainable from a financial/economic perspective.



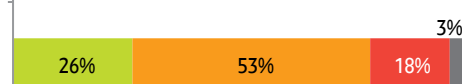
5.3 The programme is sustainable from an environmental perspective.



5.4 The programme is sustainable from a technical perspective.



5.5 The programme is sustainable from an institutional perspective.



■ Yes ■ Partially ■ No ■ Not addressed/Not clear

Q6

6.1 The programme is congruent with the MFA policy on HRBA.



6.4 Is there evidence in the report that achievements in terms of HRBA will be sustained after the programme is completed?



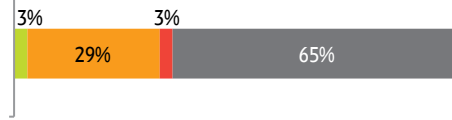
■ Yes ■ Partially ■ No ■ Not addressed/Not clear

Q7

7.1 The programme is congruent with the MFA policy on gender equality.



7.4 Is there evidence in the report that achievements in terms of gender equality will be sustained after the programme is completed?



■ Yes ■ Partially ■ No ■ Not addressed/Not clear

Q8

8.1 The programme is congruent with the MFA policy on reduction of inequalities.



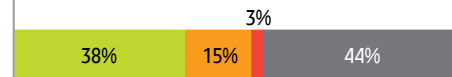
8.4 Is there evidence in the report that achievements in terms of reduction of inequalities will be sustained after the programme is completed?



■ Yes ■ Partially ■ No ■ Not addressed/Not clear

Q9

9.1 The programme is congruent with the MFA policy on climate sustainability.



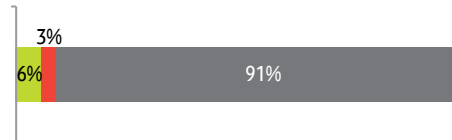
9.4 Is there evidence in the report that achievements in terms of climate sustainability will be sustained after the programme is completed?



■ Yes ■ Partially ■ No ■ Not addressed/Not clear

Q10

10.4 Is there evidence in the report that achievements in terms of the other cross-cutting objective(s) or emerging theme(s) will be sustained after the programme is completed?



■ Yes ■ Partially ■ No ■ Not addressed/Not clear

ANNEX 13 CONSULTATION PROTOCOL ON CAPACITY BUILDING SERVICES AND EVALUATION MANAGEMENT CYCLE

MFA Finland Meta Evaluation 2012-2014 Protocol for consultation with Selected MFA Stakeholders January 2015

The purpose of this meta-evaluation is to help the MFA improve the quality of evaluations, the evaluation management practices and the evaluation capacity development services. The meta-evaluation is also expected to bring forward issues and lessons learned to help the MFA improve its development co-operation. The work on this meta-evaluation has started in October 2014 and will be completed in March 2015. A first draft report has been submitted to present data on the quality of evaluations. The second phase of this mandate is ongoing.

The following protocol was developed to collect data on the MFA evaluation management practices and capacity development services.

Introduction

- Background on Meta Evaluation purpose and status (if needed)
- Purpose of this consultation
- Confidentiality
- Duration

During the Phase I review we noted that:

- TORs are sometimes based on a pre-set template without sufficient tailoring;
- Evaluation reports sometimes have severe gaps, but are approved anyways;
- Indicators or log frames in project descriptions are rare;
- Cross-cutting objectives are not systematically reflected in TOR and evaluation reports.
- Why?

Evaluation Capacity Building

We are interested in your views on the **strengths** and **areas for improvement**

related to the current evaluation capacity building support provided by MFA EVA-11 and Regional Advisors where appropriate with respect to:

- Types of support provided (in-person and virtual basic training, evaluation guides, tools, help desk, AHA guidance and templates, other support?)
- Relevance and appropriateness of support provided (is guidance provided up to date? Meeting the needs and issues you are familiar with in your experience with evaluations? Available in the right languages?)
- Adequacy of human/financial/technical resources allocated within MFA (EVA- 11 and Regional programs) to support evaluation capacity development
- Appropriateness of targeted audiences (desk officers in Helsinki, Directors/management; Regional advisors; Sector advisors; local staff overseas; evaluation consultants, others)
- Are there any institutional limitations to improving evaluation capacity (e.g., evaluation culture, coordination between EVA-11 and Regional programs, others)?
- What recommendations do you have to improve capacity building services provided by EVA-11 or others?

Evaluation Management Cycle

- Please review the current cycle - does this capture the key steps? Anything missing? Inaccurate? Is quality assurance clearly and adequately represented?
- Referring to the attached diagram, please share your views on the strengths and areas for improvement in the current evaluation management cycle in place in MFA.
- In your opinion, are roles and responsibilities among MFA stakeholders for managing evaluations in MFA:
 - Clear?
 - Appropriate?
 - Followed?
- In your opinion, are the roles and responsibilities, mechanism, incentives for engaging MFA's developing country partners in evaluations:
 - Clear?
 - Appropriate?
 - Followed?
- What recommendations do you have to improve the evaluation management cycle? What implications do these have, if any for evaluation capacity development in MFA?

Other Feedback?

Do you have any other points that you would like to make at this point in time?

Thank you!

ANNEX 14 TRAINING OFFERED

Year	Name of training session	Number of participants	Institution/ location	Other information	Language ¹	Courses/ Participants
2010	1. Training in evaluation of development co-operation	22	MFA Finland	3-module basic training. Of total 33 participants 6 completed all modules	Finnish	1 course / 33 participants
	2. Managing evaluation process	18	Finland			
	3. Quality of evaluation report	15	Finland			
2011	1. Training in evaluation of development co-operation	23	18 Finland 5 embassies	3-module basic training. Of total 31 participants, 13 completed all modules	Finnish	3 courses / 94 participants
	2. Managing evaluation process	26	24 Finland 2 embassies			
	3. Quality of evaluation report	27	22 Finland 5 embassies			
	Training in evaluation of development co-operation + Development evaluation workshop	18	9 Finland 7 embassies 1 other	Two day training	English	
	1. Training in evaluation of development co-operation	14	11 Finland 3 embassies	Basic evaluation training	Finnish	
	2. Managing evaluation process	17	14 Finland, 3 embassies			
	3. Quality of evaluation report	14	9 Finland 5 Embassies			
2012	Evaluation training for evaluators	25	25 consultants	Two day training	Finnish	4 courses/ 106 participants
	Training in evaluation of development co-operation + Development evaluation workshop	12	9 MFA Finland 4 embassies 1 consultant	Two day training	English	

¹ Language of instruction has been specified in the last column, when available. However, it is typical to mix languages (Finnish and English) during the training and meetings.

Year	Name of training session	Number of participants	Institution/ location	Other information	Language ¹	Courses/ Participants
2012	Training on impact evaluation of development co-operation- Real World Evaluation (For MFA personnel and consultants)	69	10 MFA 59 Consultants	Two day training	English	4 courses/ 106 participants
	1. Evaluation of development co-operation -why, how and what? The basic concepts and principals of evaluation	Information not available		Basic two-day training	Finnish	
	2. Managing evaluation process -preparation, implementation and use of results	Information not available				
2013	KEO-40: The basic concepts and principals of evaluation	6-8 estimated	MFA Finland	Special training tailored for the UN unit (KEO-40)	Finnish	5 courses / 104 participants
	KEO-40: the roles and responsibilities in UN evaluation	6-8 estimated	MFA Finland		English	
	KEO- 40: webinar with UN evaluation colleagues on evaluation	6-8 estimated	MFA Finland			
	Training in evaluation of development co-operation	16	4 MFA Finland 6 embassies 6 consultants	Basic two-day training	English	
	What is changing in management of development co-operation projects and evaluation?	27		Training for the new manual 2013	Finnish	
	What is changing in management of development co-operation projects and evaluation?	29	14 MFA 15 consultants	Training for the new manual 2013	Finnish	
	What is changing in management of development co-operation projects and evaluation?	25	6 MFA 19 consultants	Training for the new manual 2013	English	
2014	1. Evaluation of development co-operation -why, how and what? The basic concepts principals of evaluation	11 + 2 trainers	9 MFA 2 Consultants		Finnish	3 courses/ 63 participants
	2. Managing evaluation process -preparation, implementation and use of results	14 + 2 trainers				

Year	Name of training session	Number of participants	Institution/ location	Other information	Language ¹	Courses/ Participants
2014	3. Evaluation of human rights based development co-operation	11	8 MFA 3 consultants			3 courses/ 63 participants
	1. KEO-30: Evaluation of development co-operation -why, how and what? The basic concepts principals of evaluation	10-12 (estimated)		Special training tailored to the NGO-unit (KEO-30)	Finnish	
	2. KEO-30: Managing evaluation process -preparation, implementation and use of result	10-12 (estimated)				
	Training in evaluation of development co-operation	16	3 MFA Finland 7 embassies 2 other min.		English	

EVALUATION

**META-EVALUATION OF PROJECT AND
PROGRAMME EVALUATIONS IN 2012-2014**



**MINISTRY FOR FOREIGN
AFFAIRS OF FINLAND**