Evaluation

Meta-Evaluation of Decentralised Evaluations in 2010 and 2011



Evaluation report 2012:8

MINISTRY FOR FOREIGN AFFAIRS OF FINLAND

REPORT 2012:8 Meta-Evaluation of Decentralised Evaluations in 2010 and 2011
ISBN: 978-952-281-079-3 (printed), ISBN: 978-952-281-080-9 (pdf), ISSN: 1235-7618

REPORT 2012:7 Finland's Contribution to Building Inclusive Peace in Nepal
ISBN: 978-952-281-029-8 (printed), ISBN: 978-952-281-030-4 (pdf), ISSN: 1235-7618

REPORT 2012:6 Nordic Influence in Multilateral Organizations: A Finnish Perspective
ISBN: 978-952-281-027-4 (printed), ISBN: 978-952-281-028-1 (pdf), ISSN: 1235-7618

REPORT 2012:5 Finnish support to development of local governance
ISBN: 978-952-281-002-1 (printed), ISBN: 978-952-281-003-8 (pdf), ISSN: 1235-7618

REPORT 2012:4 Finnish Concessional Aid Instrument
ISBN: 978-951-724-995-9 (printed), ISBN: 978-951-724-996-6 (pdf), ISSN: 1235-7618

SPECIAL EDITION 2012:1 POLICY BRIEF: Country Programmes between Finland and Nepal, Nicaragua and Tanzania
ISBN: 978-951-724-655-2 (printed), ISBN: 978-951-724-659-0 (pdf), ISSN: 1235-7618

REPORT 2012:3 Country Programme between Finland and Tanzania
ISBN: 978-951-724-993-5 (printed), ISBN: 978-951-724-994-2 (pdf), ISSN: 1235-7618

REPORT 2012:2 Country Programme between Finland and Nepal
ISBN: 978-951-724-987-4 (printed), ISBN: 978-951-724-988-1 (pdf), ISSN: 1235-7618

REPORT 2012:1 Country Programme between Finland and Nicaragua
ISBN: 978-951-724-983-6 (printed), ISBN: 978-951-724-984-3 (pdf), ISSN: 1235-7618

REPORT 2011:5 Junior Professional Officer Programme of Finland
ISBN: 978-951-724-966-9 (printed), ISBN: 978-951-724-967-6 (pdf), ISSN: 1235-7618

REPORT 2011:4 Finnish Aid for Trade
ISBN: 978-951-724-964-5 (printed), ISBN: 978-951-724-965-2 (pdf), ISSN: 1235-7618

REPORT 2011:3 VERIFIN Training Programme on Verification of Chemical Weapons
ISBN: 978-951-724-951-5 (printed), ISBN: 978-951-724-952-2 (pdf), ISSN: 1235-7618

REPORT 2011:2 Results-Based Approach in Finnish Development Cooperation
ISBN: 978-951-724-941-6 (printed), ISBN: 978-951-724-942-3 (pdf), ISSN: 1235-7618

REPORT 2011:1 Finnish Support to Energy Sector
ISBN: 978-951-724-894-5 (printed), ISBN: 978-951-724-895-2 (pdf), ISSN: 1235-7618

REPORT 2010:6 Agriculture in the Finnish Development Cooperation
ISBN: 978-951-724-896-9 (printed), ISBN: 978-951-724-897-6 (pdf), ISSN: 1235-7618

REPORT 2010:5/III Forestry Sector: Preliminary Study
ISBN: 978-951-724-880-8 (printed), ISBN: 978-951-724-881-5 (pdf), ISSN: 1235-7618

REPORT 2010:5/II Finnish Support to Forestry and Biological Resources. Country and Regional Reports (Parts 1 Kenya, 2 Mozambique (Eng, Por), 3 Tanzania, 4 Zambia, 5 Lao Peoples Democratic Republic, 6 Vietnam, 7 Western Balkans, 8 Central America)
ISBN: 978-951-724-878-5 (printed), ISBN: 978-951-724-879-2 (pdf), ISSN: 1235-7618

REPORT 2010:5/I Finnish Support to Forestry and Biological Resources
ISBN: 978-951-724-876-1 (printed), ISBN: 978-951-724-877-8 (pdf), ISSN: 1235-7618

REPORT 2010:4 Sustainability in Poverty Reduction: Synthesis
ISBN: 978-951-724-874-7 (printed), ISBN: 978-951-724-875-4 (pdf), ISSN: 1235-7618

REPORT 2010:3 The Finnish Development Cooperation in the Water Sector
ISBN: 978-951-724-848-8 (printed), ISBN: 978-951-724-849-5 (pdf), ISSN: 1235-7618

Evaluation

Meta-Evaluation of Decentralised Evaluations in 2010 and 2011

Evaluation report 2012:8

Evaluation

Meta-Evaluation of Decentralised Evaluations in 2010 and 2011

Svend Erik Sørensen
Casper Thulstrup

Evaluation report 2012:8

MINISTRY FOR FOREIGN AFFAIRS OF FINLAND

This evaluation was commissioned by the Ministry for Foreign Affairs of Finland.
The Consultant authors bear the sole responsibility for the presented views.
The report does not necessarily reflect the views of
the Ministry for Foreign Affairs of Finland.

© Ministry for Foreign Affairs of Finland 2012

This report can be accessed at http://formin.finland.fi/developmentpolicy/evaluations
Hard copies can be requested from EVA-11@formin.fi
or
Development Evaluation (EVA-11)
The Ministry for Foreign Affairs of Finland
P.O. Box 451
00023 GOVERNMENT
Finland

ISBN 978-952-281-079-3 (printed)
ISBN 978-952-281-080-9 (pdf)
ISSN 1235-7618
Cover photo: Martti Lintunen
Cover design: Anni Palotie
Layout: Taittopalvelu Yliveto Oy
Printing house: Kopijyvä Oy

Anyone reproducing the content or part of the content of the report should acknowledge the source. Proposed reference: Sørensen SE & Thulstrup C 2012 *Meta-Evaluation of Decentralised Evaluations in 2010 and 2011*. Evaluation report 2012:8. Ministry for Foreign Affairs of Finland, Kopijyvä Oy, Jyväskylä, 90 p. ISBN 978-952-281-079-3 (printed).

CONTENTS

PREFACE ix
ACRONYMS AND ABBREVIATIONS xi
ABSTRACT 1
Finnish 1
Swedish 2
English 3
SUMMARY 4
Finnish 4
Swedish 7
English 10
Summary of key findings, conclusions and recommendations 13

1 INTRODUCTION 19
1.1 Evaluation context 19
1.2 Evaluation process 20
1.3 Purpose and objectives 20
1.4 Sample 21
1.5 Methodology 23
1.6 Quality management 24
1.7 Limitations 25
2 QUALITY OF EVALUATION REPORTS 26
2.1 Compliance with DAC/EU Quality Standards 26
2.2 Summary of key findings 35
3 QUALITY OF TERMS OF REFERENCE 36
3.1 Compliance with selected criteria 36
3.2 Factors related to quality reporting 38
3.3 Summary of key finding 41
4 QUALITY OF DEVELOPMENT COOPERATION 42
4.1 Overall assessment 42
4.2 Relevance 43
4.3 Coherence 46
4.4 Complementarity 47
4.5 Effectiveness 52
4.6 Efficiency 54
4.7 Sustainability 54
4.8 Impact 57
4.9 Operationalisation of poverty reduction and cross-cutting objectives . . 59
4.10 Summary of key findings 66
5 USE OF EVALUATIONS – ANALYSIS OF TEN SELECTED PROJECTS 66
5.1 Use of evaluations 68
5.2 Project design 69

5.3 Indicators . . . 70
5.4 Cross-cutting objectives . . . 71
5.5 Baseline . . . 71
5.6 Monitoring . . . 72
5.7 Summary of key findings . . . 74
6 TRENDS . . . 74
6.1 Quality of reports . . . 74
6.2 Terms of reference . . . 75
6.3 Quality of development cooperation . . . 76
6.4 Cross-cutting objectives . . . 77
6.5 Paris Declaration principles . . . 77
6.6 Use of evaluations, results-based management and risk management . . . 78
6.7 Summary of key findings . . . 79
7 CONCLUSIONS AND LESSONS LEARNED . . . 79
8 RECOMMENDATIONS . . . 82

REFERENCES . . . 85
THE EVALUATION TEAM . . . 84

ANNEX 1 TERMS OF REFERENCE . . . 85
ANNEX 2 LIST OF REPORTS . . . 102
ANNEX 3 DETAILED METHODOLOGY [Includes: Table 1 Definition of evaluation criteria as per OECD/DAC and the European Commission; Table 2 Scoring system for the quality of evaluation reports applying the DAC/EU Quality Standards; Table 3 Scoring system for the quality of Finnish development cooperation] . . . 107

TABLES

Table 1 Quality of reports measured against the DAC/EU Quality Standards . . . 27
Table 2 Number of criteria "addressed" per type of report for the Overarching consideration phase of the DAC/EU Quality Standards . . . 27
Table 3 Number of criteria "addressed" per type of report for the Purpose, planning and design phase of the DAC/EU Quality Standards . . . 29
Table 4 Number of criteria "addressed" per type of report for the Implementation and reporting phase of the DAC/EU Quality Standards . . . 33
Table 5 Ranking of the ten best scoring reports on quality . . . 39
Table 6 Ranking of the ten worst scoring reports on quality . . . 40
Table 7 Ranking of DAC/EU evaluation criteria based on assessments of all evaluation reports . . . 42
Table 8 Assessment of relevance by report type . . . 44
Table 9 Assessment of coherence by report type . . . 46

Table 10 Assessment of complementarity by report type 48
Table 11 Assessment of effectiveness by report type 50
Table 12 Assessment of efficiency by report type 52
Table 13 Assessment of sustainability by report type 55
Table 14 Assessment of impact by report type 57
Table 15 List of ten selected projects for the analysis of use of evaluation 67
Table 16 Five examples of logframe structures from ten selected projects 70
Table 17 Overview of performance monitoring systems in ten selected projects 73

BOXES

Box 1 Definition of the three report types included in the meta-evaluation . . . 23
Box 2 Selected deficiencies in the design of selected projects limiting measurement towards effectiveness 51
Box 3 Impacts of projects claimed by two impact evaluation reports 58
Box 4 Two examples of the environment-poverty nexus 61
Box 5 Example of good practice of integrating CCOs in project design 64
Box 6 Brief considerations on addressing risks and assumptions in projects 78

FIGURES

Figure 1 Interconnectedness between three practices influencing development cooperation and policies 21
Figure 2 Reports by type 21
Figure 3 Reports by region 21
Figure 4 Reports by sector 22
Figure 5 Reports by long-term partner countries 22
Figure 6 Evaluation criteria by sectors 43
Figure 7 Evaluation criteria by regions 43
Figure 8 Quality of reports across the 2007 and 2009 meta-analyses and this meta-evaluation – data normalised 75

PREFACE

The purpose of the fifth meta-evaluation since 1991 was to analyse and draw lessons from the project evaluations of 2010 and 2011 to benefit development cooperation. The meta-evaluation compared the findings with two previous meta-analysis and two other evaluations. This meta-evaluation used the OECD/DAC and EU quality standards and many cross-cutting type objectives as criteria, which were not used in the earlier meta-evaluations.

The Ministry for Foreign Affairs of Finland has commissioned meta-evaluations more regularly since 2007. The meta-evaluations have been useful, for example Helpdesk service and trainings on evaluations were organised as a result of the Meta-Analysis of 2009.

The novelty in this meta-evaluation was to study projects of the ten evaluation reports included in the sample. This study was based on other documentation, project documents, annual reports and minutes of meetings. The purpose of this additional examination was to learn of the use of evaluation results, as well as how poverty reduction and cross-cutting objectives were considered in project planning and monitoring.

The current meta-evaluation showed that the quality of the Terms of References (TORs) and evaluation reports had improved. However, it did not confirm the earlier findings that high quality TORs would predict high quality evaluation reports. Thus, other factors must be influencing the quality of reports. Many findings of the quality of development cooperation, as depicted in the evaluation reports, were similar to earlier meta-evaluations. Moreover, improvements are needed on results-based management, project planning, baseline studies, indicators, monitoring and in promoting cross-cutting objectives.

This meta-evaluation identified for example the need for system-wide frameworks in quality assurance and innovative approaches and tools for project design. It also emphasised the need to include budgets for cross-cutting activities and the importance to identify the real needs of beneficiaries.

Helsinki, 31.12.2012

Aira Päivöke
Director
Development Evaluation

ACRONYMS AND ABBREVIATIONS

%	per cent
AHA	MFA case management system
CCOs	Cross-cutting objectives
DAC	Development Assistance Committee (of OECD)
EU	European Union
EUR	€, Currency of the EU
EVA-11	Development Evaluation of MFA
ICT	Information, Communication and Technology
MDGs	Millennium Development Goals
MFA	Ministry for Foreign Affairs of Finland
MTR	Mid-Term Review
OECD	Organisation for Economic Co-operation and Development
SMART	Specific, Measurable, Attainable, Realistic and Timely criteria
TOR	Terms of Reference
UNDP	United Nations Development Programme

Other acronyms and abbreviations are explained in the context they are used.

Metaevaluointi vuosien 2010 ja 2011 hajautetuista evaluoinneista

Svend Erik Sørensen ja Casper Thulstrup

Suomen ulkoasiainministeriön evaluointiraportti 2012:8

ISBN 978-952-281-079-3 (painettu); ISBN 978-952-281-080-9 (pdf);
ISSN 1235-7618

Koko raportti on luettavissa osoitteessa http://formin.finland.fi

---

TIIVISTELMÄ

Metaevaluoinnin tarkoituksena on hyödyntää hajautettujen evaluointien tuloksia kehitysyhteistyön toteutuksessa. Metaevaluoinnissa tutkittiin 41 hajautettua evaluointiraporttia ja niiden tehtävänkuvaukset. Ne sisälsivät 10 ennakkoarviointia, 19 väliarviointia ja 12 evaluointia ja niiden tehtävänkuvaukset. Tulokset osoittivat, että raportoinnin laatu oli hieman parantunut. Tehtävänkuvaukset olivat hyvin suunniteltuja ja sisälsivät oleelliset arviointikriteerit. Läpileikkaavien tavoitteiden kohdalla oli havaittavissa huomattavaa parannusta sekä tehtävänkuvauksissa että raporteissa. Myös köyhyydenvähentämistavoitteet oli otettu jossain määrin huomioon. Tehtävänkuvauksen laadun ja raportoinnin laadun välillä ei havaittu riippuvuussuhdetta. Kehitysyhteistyön laatu oli yleisesti ottaen heikkoa eikä osoittanut viitteitä kestävistä tuloksista. Projektien suunnittelu ja tulosjohtaminen olivat riittämättömiä. Niiden perusteella ei yleisesti ottaen pystynyt seuraamaan hankkeen edistymistä, mittaamaan tuloksia tai analysoimaan riskejä. Hankkeiden evaluointituloksia ei systemaattisesti toimeenpantu.

Tämä metaevaluointi tarjoaa parannusehdotuksia näihin puutteisiin. Ehdotukset sisältävät a) ottaa käyttöön johdonmukaisen laadunvarmistusjärjestelmän, mikä keskittyy hankkeen suunnitteluun, tulosperustaisuuteen, riskientunnistamiseen ja raportointiin, sekä b) kehittää yksinkertaisia ja innovatiivisia lähestymistapoja ja työkaluja hankkeiden suunnitteluun ja tulosjohtamiseen. Lisäksi raportoinnin hyväksymiskäytäntöjä tulisi parantaa; työkaluja hyödynsaajien todellisten tarpeiden tunnistamisen tulisi käyttää; horisontaalista politiikkajohdonmukaisuutta valtavirtaistamisen tehostamiseksi tulisi vahvistaa; ja läpileikkaavien aktiviteettien budjettien tulisi sisältyä hankkeisiin. Lisäksi konsulttien raportoinnin laadunvarmistustaitoja sekä hankesuunnittelutaitoja tulisi parantaa. Lopuksi keskeiset tekijät jotka parantavat tai heikentävät laadukasta raportointia tulisi identifioida.

Avainsanat: metaevaluointi, projektien evaluoinnit, tulosjohtaminen, laadunvarmistus, Suomen kehitysyhteistyö

Metautvärdering av decentraliserad utvärderingar under 2010 och 2011

Svend Erik Sørensen och Casper Thulstrup

Utvärderingsrapport for Utrikesministeriet i Finland 2012:8

ISBN 978-952-281-079-3 (tryckt); ISBN 978-952-281-080-9 (pdf);
ISSN 1235-7618

Rapporten finns i sin helhet på adressen http://formin.finland.fi

---

ABSTRAKT

Syftet med denna metautvärdering var att dra lärdom av de decentraliserade utvärderingarna som skulle gynna utvecklingssamarbetet. Metautvärderingen bestod av en skrivbordstudie som inkluderade 41 decentraliserade rapporter samt tillhörande uppdragsbeskrivningar. Tio granskningar, 19 halvtidsutvärderingar och 12 utvärderingar analyserades. Resultaten visade på en knapp förbättring av kvaliteten i rapporteringen. Uppdragsbeskrivningarna var korrekt utformade och innehöll relevanta utvärderingskriterier. De tvärgående målen var betydligt bättre adresserade i rapporterna och uppdragsbeskrivningarna. Aspekter relaterade till fattigdomsbekämpning var, till viss del, adresserade i rapporterna och uppdragsbeskrivningarna. Något samband mellan god kvalitet på uppdragsbeskrivningen och god kvalitet på rapporteringen kunde inte urskiljas. Kvaliteten på utvecklingssamarbetet var överlag låg och visade föga effekt för att nå hållbara resultat. Även projektutformningen och resultatstyrningen (RBM) var otillräckligt utvecklade och de gav generellt sett ingen möjlighet att följa upp processer, mäta prestationer och analysera risker. Det fanns inte heller någon systematisk uppföljning av utvärderingsresultat i projekten.

I metautvärderingen föreslås förbättringar till ovan nämnda brister: dessa inkluderar (a) tillämpning av ett sammanhängande och systemtäckande ramverk för kvalitetssäkring med fokus på projektutformning, strategibaserade resultat, risker och rapportering, och (b) utveckling av förenklade och nyskapande lösningar och verktyg för projektutformning och resultatstyrning (RBM). Vidare bör godkännandeförfarandet för rapportering förbättras. Verktyg för att identifiera verkliga behov hos projektens förmånstagare bör även tillämpas, horisontell policykoherens för effektiv samstämmighet (*mainstreaming*) bör stärkas, och en budget för tvärgående aktivitetet bör genomgående inkluderas i projekten. Även konsulters kvalitetssäkring av rapporter och projektutformning bör stärkas. Till sist bör viktiga faktorer som bidrar till eller hindrar rapportering av god kvalitet identifieras.

Nyckelord: metautvärdering, projektbaserad utvärdering, resultatstyrning (RBM), kvalitetssäkring, Finlands utvecklings samarbete

Meta-evaluation of decentralised evaluations in 2010 and 2011

Svend Erik Sørensen and Casper Thulstrup

Evaluation report of the Ministry for Foreign Affairs of Finland 2012:8

ISBN 978-952-281-079-3 (printed); ISBN 978-952-281-080-9 (pdf);
ISSN 1235-7618

The full report can be accessed at http://formin.finland.fi

ABSTRACT

The purpose of the meta-evaluation was to draw lessons from the decentralised evaluations to benefit development cooperation. The meta-evaluation was a desk study only, including 41 decentralised reports and their Terms of Reference (TORs). Ten appraisals, 19 mid-term reviews and 12 evaluations were analysed. Findings showed a slight improvement in the quality of reporting. TORs were well designed and included relevant evaluation criteria. There was a significant improvement in addressing the cross-cutting objectives in TORs and reports. Poverty reduction issues were addressed to some extent in the reports and TORs. Correlation between high quality TORs and high quality reporting was not found. Quality of development cooperation was overall poor showing little effect towards sustained outcomes. Project design was generally unable to monitor progress, measure achievements and analyse risks. Early quality assessment of project design was missing. There was no systematic follow-up on evaluation results in projects.

The meta-evaluation suggests improvements to the above shortcomings: (a) applying a coherent and system-wide framework for quality assurance with particular focus on project design, results-based approach, risks and reporting, and (b) developing simplified and innovative approaches and tools for project design and results-based management. Furthermore, approval procedures for reporting should be improved. Also, tools for identifying real needs of beneficiaries in projects should be applied; horizontal policy coherence for effective mainstreaming should be strengthened; and budgets for cross-cutting activities included in projects. In addition, consultants' quality assurance on reporting and their project design skills should be improved. Finally, key factors that contribute to or impede quality reporting should be identified.

Keywords: meta-evaluation, project-based evaluation, results-based management, quality assurance, Finland's development cooperation

YHTEENVETO

Metaevaluoinnin päätarkoitus on hyödyntää hajautettujen evaluointien tuloksia kehitysyhteistyön toteutuksessa. Metaevaluointi sisälsi 41 hajautettua evaluointiporttia sekä niiden tehtävänkuvaukset. Raportteihin kuului 10 ennakkoarviointia, 19 väliarviointia ja 12 evaluointia.

Raporttien ja tehtävänkuvausten laatu
Suurin osa raporteista oli hyvin kirjoitettuja ja ymmärrettäviä. Ne sisälsivät verrattain selkeän analyysin ja viittauksen tehtävänkuvaukseen. Puolta 41 raportista voidaan pitää laadultaan joko hyvinä (good) tai tyydyttävinä (adequate); loppuja voidaan pitää laadultaan heikkoina (poor). Suurin osa raporteista oli sisällöltään puutteellisia. Esimerkiksi niissä oli:

- puutteellinen kuvaus evaluoinnin kulusta, erityisesti liittyen hyödynsaajien ja sidosryhmien osallistumiseen (evaluoinnin suunnitteluun, henkilöstön koulutukseen ja yhteisevaluointeihin) sekä laadunvarmistusjärjestelmien käyttöön;
- riittämätön yhteenveto sekä kontekstin ja loogisen viitekehyksen/tuloskehyksen kuvaus;
- pinnallinen ja usein sekava DAC:n (OECD:n kehitysyhteistyökomitea) ja Euroopan Unionin evaluointikriteerien käyttö.

Havaittuihin heikkouksiin tulee puuttua noudattamalla paremmin DAC/EU laatustandardeja ja kouluttautumalla niissä. Vuoden 2009 meta-analyysiin verrattuna nyt havaittiin lievä parannus raportoinnin laadussa. Raportoinnin heikkouksiin tulee kuitenkin puuttua tehostamalla Ulkoasianministeriön ja konsulttien laadunvarmistusjärjestelmiä sekä vahvistamalla Ulkoasianministeriön raporttien hyväksymismenettelyä.

Tehtävänkuvaukset olivat tyydyttävästi laadittuja. Lähes kaikki sisälsivät oleelliset DAC evaluointikriteerit, mutta evaluointikysymyksiä oli usein liian paljon ja niitä ei ollut laitettu tärkeysjärjestykseen. Vaikka läpileikkaavia tavoitteita käsiteltiin tehtävänkuvauksissa, niiden tärkeänä pitäminen puuttui. Silti voidaan todeta, että läpileikkaavien tavoitteiden osalta on tapahtunut huomattavaa parannusta vuoden 2007 meta-analyysiin verrattuna, sillä tuolloin niitä harvoin mainittiin tehtävänkuvauksissa. Köyhyyden vähentäminen mainittiin hieman alle puolessa tehtävänkuvista, kun taas Pariisin julistuksen periaatteita käsiteltiin pääasiassa muun kriteeristön keinoin (esimerkiksi tuloksellisuus). Tehtävänkuvausten ohjeistusta tulee tarkistaa sekä evaluointikriteereitä ja -kysymyksiä laittaa niissä tärkeysjärjestykseen.

Sekä 2007 että 2009 meta-analyysit osoittivat laadukkaan tehtävänkuvauksen ja laadukkaan raportoinnin välillä olevan vahvan riippuvuussuhteen. Samaa ei havaittu tässä metaevaluoinnissa. Muut tekijät vaikuttavat raportoinnin laatuun, kuten:

- ulkoasianministeriön yleinen institutionaalinen tilanne ml. henkilöstön taidot, kuormitus ja vaihtuvuus;
- poliittisten linjausten, käytännön ohjeistusten ja työkalujen laatu sekä niiden käyttö hankkeiden suunnittelussa ja toteutuksessa;
- hajautettujen evaluointiraporttien hyväksymismenettely;
- evaluointitiimi sekä sen sisältö- ja raportointitaidot.

Kehitysyhteistyön laatu
Kehitysyhteistyön laatu oli yleisesti ottaen heikkoa ja osoitti vähän viitteitä kestävistä tuloksista. Tarkoituksenmukaisuus (relevance) sai korkeimmat pisteet, johdonmukaisuus (coherence) toiseksi korkeimmat ja täydentävyys (complementarity) seurasi lähellä perässä. Sen jälkeen tulivat matalimmat pisteet saaneet tuloksellisuus (effectiveness), tehokkuus (efficiency), kestävyys (sustainability) ja vaikuttavuus (impact).

Heikon hanketoteutuksen ja heikkojen tulosten (tehokkuus ja tuloksellisuus) välillä havaittiin vahva riippuvuussuhde. Lisäksi tulokset hankkeiden vaikuttavuudessa ja kestävyydessä eivät olleet toivotun kaltaisia. Nämä varmistavat tarpeen:

- parantaa tuloskehyksen laadunvarmistusta ja
- etsiä uusia ja innovatiivisia työkaluja seuraamaan hankkeiden edistymistä ja tuloksien saavuttamista.

Metaevaluoinnissa nousi esiin useita tärkeitä osa-alueita, joiden avulla on mahdollista parantaa kehitysyhteistyön tämänhetkistä tilaa. Näitä ovat:

- hankesuunnittelun yksinkertaistaminen ja loogisen viitekehyksen/tuloskehyksen käyttö;
- tarkoituksenmukaisuuden (relevance) ja köyhyysvaikutuksien vahvistamiseksi hyödynsaajien voimakas mukaan ottaminen;
- keskittyminen horisontaaliseen politiikkajohdonmukaisuuteen;
- projektin suunnittelun laadunvarmistus projektin identifioinnin ja suunnittelun aikana kehitysyhteistyön parempaa arviointia varten;
- yksinkertaistettu ja systemaattinen riskianalyysi tukemaan paremmin hankkeen kestävyyttä ja tuloksia;
- sellaisten hyödynsaajien tukeminen, joilla on vahva oma tarveagenda;
- läpileikkaavien tavoitteiden budjetointi ja sellaisten sektorien tukeminen, jotka ovat tuloksellisia köyhyyden vähentämisessä.

Köyhyyden vähentäminen ja läpileikkaavat tavoitteet
Köyhyyden vähentäminen mainittiin puolessa raporteista. Tehokkaimmin köyhyyden vähentämistä oli lähestytty maaseudunkehitys-, vesi- ja sanitaatio-, metsä- sekä ympäristösektoreilla. Köyhyyden vähentämisen kolmea ulottuvuutta (yhteiskunnallinen, taloudellinen ja luonnontaloudellinen) ei ollut sovellettu usein.

Keskeisten hyödynsaajaryhmien kokonaisvaltainen osallistuminen oli puutteellista ml. köyhimmät ja helpoiten syrjäytyvät. Tehokkaan köyhyyden vähentämisen kannalta on tärkeää, että hyödynsaajien tarpeet tunnistetaan paremmin.

Tässä metaevaluoinnissa näkyi, että läpileikkaavat tavoitteet olivat osa raportointia, mutta niitä ei systemaattisesti edistetty. Yhtenä poikkeuksena oli sukupuolten tasa-arvo. Demokratiakehitystä, ihmisoikeuksia ja ilmastonmuutosta käsiteltiin erittäin vähän. Merkittävää parannusta oli kuitenkin tapahtunut läpileikkaavien tavoitteiden käsittelyssä vuoden 2007 meta-analyysiin verrattuna.

Kehitysyhteistyön onnistumisen kannalta keskeistä riskianalyysiä ei ollut käytetty tarpeeksi. Riskien kartoituksen ja hallinnan tulisi olla hankesyklin joka vaiheessa keskiössä ja erityisenä prioriteettina hankesuunnittelussa.

Pariisin julistuksen periaatteet ja vuosituhattavoitteet olivat harvoin näkyvissä hankkeissa. Suomalaista lisäarvoa ei pystynyt arvioimaan, sillä raporteissa ei ollut tarpeeksi tietoa siitä.

Evaluointiraporttien käyttö

Tarkemman tarkastelun kohteena oli kymmenen hanketta. Niistä tarkasteltiin sitä, miten evaluointiraporttien suosituksia oli käytetty projektisuunnitelmien tai projektien jatkon valmistelemisessa. Evaluointien tulosten käyttöä ilmeni, mutta se ei ollut systemaattista. Loogiset viitekehykset olivat heikkoja ja indikaattorit sekä lähtökohta-aineisto puuttuivat usein. Läpileikkaavat tavoitteet eivät näkyneet budjetoinnissa. Vaikka sukupuolten tasa-arvo oli hankesuunnitelmassa, se ei kuitenkaan näkynyt tuloksissa. Hankkeiden seuranta ja evaluointi olivat heikkoja. Suosituksena on kehittää standardimalli hajautettujen evaluointien tulosten toimeenpanoa varten (management response) ulkoasiainministeriössä.

Opetukset ja kokemukset

Organisaation institutionaalista ja henkilöresurssien mallia täytyy katsoa laajemmassa viitekehyksessä, jotta ymmärretään laadukkaan raportoinnin, tehtävänkuvausten ja evaluointien hyödyntämisen yhteys. Vain tällä tavalla voidaan tehokkaasti tukea kehitysyhteistyötä.

Kehitysyhteistyön tulosten evaluointi tulee jäämään heikoksi, jollei tämän metaevaluoinnin suosituksia huomioida tosissaan. Onnistumisen edellytys on, että ulkoasianministeriön johto on halukas ja kyvykäs tukemaan tällaista prosessia.

SAMMANFATTNING

Syftet med denna metautvärdering var att dra lärdom av det finska utrikesministeriets (MFA) decentraliserade utvärderingar. Metautvärderingen bestod av en skrivbordstudie som inkluderade 41 decentraliserade rapporter med tillhörande uppdragsbeskrivningar, däribland tio granskningar, 19 halvtidsutvärderingar och 12 utvärderingar.

Kvalitet på rapporter och uppdragsbeskrivningar
De flesta rapporter ansågs välskrivna och lättförståeliga. I de flesta fall inbegriper de en relativt tydlig analys och hänvisning till uppdragsbeskrivningen. Hälften av de 41 rapporterna kan anses vara av antingen god eller tillfredsställande kvalitet, den andra hälften av otillräcklig kvalitet. De flesta rapporterna saknade tillräcklig information om utvärderingsprocessen och var ofullständiga i fråga om innehåll. Exempel:

- otillräcklig beskrivning av utvärderingshanteringen, speciellt relaterat till involvering av olika parter (utvärderingsplanering, personalutbildning, och gemensamma utvärderingar) samt användningen av kvalitetssäkringssystem;
- otillräcklig redogörelse av sammanfattningar, innehåll och logframe/resultatmodell (verksamhetslogik);
- ytligt och ofta förvirrande användning av DAC/EU-utvärderingskriterier.

Utifrån dessa identifierade svagheter kan slutsatsen dras att förståelsen för och användandet av DAC/EU-kvalitetsnormer för utvärderingsrapporter måste förbättras genom träning/utbildning. En mindre förbättring i kvaliteten på rapporteringen jämfört med 2009 års metautvärdering har dock kunnat urskiljas. De konstaterade bristerna i rapporteringen ger vid handen att en mer effektiv kvalitetssäkring av det finska utrikesministeriet och av konsulterna bör ske, liksom att utrikesministeriets godkännandeförfarande av rapporter bör förbättras.

Uppdragsbeskrivningarna var utformade på ett tillfredställande sätt. Nästan alla uppdragsbeskrivningar inkluderade relevanta DAC-utvärderingskriterier, men i flera av uppdragsbeskrivningarna var antalet utvärderingsfrågor för många och utan prioritetsordning. De tvärgående målsättningarna hade behandlats men de var inte prioriterade. Dock kunde en avsevärd förbättring i behandlingen av de tvärgående målsättningarna konstateras i jämförelse med 2007 års metautvärdering, då de tvärgående målsättningarna sällan presenterades i uppdragsbeskrivningen. Fattigdomsbekämpning nämndes direkt i färre än hälften av uppdragsbeskrivningarna. Parisdeklarationens principer behandlades mestadels genom andra kriterier, som t.ex. effektivitet (måluppfyllelse). Det föreslås därför att man granskar den nuvarande uppdragsbeskrivningsguiden, och att utvärderingskriterier och utvärderingsfrågor prioriteras i utformningen av uppdragsbeskrivningar.

Enligt tidigare metautvärderingar finns det ett samband mellan god kvalitet på uppdragsbeskrivningen och god kvalitet på rapporteringen. Sådana samband kunde inte

konstateras i denna metautvärdering. Andra faktorer än uppdragsbeskrivningen anses viktigare för att skapa god kvalitet på rapporteringen. Dessa inkluderar:

- den övergripande institutionella kontexten vid det finska utrikesdepartementet, däribland personalens kunskap, tid och personalomsättning;
- kvaliteten och användandet av policy samt praktiska guider och verktyg för effektiv projektutformning och implementering;
- godkännandeprocessen för rapporteringen kring de decentraliserade utvärderingarna;
- kvaliteten på utvärderingsteamet och dess tekniska- och rapporteringskompetens.

Kvalitet på utvecklingssamarbetet
Kvaliteten på utvecklingssamarbetet ansågs vara låg och visade endast ha liten effekt för hållbara resultat. Relevans fick högst poäng, koherens näst högst poäng, följt av komplementaritet och till sist effektivitet, kostnadseffektivitet, hållbarhet och långsiktliga effekter.

Den relativt starka korrelationen mellan svag projektimplementering och resultat [kostnadseffektivitet och effektivitet (måluppfyllelse)] samt bristen på förändringseffekter (hållbarhet och impact/långsiktliga effekter) bekräftar behovet av:

- ett starkt fokus på stärkt kvalitetsgranskning genom effektiv utformning av resultatramverk;
- ett starkt behov att identifiera nya och innovativa verktyg för processuppföljning och uppföljning av uppnådda resultat.

I metautvärderingen identifierades ett flertal områden som skulle kunna förbättra rådande status av projekt i utvecklingssamarbetet. Dessa inbegriper:

- förenkling av projektdesign samt användandet av logframe/resultatramverk;
- ökad involvering av mottagare, för ökad relevans;
- fokuserade ansträngningar för horisontell policykoherens;
- tidig kvalitetsgranskning för förbättrad bedömning av utvecklingssamarbetet;
- tillförsäkring av hållbara projektresultat genom förenklade och mer systematiska riskanalyser;
- stöd till mottagargrupper som sätter en tydlig efterfrågestyrd agenda;
- budgetering för att uppnå tvärgående målsättningar och stödja sektorer som uppnår effektivitet vad gäller fattigdomsbekämpning.

Fattigdomsbekämpning och tvärgående målsättningar
Fattigdomsbekämpning nämndes direkt i mindre än hälften av rapporterna. Mest effektivt var fattigdomsbegränsning adresserat när det gäller landsbygdsutveckling, vattenförsörjning och -rening, samt skog och miljö. De tre dimensionerna av fattigdomsbegränsning – social, ekonomisk och ekologisk – applicerades sällan.

Involvering av huvudmottagare, inklusive de fattigaste och mest utsatta grupperna, var svag. Användandet av verktyg för att tydligt identifiera uppfattade behov hos mottagarna är viktigt för en effektiv fattigdomsminskning.

I denna metautvärdering utgjorde tvärgående målsättningar/ämnesöverskridande mål (*cross-cutting objectives*) en integrerad del av rapporteringen men dessa applicerades inte på ett systematiskt vis. Ett undantag var jämställdhet. Demokratisk samhällsutveckling, mänskliga rättigheter och klimatförändringar adresserades endast ett fåtal gånger. Jämfört med resultaten från metaanalysen 2007 kunde dock en avsevärd förbättring vad gäller adressering av tvärgående målsättningar konstateras.

Trots att riskanalyser är en viktig förutsättning för ett lyckat utvecklingssamarbete kan man konstatera att sådana inte adresserats på ett adekvat sätt. Identifiering och hantering av risker bör ske på varje projektnivå och prioriteras högt i utformningen av projekt.

Parisdeklarationens principer och Millenniummålen (MDGs) förekom sällan i projekten och finska mervärden kunde inte adresseras på grund av brist på data.

Användning av utvärderingsrapporter
Rekommendationer från tio av de undersökta rapporterna användes i beredningen av projektförslag, om än dock inte på ett systematiskt sätt. Svaga resultatramverk samt avsaknad av indikatorer och *baseline*-data dominerade rapporterna. Tvärgående målsättningar och aktiviteter reflekterades inte i budgeten. Medan jämställdhet ofta adresserades i projektdokument var det inte synligt i outputs. Uppföljning och utvärdering av projekten var svag. Det fanns ingen effektiv uppföljning av förbättringar i projekt. Ett standardiserat format för utrikesdepartementets *management response* föreslås därför.

Lärdomar
Ett bredare beaktande av såväl institutionella som personliga mönster i organisationerna är nödvändigt för att förstå den ömsesidiga kopplingen mellan rapportkvalitet, uppdragsbeskrivningar och användandet av utvärdering som stöd i ett effektivt utvecklingssamarbete.

Mätningen av utvecklingssamarbetets resultat kommer även fortsättningsvis vara undermålig om rekommendationerna i denna metautvärdering inte tas på allvar. Vilja och kapacitet i Utrikesdepartementets ledarskap att möjliggöra denna process är en förutsättning för förbättringar.

SUMMARY

The purpose of the meta-evaluation was to draw lessons from decentralised evaluations of the Ministry for Foreign Affairs of Finland (MFA). The meta-evaluation was a desk study only of 41 decentralised reports and their Terms of References (TORs), covering ten appraisals, 19 mid-term reviews and 12 evaluations.

Quality of reports and terms of reference
Most reports were well written and understood. They often included a relatively clear analysis and reference to TOR. Half of the 41 reports could be considered of either good or adequate quality, the other half of inadequate quality. A majority of the reports lacked sufficient information on the evaluation process and were incomplete in terms of content. This included:

- inadequate description of evaluation management, particular related to inclusiveness of stakeholders (evaluation planning, training of staff, and joint evaluations) and use of quality control systems;
- inadequate description of summaries, context and logframes/results frameworks;
- superficial and often confusing use of the Development Assistance Committee (DAC) and European Union (EU) evaluation criteria.

The identified weaknesses call for improved compliance with the DAC/EU Quality Standards for evaluation reporting through skills training. A small trend towards improvements in the quality of reporting was observed compared with the 2009 Meta-analysis. Deficiencies identified in reporting suggest a need for more effective quality assurance by MFA and consultants, and strengthened approval procedures for reporting by the MFA.

TORs were satisfactorily designed. Almost all TORs included relevant DAC evaluation criteria but in several TORs the number of evaluation questions was too many and not prioritised. While cross-cutting objectives (CCOs) were addressed they were not prioritised. Still there was a significant improvement in addressing the CCOs compared with the 2007 Meta-analysis, where CCOs were rarely presented in TORs. Poverty reduction was mentioned directly in less than half of the TORs. The Paris Declaration principles were addressed mainly through other criteria, e.g. effectiveness. It is suggested that current TOR guidelines are reviewed and evaluation criteria and evaluation questions prioritised in the design of TORs.

Previous meta-analyses carried out in 2007 and 2009 claimed correlation between high quality TORs and high quality reporting. Such correlation was not found in this meta-evaluation. Other factors than TORs seem important for the quality of an evaluation report, including:

- the overall institutional context of the MFA, including staffing 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.

Quality of development cooperation
The quality of the development cooperation was considered to be poor showing little effect towards sustained outcomes. Relevance received the highest score, coherence second highest, followed by complementarity and a group ranked lower, comprising effectiveness, efficiency, sustainability and impact.

The relatively strong correlation between weak project implementation and results (efficiency and effectiveness) and the lack of effects towards change (sustainability and impact) confirms the strong need to:

- improve quality assurance through effective design of results framework and
- identify new and innovative tools for progress monitoring and results achievement.

The meta-evaluation identified several areas of importance that would enhance the current state of projects in development cooperation. These include:

- simplifying project design and use of logframe/results framework;
- stronger involvement of beneficiaries for improved relevance and better addressing poverty;
- focused efforts on horizontal policy coherence;
- early quality assurance of project design – during project identification and formulation – for improved assessment of development cooperation;
- sustaining project benefits and results better through simplified and systematic risk analysis;
- supporting dedicatedly beneficiary groups that set a strong demand-driven agenda;
- budget for CCO activities and support sectors that achieve strong effectiveness towards poverty reduction.

Poverty reduction and cross-cutting objectives
Poverty reduction was directly mentioned in less than half the reports. Poverty reduction was most effectively addressed in rural development, water supply and sanitation, forestry and environment. The three dimensions of poverty reduction: social, economic and environmental, were not often applied.

Full involvement of key beneficiary groups lacked, including the poorest and most vulnerable. Applying tools for clearly identifying perceived needs of the beneficiaries is important for successfully alleviating poverty.

In this meta-evaluation CCOs comprised an integrated part of the reporting but were not systematically applied. One exception was gender equality. Democratic development, human rights and climate change were hardly addressed. Compared with 2007 Meta-analysis there was however a significant improvement in addressing the CCOs.

Risk analysis was inadequately addressed though crucial for development cooperation to be successful. Identifying and managing risks should be targeted at every level of the project and receive high priority in project design.

The Paris Declaration principles and the Millennium Development Goals (MDGs) were rarely visible in projects, and Finnish value added could not be assessed, due to lack of data about it in the reports.

Use of evaluation reports

Recommendations from ten reports investigated were generally used in the preparation of project proposals or extensions, though not systematically. Weak logframes, lack of indicators and baseline data dominated the reports. CCOs were not reflected in the budget. While gender equality was often addressed in project documents, it was not visible in outputs. Monitoring and evaluation application in the projects was weak. There was no effective follow-up on improvements in projects. Standardised formatting for MFA management response of decentralised evaluations is suggested.

Lessons learned

A broader framework of institutional and human resources patterns across organisations is needed to understand the interconnectedness between the quality of reports, TORs and the use of evaluations. Only in this way can development cooperation be effectively supported.

Development cooperation results will remain poorly measured in years to come if recommendations from this meta-evaluation are not seriously considered. Willingness and ability of the MFA leadership to facilitate this process is a prerequisite for success.

Summary of findings, conclusions and recommendations

Findings	Conclusions	Recommendations
Strategic considerations		
The quality of development cooperation, results-based management and reporting was inadequate.	Finnish development cooperation results will remain deprived in years to come if recommendations from this and previous meta-evaluations are not seriously considered.	1 Establishment and maintenance of a **coherent and system-wide quality assurance** framework in the Ministry for Foreign Affairs (MFA) for improving quality in development cooperation – with a special focus on simplified project design, results-based approach, risks and reporting. 2 **Willingness of MFA leadership** is required to develop and maintain quality assurance. This must include testing and applying new, innovative approaches and tools – supported by institutional changes, human resources management and staff training.
Project design and results-based approach		
Reports showed that results frameworks of projects were generally weak, i.e. the definition of results targets at outcome and output level were inadequate, indicators not measurable and baseline studies absent. Progress reporting was primarily activity based.	Measurement of effectiveness, sustainability and impact is challenged by lack of a clear results chain, baseline and monitoring data which affect all levels of planning (MFA, partner country, projects). Therefore, managers and decision-makers cannot correct activities properly and learn for future activities.	3 Existing **tools and guides** on result-based management should be reviewed and improved, e.g. *Manual for Bilateral Programmes*. 4 **Simplification** of the results chain/logframe should be targeted on a **pilot** basis. This will enable MFA/project staff and consultants to test new, easy and time-saving

	Simplification of systems could provide an effective fast-track mechanism towards improvements in results-based management in the MFA. Simplification of the results-based tools will strengthen consultants' and MFA staff's skills in management of projects. Intervening at an early stage would allow for targeting better goals of poverty reduction and cross-cutting objectives (CCOs). This will raise the quality of appraisals, the project document and the measuring of progress and achievements.	tools for project monitoring and results achievements. Simplifications should include: phrase output narratives in generic terms, apply easy Quantity, Quality and Time indicators, strengthen risk mitigation by incorporate risk issues into project activities, etc. 5 MFA staff and consultants must strengthen their project **design skills** and MFA its **approval procedures**. 6 The decentralised evaluations under the responsibility of the regional departments and embassies should effectively deal with project design discrepancies, particularly during the **early stages** of the project cycle. The Quality Assurance Group must provide effective supervision of adherence to quality and approval procedures.
Relevance of projects was overall met. There was great variation in relations to the priorities and policies of the target group, recipient and donor. **Poverty concerns** and priorities of the beneficiaries were often insufficiently addressed.	If real needs and priorities of primary beneficiaries are not adequately targeted, poverty reduction cannot be realistically tackled. This reduces project relevance. In assessing relevance particular attention should be given to detecting any chang-	7 **Project beneficiaries**, particularly the poor and vulnerable and their institutions, should have absolute priority in project design. They should constitute the basis upon which policy priorities and country strategies are developed. 8 **Specific tools**, e.g. classic grounded theory and

In some projects objectives and purpose drifted. This caused relevance to alter with often negative consequences for project outputs and implementation.	es or drifting in the project objectives and purpose.	participatory appraisals, should be used during project identification and formulation to identify the **main concerns** and real needs of beneficiaries. 9 TORs for mid-term evaluations should always include an assessment of **changes** to objectives and the consequences hereof to project results chain/ logframe.
Policy **coherence** is addressed primarily vertically (i.e. international, regional, national and local levels), while horizontal policy coherence (e.g. across governmental departments) is almost not addressed in projects.	Mainstreaming policies, especially cross-cutting policies, do make sense only if horizontal policy coherence is addressed in projects.	10 **Horizontal policy coherence** must be strengthened in project design to support improved mainstreaming of policies. The Organisation for Economic Cooperation and Development (OECD) guidelines for policy coherence for development should be applied in this process.
Complementarity received a lower score in appraisals than in mid-term reviews and evaluations. This indicated that complementarity was inadequately addressed in project preparation and design.	The lack of attention paid to complementarity in appraisals is likely to backfire during project implementation. Parallel project structures and lack of coordination have previously demonstrated unsustained development.	11 Complementarity must be thoroughly addressed at the **early stages of the project design** to support harmonisation and avoid duplication.

Sustainability of projects is weakened when project designs are based on subsidisation, poor cost recovery and high operation and maintenance cost. Excessive dependence on Finnish donor contribution and imbalances where beneficiaries' needs are not sufficiently targeted also weaken sustainability. Reports show that participation and awareness are often seen as important properties of sustainability.	Unsustained project outcomes and outputs are the result of poor design and poor risk analysis. Participation and awareness should not on their own merits justify the initiation of a project or the continuation of on-going projects into a new phase.	12 MFA/project staff and consultants must undertake thorough **risk analysis** and avoid complacency when dealing with assumptions. Effective risks mitigation measures must be applied. Influential and determining factors that support sustainability of outputs and outcomes must be identified. 13 When justified for continuation of a project, participation, awareness and similar processes should be **clearly linked to the output and outcome levels**.
Reasonably strong **effectiveness** at the output level and to some extent the outcome level was found in projects in water and sanitation, rural development, community forestry and environment. Delivery of products and services had an immediate effect on the ability of the beneficiaries to improve their livelihoods.	A strong and more direct impact on poverty reduction is achieved through funding to projects in water and sanitation, rural development, community forestry and environment. Lessons learned and best practices from these sectors should be identified.	14 MFA should consider increasing its financial support to those sectors that prove to achieve **high effectiveness on reducing poverty**.
Cross-cutting objectives' (CCOs) activities are not reflected in budgets. Gender equality is only addressed in few project outputs.	CCOs are less likely to be given priority if not addressed through specific outputs.	15 Prioritised CCOs should be incorporated into all projects, at outcome and output levels, with a designated **budget line**.

Recommendations from the mid-term reviews and evaluation reports are generally discussed and addressed in preparation of subsequent project phases. However, there is no systematic approach for **follow-up on recommendations**.	Recommendations from mid-term reviews and evaluations should be addressed systematically in a format that allows for follow-up on status and clarity of measures taken to address recommendations. Only in this manner is value for money guaranteed and learning for future interventions or phases enhanced significantly.	16 Improved management tools for the use of evaluations should be developed. **Standardised** management response, follow-up on recommendations and back-reporting on the decentralised evaluations should be designed.
Quality of evaluation reporting		
The compliance with the DAC/EU Quality Standards for evaluation reporting is reduced by lack of information on the evaluation process and incompleteness in report contents. There is no formal approval process of decentralised evaluation reports in the MFA.	There is a gap between current decentralised evaluation practices and what is perceived as the best evaluation practice (=DAC/EU Quality Standards). The utility of the decentralised evaluation reports is not likely to happen without a formal approval process in place.	17 **Capacity building support** should be provided to MFA/project staff and consultants on how to comply with the DAC/EU Quality Standards. 18 MFA must review and improve its **approval and quality assurance procedures** for decentralised evaluation reporting.
There is no clear correlation identified between quality of TORs and quality of reports.	Identification of factors that influence high quality reporting is required for improving the design of TORs, evaluation reporting and assessing development cooperation.	19 **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.

Terms of reference		
TORs include relevant DAC evaluation criteria but often there are too many evaluation questions and these are not prioritised. CCOs, poverty reduction and Paris Declaration principles are addressed in TOR, yet not methodically.	TORs are satisfactorily designed with some flaws related to prioritisation and numbers of evaluation questions. Without systematically addressing important development policies in TOR the possibilities for meeting policy targets/statements are diminished.	20 Current TOR guidelines should be reviewed and clear instructions prepared on how to use evaluation criteria and questions. Depending on the scope and budget allotted for an appraisal, a mid-term review or an evaluation, focus should be on **prioritising** development and evaluation criteria and questions.
Meta-evaluation recommendations		
Improvements to the quality of development cooperation, quality of evaluation reporting and TORs, and use of evaluations have been identified.	Without proper management and administration of recommendations suggested from this meta-evaluation the MFA will lose the opportunity to strengthen development cooperation.	21 Recommendations from this meta-evaluation should be integrated into the MFA **case management system** (AHA), country programming, country strategy papers and feed into **policy guidelines and tools**.

1 INTRODUCTION

1.1 Evaluation context

Evaluations of development cooperation comprise an important part of the internal control and monitoring of effectiveness of the Ministry for Foreign Affairs of Finland (MFA). The purpose is to produce independent information on the results, efficiency, effectiveness, sustainability and impact of development cooperation. This will ensure accountability towards the development policy. Also, the purpose includes administrative procedures and organisational processing for improvement of development cooperation.

The evaluation function of the MFA is divided into two: (a) the Development Evaluation (EVA-11) which is attached to the Under-Secretary of State for development cooperation and development policy. EVA-11 is responsible for extensive and strategically important evaluations, including meta-evaluations; (b) evaluations of projects and regional programmes, which fall under the scope of decentralised evaluations performed by departments of MFA or the Finnish embassies. They can be appraisals, mid-term reviews, final evaluations and *ex-post* evaluations of projects.

Previous meta-analyses of decentralised evaluations were carried out under the responsibility of EVA-11, in 1996, 2007 and 2009. As a result hereof, steps have been taken to strengthen the quality and use of evaluations, including development of tools for guiding evaluations. A "help-desk service" was established in 2009 and the provision of regular evaluation training events was initiated in 2010, offered by EVA-11. Evaluation training is now integrated into the advanced development cooperation training, held twice a year. As such, an evaluation culture is strengthened within the MFA (MFA 2012e, 41).

Meta-evaluations are, according to the Organisation for Economic Co-operation and Development (OECD), "used for evaluations designed to aggregate findings from a series of evaluations. [They] can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators" (OECD 2002, 27). The present meta-evaluation entails both elements. It not only covers decentralised evaluations but also mid-term reviews and appraisals.

The rationale for this meta-evaluation was stated in the Terms of Reference (TOR, 3; Annex 1): "The meta-evaluation of evaluations is an excellent means to bring together the otherwise scattered knowledge and lessons learned from the decentralised and centralised evaluation systems. At the time of the emergence of the 2012 development policy and the imminent commitment of the strategic planning of the development programmes, it is important that lessons from the past experience are assessed, made available and utilized".

1.2 Evaluation process

This meta-evaluation was performed between May and December 2012 by two independent evaluation consultants. An initial meeting was held between the evaluation team, the contractor and EVA-11 on 7 May 2012. A Start-Up Note was prepared, commented by EVA-11 and, in the revised form approved on 22 May 2012. The preparation of the Inception Report involved (a) discussions on the elaboration of a common understanding of TOR and defining a framework for a suitable approach to accomplish the purpose and objectives of the evaluation, (b) feed-back from a quality assurance team to draft versions of the Inception Report during the period 22 May-17 June 2012, and (c) EVA-11's comments to versions of inception reports, dated 26 June and 4 July 2012. The Inception Report was accepted by EVA-11 on 12 July 2012.

This meta-evaluation has strived towards compliance with the Development Assistance Committee (DAC)/European Union (EU) Quality Standards for high quality evaluation reporting (MFA 2011a).

Throughout this report the term 'project' has been applied to all development interventions and activities as the conditions for using a 'programme' term as defined in the *Manual for Bilateral Programmes* is, in most cases, not fulfilled (MFA 2012b, 12).

1.3 Purpose and objectives

The overall purpose of this meta-evaluation was twofold: to contribute to strengthening institutional learning and to be accountable towards the general public and professionals. Looking across decentralised evaluation of Finnish development cooperation, the meta-evaluation should provide the MFA with an overview of the quality of evaluation reports and the projects being evaluated across different regions and sectors.

The specific objectives of the meta-evaluation were to:

- build a comprehensive and independent analysis of the quality of Finnish development cooperation, the quality of decentralised evaluation reports from 2010 and 2011 and their TORs;
- assess the operationalisation of poverty reduction and Finnish cross-cutting objective (CCOs);
- identify possible trends of change and address trends related to best practices, challenges and obstacles;
- study the effects and use of evaluations at the practical level.

The above purpose and objectives reflect the fundamental assumption that quality of development cooperation can be improved through good evaluation practices, i.e. high quality of TORs and reports (appraisals, mid-term reviews and evaluations), and the proper management of knowledge captured in reports. The presence of these three practices may have a positive effect on development activities and they are

therefore relevant to investigate in one coherent meta-evaluation. Figure 1 illustrates the implicit interconnectedness between the three practices in influencing development cooperation and policies.

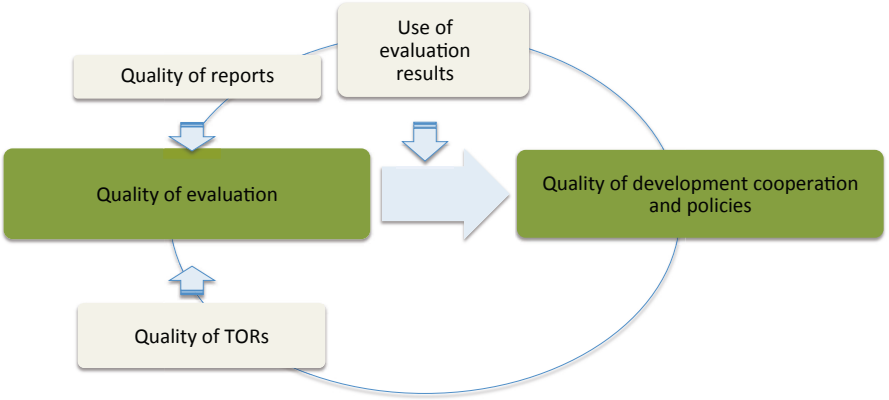


Figure 1 Interconnectedness between three practices influencing development cooperation and policies.

1.4 Sample

The meta-evaluation involved 41 decentralised sample reports (hereafter "reports"), including ten appraisals, 17 mid-term reviews (MTRs) and ten final, *ex-post*/impact evaluations. In addition two combined evaluations/appraisals and two combined mid-term reviews/pre-appraisals were included. The former was categorised as evaluations in the analysis, the latter as mid-term reviews. All reports were available electronically and accessed from the internet service Dropbox.

Ten appraisals (25%), 19 mid-term reviews (46%) and 12 evaluations (29%) constituted the sample. The sample also constituted 87% of the total number of appraisals, mid-term reviews and evaluations carried out in the period 2010-2011, or 41 out of 47 reports of bilateral and multi-bilateral projects. The spread of the reports in terms of report type and geographical region is illustrated in Figure 2 and Figure 3.

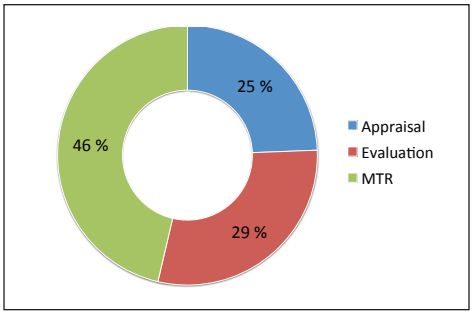


Figure 2 Reports by type.
Source: Meta-evaluation database.

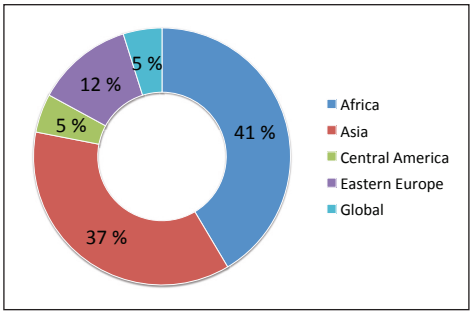


Figure 3 Reports by region.
Source: Meta-evaluation database

The sectors of the reports covered rural development (17%), environment (20%), water (17%), forestry (12%), education (10%), human rights, including gender equality and ethnic minorities (10%), information, communication and technology (ICT) and innovation (7%) and others (7%), including law and private sector development (Figure 4).

The reports covered different continents, regions and countries that receive Finnish development assistance. 17 reports covered the African continent (41%), 15 the East and South-East Asian region (37%), five Western Balkan and Eastern Europe and Central Asia (12%), two Central America (5%) and two global (5%). Country focus of the reports constituted 26 (67%) and regions and global focus 15 reports of the sample (33%).

Reports targeting Finnish long-term partner countries constituted 19 out of 41, or 46% (Figure 5). This included Tanzania (three projects), Kenya (four projects), Mozambique (one project), Zambia (two projects), Nepal (two projects), Vietnam (six projects) and Nicaragua (one project). Other countries included South Africa (one project) and Lao People's Democratic Republic (Lao PDR) (three projects). Furthermore, 12% of sample reports targeted Fragile States including Western Balkan (three), Eastern Europe (one) and Central Asia (one).

In addition, documents were selected by EVA-11 for ten of the 41 reports for the analysis of use of evaluations, including project documents, progress reports and minutes from steering group meetings. No information was provided regarding which MFA departments and/or embassies commissioned which appraisals, mid-term reviews and evaluations. Annex 2 presents a fact sheet for all the 41 reports with easy reference to be used forward in this report (R1, R2, R3, etc.). Definitions of the respective types of reports are presented in Box 1.

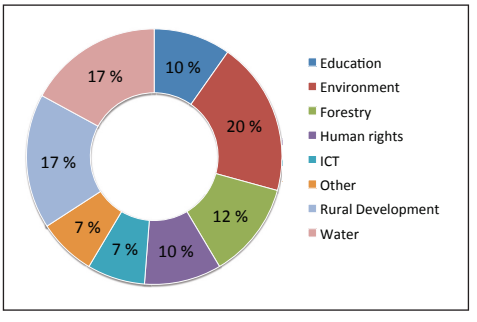


Figure 4 Reports by sector.

Source: Meta-evaluation database

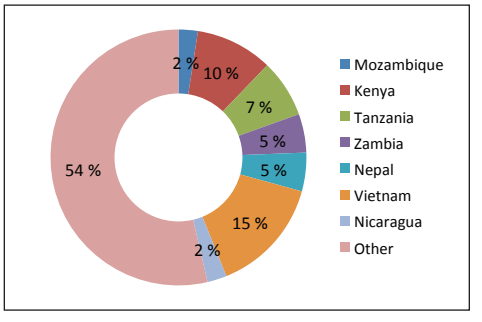


Figure 5 Reports by long-term partner countries.

source: Meta-evaluation database

The assessment relied on broad based information sources. Five reports were referred to between two and four times, 13 reports were referred to between five and

ten times, 20 reports were referred to between 11 and 20 times, and three reports were referred to more than 20 times. This "normal distribution" characteristic of the use of reports indicates strength of the meta-evaluation methodology.

The reports ranged from a power point presentation of 18 pages to comprehensive reporting of several hundred pages. Eight of the 12 projects that were subject to evaluations were initiated in the mid-2000s, two in the late 1990s, and two in 2008 and 2010, respectively. All project start-ups for the mid-term reviews were initiated in 2009 apart from two, in 2002. For appraisals most projects were initiated in 2010 and 2011.

Box 1 Definition of the three report types included in the meta-evaluation.

Appraisal
An "[a]ppraisal is 'an overall assessment of the relevance, feasibility and potential sustainability of a development intervention prior to decision of funding' (OECD/DAC). Appraisal is also called *ex-ante* evaluation. An appraisal provides an independent view – a second opinion – of the draft project document before implementation starts" (MFA 2012b, 54). The feasibility primarily aims to address adequacy in background analysis, the intervention logic, project management and aid effectiveness.

Mid-term review (MTR), mid-term evaluation
A mid-term review or mid-term evaluation "is an evaluation that is performed towards the middle of the implementation of the project (e.g. after two years of implementation in a 4-year project). A mid-term evaluation typically focuses on issues that are relevant for improvements in remaining years of implementation. It also often gives recommendations on the continuation of support" (MFA 2012b, 77).

Evaluation
An evaluation "is a systematic and objective assessment of either an on-going or already completed development programme. An evaluation focuses on the programme's design, implementation and its achievements. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors" (MFA 2012b, 76).

1.5 Methodology

The main evaluation criteria applied in the meta-evaluation were (a) the DAC/EU evaluation criteria: relevance, effectiveness, efficiency, sustainability, impact, coherence and complementarity (EU 2012; OECD 2012a, 3; OECD 2012b), and (b) the DAC/EU Quality Standards for evaluation reporting in development cooperation

(MFA 2011a). As a member state of the OECD and the EU Finland is committed to comply with these evaluation criteria and quality standards. Additional evaluation criteria applied included Finnish development cooperation's policy priorities of poverty reduction and CCOs, as well as the Paris Declaration principles, the Millennium Development Goals (MDGs) and Finnish value added.

The specific objectives of the meta-evaluation cover four major, interrelated themes:

- quality of TORs and evaluation reports;
- quality of Finnish development cooperation;
- the use of evaluation findings in projects;
- trends that can be identified from the assessment of decentralised evaluation reports, TORs, and findings in decentralised evaluation reports on Finnish development cooperation.

To guide the meta-evaluation an evaluation matrix was developed with an outset in the evaluation questions defined in the TOR. The evaluation matrix provided the overarching framework for the screening of the reports and data collection against the applied evaluation criteria. All data from the screening of the reports were subsequently added to a database. This enabled an analysis, including data handling, narrative assessment and scoring, of the individual reports and across the different types of reports, sectors, etc.

The analysis of the reports in this meta-evaluation is compared against the findings and recommendations of four previous evaluation reports, to identify trends. They are referred to as follows:

- 2007 Meta-analysis (White & Stenbäck 2007);
- 2009 Meta-analysis (Williams & Seppänen 2009);
- 2008 CCO evaluation (Kääriä, Poutianen, Santisteban, Pineda, Munive, Pehu-Voima, Singh & Vuorensola-Barnes 2008);
- 2010 Synthesis of evaluations (Caldecott, Halonen, Sørensen, Dugersuren, Tommila & Pathan 2010).

A detailed description of the methodology applied in this meta-evaluation is presented in Annex 3.

1.6 Quality management

Quality of the evaluation process was assured using the quality management system defined by MFA and Finnish Consulting Group and guidelines provided by the MFA. A quality assurance group – external to the evaluation team – provided systematic quality assessments of the inception report, draft final report and final report. The quality assessment report and the quality assurance grid were submitted at the same

time as the final report, using the EU evaluation report quality assurance criteria and the DAC/EU Quality Standards.

The evaluation team applied its own internal quality assurance mechanisms to guarantee consistency and reliability in the assessment of the 41 reports. The assessment and scoring given to the reports were discussed to clarify any divergences in assessments, including data handling, narrative assessments and scoring. The consultants eventually 'agreed' to a final assessment. Quality assurance was also applied during the course of the evaluation process. The consultants communicated on any issues of importance that emerged from the analyses and discussed and assessed accordingly. This on-going and constant quality assurance process has contributed to strengthening the quality of the meta-evaluation.

1.7 Limitations

The meta-evaluation was a desk study of 41 reports only. Without an opportunity for verifying findings in reports through interviews with MFA officials, evaluators, beneficiaries and other stakeholders, the validity of findings and conclusions regarding quality of Finnish development cooperation is weakened. However, by limiting the meta-evaluation to a desk study only, the quality of the reports became the centre of the analysis. This has exposed the strengths and weaknesses of the reports and whether they are of an acceptable quality that will enable decision-makers to take appropriate action.

The differences in purpose of the reports (appraisals, mid-term review or evaluation) are reflected in differences in terms of the importance of individual criteria included in the DAC/EU evaluation criteria and the DAC/EU Quality Standards. The meta-evaluation team has done its best to take into account these differences in the assessment of the reports and their TORs. This included assessing each of the three reporting types against the DAC/EU evaluation criteria and Quality Standards. The TOR models for the different types of reporting which was developed and applied for MFA/project staff and consultants in mid-2011 were not applied to any of the 41 reports.

The reports were commissioned by the regional departments and embassies of the MFA (regional departments or embassies) or as part of Finnish support to a multi-donor appraisal, mid-term review or evaluation. In case of the latter, the TORs were not necessarily prepared by the decentralised units nor were the evaluation teams selected by them.

Any attempt to extract general conclusions about quality across development activities has to be made with respect for the specific circumstances of each individual project – from financial support to a small research organisation, such as the International Council on Human Rights Policy based in Geneva, to longstanding project

support to sustainable management of natural forests in Lao PDR. The sample was limited to bilateral and multi-bilateral projects. Evaluations of budget and sector support programmes and special instruments, such as the Institutional Cooperation Instrument and cooperation with non-governmental organisations, were not included in the meta-evaluation. Typically these evaluations suit well for joint evaluations or are evaluated separately. Also, investment projects and allocations to EU agencies, EU instruments and the World Bank Group were not included. Therefore, it would not be fair to make generalisations about the quality of Finnish development cooperation based on the assessment of the reports in this meta-evaluation alone. However, the high representativeness of decentralised evaluations (87%) covering bilateral, multi-bilateral and some jointly financed projects over a full two year period provides a solid base for and a significant insight into the decentralised development efforts made by Finnish development cooperation.

The projects operate within the institutional context of the MFA and the appraisals, mid-term reviews and evaluations commissioned by the MFA reflect and are influenced by this context. However, this meta-evaluation was not assigned to undertake an institutional study within which the analysis of the qualities of development cooperation, reporting and TOR was a part.

2 QUALITY OF EVALUATION REPORTS

This section assesses the compliance of reports with the DAC/EU Quality Standards for evaluation reporting (MFA 2011a). This covers an overall assessment followed by an analysis of the extent to which reports address 33 criteria in the DAC/EU Quality Standards. These include *Overarching consideration* (2.1.2) – six criteria; *Purpose, planning and design* (2.1.3) – 12 criteria; and *Implementation and reporting* (2.1.4) – 15 criteria.

2.1 Compliance with DAC/EU Quality Standards

2.1.1 Overall assessment

None of the 41 reports achieved a score equivalent of Very Good quality (score 6). Seven reports achieved a score equivalent of Good quality (score 5) and 13 reports achieved a score equivalent to Adequate quality (score 4). 15 reports achieved a score equivalent to Inadequate quality (score 3) while the remaining six reports achieved a score between 2 and 1 equivalent to Very Poor quality and Unacceptable quality. Eight out of ten appraisal reports scored below adequate, half of the mid-term reviews scored below adequate, while three out of 12 evaluations scored below adequate. As such there was a clear trend from poor quality in appraisals to average quality in mid-term reviews and higher quality in evaluations. The ranking of projects based on score and type of report is illustrated in Table 1.

Table 1 Quality of reports measured against the DAC/EU Quality Standards.

Quality	Appraisal	MTR	Evaluation	Total
6 Very Good	0	0	0	0
5 Good	1	4	2	7
4 Adequate	1	5	7	13
3 Inadequate	4	8	3	15
2 Very Poor	3	1	0	4
1 Unacceptable	1	1	0	2
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

2.1.2 Overarching considerations

Overarching considerations apply to the entire evaluation process, i.e. the systematic and objective assessment of an on-going or completed project, its design, implementation and results (Table 2).

The MFA by-law, Norm 4/2011 (MFA 2011b, 3), states that: "the evaluations are commissioned to external, independent consultants and/or experts". However, few reports and TORs clearly described or substantiated the **evaluation process as free and open** (criterion 1.2). In 14 out of 41 sample reports descriptions were found as being open and fair, i.e. reference was made in the report or the TOR to, e.g. "invitation to tenders" or the report was carried out by "independent external evaluators".

Table 2 Number of criteria "addressed" per type of report for the *Overarching consideration phase* of the DAC/EU Quality Standards.

Criterion	Appraisal	MTR	Evaluation	Total
1.2 Free and open evaluation process	0	5	9	14
1.3 Evaluation ethics	0	1	2	3
1.4 Partnership approach	2	2	1	5
1.5 Coordination and alignment	0	1	0	1
1.6 Capacity development	1	0	0	1
1.7 Quality control	0	3	2	5

Source: Meta-evaluation database.

In a few of the reports there were indications that a free and open evaluation process was not strictly adhered to. These include situations where documentation had been withheld from the evaluation team (R33), or where interviews were "interfered" with by the main implementing partner, the Regional Environmental Center (R37). In one report the independence of the mid-term review team could be seriously questioned, because the design of TORs and team selection was executed by a commissioning group, in which one partner had a 20% stake in the project (R20).

Evaluation ethics (criterion 1.3) were addressed in a few reports but they were not addressed in any TORs. In these reports reflections were made, such as "the team shall exercise discrete manners". Also, considerations regarding ethics were explicitly described in relation to confidentiality, impartiality and objectivity (R35). In one report the evaluation team showed a commendable level of honesty and integrity in openly discussing observed favouritism in a project (R37). In this context it is noteworthy that Finnish development cooperation applies a zero-tolerance in relation to corruption behaviour. This is clearly stated in the following: "The Ministry for Foreign Affairs is committed to maintaining its reputation as an organization that will not tolerate abuse of position for personal or organizational gain" (MFA 2012c).

Considerations were described in regard to **evaluation partnership approach** (criterion 1.4) in a few reports. In two of them, and only sketchily, were there clear signs of inclusiveness and early involvement of stakeholders (R15, R20). The limited efforts to include stakeholders early in the evaluation process were also mirrored in criteria related to stakeholder involvement in the evaluation (criterion 2.5) and considerations for joint evaluations (criterion 2.6).

No clear description was found in reports or TORs that included **coordination or alignment** (criteria 1.5) with national or local monitoring and evaluation systems – except for one mid-term review (R20). This lack of focus on coordination and alignment with local systems indicates shortcomings in developing sustainable institutional structures in projects, including monitoring and evaluation systems. Similarly, **capacity development** (criterion 1.6) in evaluation was not described in the reports or TORs.

There was no mentioning of **quality control systems** (criterion 1.7) in almost any of the reports. The commissioner of the report delivered comments through the project steering committees or supervisory boards, in which partner groups and stakeholders are represented. The only well described quality control system was in the Education for All-Fast Track Initiative project for Mozambique (R13) – a huge global project on educational support commissioned by a large group of donors, including Finland. Efforts made to ensure effective quality control were also exercised in the Organisation for Security and Cooperation in Europe (OSCE) Academy project (R40). However, this was not entirely successful as the alumni organisation – a key stakeholder group – was not involved in the evaluation and only five out of a total of 175 alumni were interviewed.

Elements of informal quality assurance mechanisms were observed in the Sustainable Forestry and Rural Development project in Lao PDR (R24). In the mid-term review there was a description of major milestones in the evaluation process, e.g. wrap up meeting, circulation of draft, incorporation of comments in final report, etc. The milestones indicated that there had been a "window of opportunity" for key stakeholders to take stock of the evaluation work. Still, very few effective quality control mechanisms were described in the reports.

2.1.3 Purpose, planning and design

The Purpose, planning and design phase of the DAC/EU Quality Standards was addressed in a significantly higher number of reports than that of the Overarching considerations. *The Purpose, planning and design phase* comprised 12 criteria that guide the preparation of an evaluation (Table 3).

Table 3 Number of criteria "addressed" per type of report for *the Purpose, planning and design phase* of the DAC/EU Quality Standards.

Criterion	Appraisal	MTR	Evaluation	Total
2.1 The rationale and purpose of the evaluation	7	16	11	34
2.2 The specific objectives of the evaluation	6	14	12	32
2.3 Evaluation object and scope	5	15	9	29
2.4 Evaluability	4	13	7	24
2.5 Stakeholder involvement	3	5	2	10
2.6 Systematic consideration of joint evaluation	1	5	2	8
2.7 Evaluation questions	5	8	10	23
2.8 Selection & application of evaluation criteria	8	18	12	38
2.9 Selection of approach and methodology	5	10	11	26
2.10 Resources	6	3	3	12
2.11 Governance and management structures	1	1	1	3
2.12 Document defining purpose and expectation	7	15	11	33

Source: Meta-evaluation database.

In most TORs and reports the **rationale and purpose** (criterion 2.1) of the assignment were reasonably described. In a few evaluation reports the rationale and purpose was not mentioned at all (R2, R14, R39). It was not clear in the evaluation of the Institute for Human Rights and Business project (R36) whether it was the organisation itself or the Finnish supported projects that were assessed – or both. In the appraisal of the Ethiopian Water, Sanitation and Hygiene project (R9) and the mid-term review of the Environmental and Natural Resources Management project in Zambia (R15) the rationale and purpose was insufficiently described. In most cases the **specific objectives** (criterion 2.2) were defined and clearly presented in the reports and TORs. The exception included R2, R9, R11, R15 and R39 and to some extent R34 and R36.

The **object of the evaluation** (criterion 2.3) was described in most reports, often in the form of an incomplete intervention logic. The **description of scope** (criterion 2.3) was not fully covered in the majority of reports and TORs. When the scope is not described well it leaves the reader with an incomplete and fragmented understanding of the context in which the evaluation is carried out. This was confirmed in the description of criterion 3.6, Clarity and representativeness of summary.

The Southeast Asian Climate Change Network project (R34) was the only report that included clear and explicit reflections on the **evaluability** (criterion 2.4) of the project. However, more than half of the reports reflected upon or explicitly assessed the quality of the results-based framework and monitoring practices, which are determining factors for the evaluability of any project. In some of these reports the logframe or intervention logic were presented very clearly (e.g. R18, R20, R21, R25). An almost equally large part of the reports had no or very poor reflections on the project document, monitoring framework and monitoring practices. This was the case in the Biodiversity and Ecosystem Services for Local Sustainable Development in Western Balkans (R38). In this mid-term review the lack of quality in the project logframe should have been more critically assessed.

Stakeholder involvement (criterion 2.5) in the evaluation planning process was limited. It included stakeholders' comment on a draft evaluation report (R2), a comment on an "impact matrix" (R8) and a mission preparation note prior to the execution of the evaluation (R15). In a few large multi-donor funded projects (R13, R20) there was strong evidence of stakeholder participation in the evaluation planning, including the establishment of an Evaluation Overseeing Committee. From the initiation of the evaluation the Committee engaged all relevant partners through partnership meetings (R13).

In most projects where Finland and its recipient partner countries were *both* contributing to projects (financially and/or in kind) the assignments were commissioned solely by Finland. This included projects in Ethiopia (R10), Vietnam (R29, R30), Nepal (R22), Nicaragua (R33) and Zambia (R15). Also, in several regional and global projects to which Finland contributed there were no **joint evaluations** (criterion 2.6),

e.g. projects in the Mekong and in Western Balkans (R17, R18, R36, R37, R39). Joint evaluations of projects were however carried out with United Nations Development Programme (UNDP) (R1, R6), Denmark (R15) and United Kingdom (R35) and in collaboration with groups of donors (e.g. R5, R13, R20). Nevertheless, only limited efforts were made to embark on joint evaluation practices with stakeholders. This is in contrast to other MFA endeavours, including EVA-11's strong commitment in partner-led and joint evaluations (OECD 2012c, 68).

Regarding the use of **evaluation questions** (criterion 2.7) only a few reports translated evaluation objectives into relevant evaluation questions applying research questions and evaluation matrices (e.g. R19, R41). In the case of R41 this ensured a harmonised approach by the team of evaluators when assessing the three separate components included in the project. A great majority of reports had no evaluation questions defined to specify how evaluation criteria would be assessed or which dimensions of an evaluation criterion would be prioritised.

With the exception of two reports (R9, R39) one or more of the main **DAC evaluation criteria** (criteria 2.8) were applied in the reports. These criteria were often supplemented by other criteria, such as CCOs, coherence, complementarity, compatibility, and specific criteria, such as risk management, project design, alignment and coordination, stakeholder satisfaction and Finnish value added. For the appraisals the main criteria were relevance, feasibility and sustainability. The evaluation criteria were interpreted differently in the reports, though they should be defined "unambiguously" – according to the OECD definition of a criterion (OECD 2002). These issues on evaluation criteria are dealt with in more details in Section 4, Quality of Development Cooperation.

Some reports (R2, R3, R10, R22, R33) did not provide a sufficient description of **approach and methodology** (criterion 2.9). Yet, in up to half of the reports the approach and methodology were described in subsequent report sections. All evaluation reports included some kind of description of approach and methodology. Some midterm reviews and appraisals either omitted such a description or briefly mentioned that interviews were conducted and documents reviewed. An example of good practice is the evaluation of the International Council on Human Rights Policy (R35). In this evaluation the description of approach and methods was comprehensive and targeted. This included, e.g. measuring sustainability and impacts of the Finnish support to a research entity that produced reports. Such targeted approach and methodology represents the exception rather than the rule.

Assessing available **resources** (criterion 2.10) for the evaluations against effective implementation and meeting objectives was difficult. Data regarding evaluation budget, number of days assigned and skills of evaluators were only included in a few TORs and reports. Certain extreme situations were shown in the reports: from five days of fieldwork assigned for two persons to cover three countries – including travel (R37) – to the use of consultants in abundance (R20). However, it was difficult to assess

whether a good report with a weak summary (e.g. R33) or evaluation questions not being answered (e.g. R3) was the result of poor resource allocation or inefficient work by the evaluation team.

The MFA **governance and management structures** (criterion 2.11) of the evaluations process and phases were overall unsatisfactory. This was clearly evidenced in the lack of adopting stakeholder inclusiveness, lack of ensuring sufficient quality control and limited efforts made to embark on joint evaluations in partner countries. Lack of support to the review team (R37) and inappropriate time management (R26) were also found. Only the global fund project for Fast Track Initiative on Education for All in Mozambique (R13) showed a near ideal evaluation governance practice. The **document defining purpose and expectation** (criterion 2.12) was presented as a part of assessing the quality of the TORs (Section 3).

2.1.4 Implementation and reporting

The Implementation and reporting phase of the DAC/EU Quality Standards describes the evaluation process from the composition of the evaluation team and preparatory arrangements with stakeholders to the production of the report, the latter covering criteria 3.5 to 3.15. *The Implementation and reporting phase* comprised a total of 15 criteria (Table 4).

The information available in the reports and the TORs did not give a clear picture of the recruitment processes, gender balance or professional composition of the **evaluation teams** (criterion 3.1). In some reports it was mentioned that the official tender procedures were applied. The assignment forms used in two reports (R38, R39) were clear on recruitment conditions.

The degree of **independence of evaluators vis-à-vis stakeholders** (criterion 3.2) was not explicitly reflected upon in the reports. Indications of independency were found in various statements, where the evaluation was performed by "an independent international consultant" (R25), "an external independent observer" (R22) or "an external, independent and objective view" (R17). Since no clear descriptions were provided for in the reports on the selection of evaluation teams, the independence of evaluators could not be verified objectively.

Stakeholder consultation (criterion 3.3) addressed stakeholders' participation in and contribution to the implementation of the evaluation. Consultations were satisfactorily described in a majority of the reports. Different methods were used, e.g. interviews, field visits, observations, questionnaires and workshops. However, in some reports the beneficiaries appeared to have been left out. For example, potato farmers and seed farmers were not consulted in the Potato Sector Development project in Tanzania (R3), far too few villages were consulted in the appraisal of the Mama Misitu Forestry Governance project in Tanzania (R4) and the Mekong Water Dialogue project (R10) showed a clear imbalance in the information base of the evaluation. In addition, rel-

evant business communities or firms were not sufficiently consulted in the evaluation of the Institute for Human Rights and Business (R35). Issues concerned with the implementation of evaluation within **allotted time and budget** (criterion 3.4) are dealt with in Section 4 and criterion 2.9 in Section 2.

Table 4 Number of criteria "addressed" per type of report for *the Implementation and reporting phase* of the DAC/EU Quality Standards.

Criterion	Appraisal	MTR	Evaluation	Total
3.1 Evaluation team	2	8	7	17
3.2 Independence of evaluators vis-à-vis stakeholders	2	10	5	17
3.3 Consultation with and protection of stakeholders	5	17	12	34
3.4 Implementation of evaluation within allotted time and budget	5	5	5	15
3.5 Evaluation report	7	16	12	35
3.6 Clarity and representativeness of summary	4	13	9	26
3.7 Context of the development intervention	5	15	8	28
3.8 Intervention logic	6	14	6	26
3.9 Validity and reliability of information sources	4	14	12	30
3.10 Explaining the methodology used	5	12	11	28
3.11 Clarity of analysis	8	15	12	35
3.12 Evaluation questions answered, meeting needs	4	10	11	25
3.13 Acknowledgement of changes and limitations of the evaluation	1	8	5	14
3.14 Acknowledgement of disagreements within the evaluation team	0	0	0	0
3.15 Incorporation of stakeholders' comments	3	6	4	13

Source: Meta-evaluation database.

Almost all **reports were well composed** (criterion 3.5), well written and reasonably understood. However, in nearly half of the reports the list of acronyms was incomplete, and in a couple of reports not produced at all. A less clear picture was found with regard to clarity of **executive summaries** (criterion 3.6) in the reports. High quality summaries were characterised by a clear distinguishing between findings, conclusions and recommendations (e.g. R5, R6, R8, R20, R24, R30, R31, R34, R35) though not all of them were able to present a clear interconnectedness between the three. Some reports were without a clear summary (R3, R14, R15, R23 and partly R18). Poor quality summaries were present in more than half of the reports. They were too short, or too long, or lacking focus or – in too many cases – lacking important information described in the reports. In the Rural Water Supply and Sanitation Programme in Nepal (R22) findings, conclusions and recommendations were well structured and yet the summary was of low quality.

The context of the project (criterion 3.7) was often inadequately described in the reports. This included policy and institutional structures, the capacity and specific role of the implementing partners, and national and donor related activities in the sector. For example, the level of funding and the geographic spread of other donors' support were often not described. It was often difficult to assess the value and use of findings, conclusions and recommendations without a suitable contextual description. Therefore, the overall quality of the report would be diminished. This situation indicated problems related to the time allotted to the evaluation, the evaluators' knowledge of the project or their ability to get access to important information.

The **intervention logic** (criterion 3.8) is important for understanding the interconnectedness between the project's objectives, purpose and deliverables. In more than half of the reports the project logframes were insufficient. They did not clearly present the project's purpose, objectives and deliverables in a logical context. Where the project logframes were more clear and verifiable (e.g. R7, R18, R19, R20, R24, R25) reports also included assessments of the intervention logic as described in the project document. See also criterion 2.4 on Evaluability.

Validity and reliability of information sources (criterion 3.9) were generally well described in the reports. In most reports information was analysed using cross validation and triangulation, e.g. use of statistics, documents, field visits, observations, interviews and workshops. In some reports this was excellently performed (e.g. R4, R8, R30) in others less well performed, mainly due to lack of data, time constraints or the complexity of the project.

Some problems were observed with regard to the use (or lack) of information sources in reports. For example, key project beneficiaries were not brought sufficiently into the assessment of the projects, which questions the reliability of the reports (R29, R36, R37). Also, stakeholder participation and ownership in projects were occasionally used – often unsubstantiated – to meet a certain DAC criteria, e.g. sustainability or efficiency (R17, R25). Further discussion on this issue is presented in Section 4. In

the Rural Water Supply and Sanitation project in Nepal (R22), there was a mismatch between insufficient analysis and recommendations for continuation of the project.

Methodologies (criterion 3.10) were presented in almost all reports, in some very sketchy and brief, in others very detailed, and in a few comprehensively describing the use of qualitative and quantitative data. Two reports did not describe a methodology at all (R2, R3). Most reports **analysed** (criterion 3.11) their evaluation object well, but they varied in degree of diligence and clarity. Yet, they provided satisfactory evidence for conclusions and recommendations. Problems were observed in some reports, including the following: the absence of analysing according to DAC criteria (R37), the superficial use of the DAC criteria in the analysis (R25), the missing analysis of critical problems that should underpin the project (R4), and important findings not reflected in conclusions and recommendations (R29). Regarding **questions and needs met** (criterion 3.12) all issues are covered in criterion 2.7 and in Section 4.

Limitations to the implementation of the evaluation (criterion 3.13) were acknowledged in several reports. Limitations primarily related to lack of time (R5, R6, R18, R20, R26), lack of data for measuring outcomes, sustainability and impacts (R20, R27) or the inability to address the logframe in a suitable manner (R22, R32). Withholding or unavailability of data and important documentation also limited the work efforts of some evaluation teams (R29, R33, R39). **Disagreements** (criterion 3.14) were not addressed in the reports.

A majority of the reports did not provide sufficient evidence of clearly incorporating **stakeholder comments** (criterion 3.15) prior to their completion. The reports showed surprisingly little interest or systematic incorporation of views by stakeholders, though in half of the cases some kind of indication was mentioned – some clearly (R19), most vaguely.

2.2 Summary of key findings

- The reports included acceptable descriptions of the selection and application of the evaluation criteria, the rationale and purpose of the evaluation, specific objectives of the evaluation and consultations with stakeholders during the implementation of the evaluation.
- The reports were well written and could be reasonably understood, often included a method description and a clear analysis and referred to the TOR.
- The reports included insufficient information in relation to key areas of the evaluation process, particularly on inclusiveness of stakeholder and use of quality control systems.
- The reports included no or inadequate reflections on coordination and alignment with evaluation systems of partner organisations, the possibility of joint evaluation with partner countries and how the evaluation process could support building the evaluation capacity of development partners.

3 QUALITY OF TERMS OF REFERENCE

This section assesses the compliance of TORs with different criteria and topics as requested in the TOR of the meta-evaluation. This included an assessment of the TOR's objectives, purpose, scope and methodologies; their use of evaluation criteria; descriptions and use of key policy priorities of Finnish development cooperation (poverty reduction and CCOs); and how the Paris Declaration principles were addressed. Finally, a comparative analysis that determines TOR's role in facilitating or impeding the quality of reporting is presented.

3.1 Compliance with selected criteria

3.1.1 Objectives, purpose, scope and methodologies

The TORs varied in their descriptions of objectives, purpose, scope and methodologies and were not always clear in defining the difference between them. However, in most TORs the rationale, objectives and purpose of the assignment were sufficiently described. Almost all TORs included descriptions of approach and/or methodology, from very brief to comprehensive and detailed, varying significantly in quality and operationalisation. Approach and methodology were often synonymous with the DAC evaluation criteria but also included descriptions of the data gathering processes, e.g. desk reviews, field visits, interviews, etc. Consultants were asked to provide detailed and elaborated methodologies in the tendering process. This requirement was, however, not always fulfilled in a satisfactory manner (e.g. R22, R29, R32).

3.1.2 DAC evaluation criteria

Almost all TORs included references to the DAC evaluation criteria as well as special criteria for appraisals, often focusing on feasibility, effectiveness and sustainability. The evaluation criteria were presented in a separate section or as part of different sections of the TOR, e.g. in the approach or methodology section, in the scope section or under "issues to be studied" or "assessing the performance of the study". The description of each evaluation criterion differed: from general references to one or more evaluation criteria to very detailed break-down of each criterion, covering a vast number of evaluation questions.

When comparing the TOR with their respective reports the assessment showed that most reports answered the questions presented in the TOR. Often, the evaluation questions from the TOR were "copied" into the report. However, too many and unprioritised questions were asked for in TOR. In some cases the evaluation team did not answer sufficiently questions from the TORs, often due to lack of time (R4, R29, R36). The TOR for the appraisal of Mama Misitu Forest Governance project in Tanzania (R3) listed 37 not prioritised questions. Similarly, the TOR for the evaluation of

phase I and appraisal of phase II of the Core Environmental Programme and Biodiversity Corridors Initiative in the Greater Mekong Sub-region (R19) listed 30 questions for the evaluation and 40 questions for the appraisal. Many of these questions were defined as issues to be investigated. Other similar examples were identified (R15, R24, R32). A prioritisation of questions would help to guide the consultants to the issues considered most relevant.

3.1.3 Resources and time allotted

The assessment showed that the relationship between budgets and time spent on the assignments could not be assessed in a satisfactory manner. Often there was no budget stated in the TOR and time to be spent on the assignment was also not informed. Very few TORs included descriptions of both resources and time allotted. In those cases no discrepancy was identified. Timeliness in report delivery was overall kept.

3.1.4 Poverty reduction

Poverty reduction was mentioned directly in almost half of the TORs. Poverty reduction was often directly addressed in the objectives or evaluation questions where poverty concerns were at the very core of the project itself (R3, R33). Poverty reduction was not directly mentioned in several integrated projects on water supply, forestry and rural development (e.g. R4, R11, R22, R23, R30, R32) but referred to indirectly, mentioning "livelihood improvements for communities" (R4) or mentioning the target group, such as vulnerable groups (R8). Poverty reduction issues were not mentioned in projects on gender equality (R5), education (R13, R40), knowledge partnership (R16), regional water, private sector and security (R17, R18, R20, R41), international law (R25) and human rights (R35, R36).

3.1.5 Cross-cutting objectives

In most TORs the CCOs comprised an integrated part of the content. They were described either in a separate section or as an integrated element of the evaluation criteria section or the approach and methodology sections. Different projects prioritised different CCO issues. For example in the TOR for the ICT project in Tanzania (R2) focus was on sustainable development, gender and social equality. In the TOR for the Potato Sector Development project in Tanzania (R3) focus was on sustainable and local development. In the TOR for the Environmental and Natural Resources project in Zambia (R15) risk management was included as a priority. Gender equality and environment were the focus in the TOR for the Finnish Knowledge Partnership project in South Africa (R16).

While acknowledging the presence of a CCO focus in TORs there was a lack of systematically addressing CCOs as assessment criteria. There was often no clarity or justification for why particular CCOs were selected for assessment and others not in the TORs. CCOs were not included as assessment criteria at all in TORs in some projects (R20, R25, R26).

3.1.6 Paris Declaration principles

The Paris Declaration principles on aid effectiveness constitute an important challenge to development cooperation and evaluation practices (Wood, Betts, Etta, Gayfer, Kabell, Ngwira, Sagasti & Samaranayake 2011). The principles were mentioned and addressed in the TORs, however, to a lesser degree and less directly than CCOs and poverty reduction. Rather, they were addressed as part of the DAC/EU evaluation criteria, where ownership and harmonisation were considered as important elements of coordination, complementarity and coherence. The direct relevance of the Paris Declaration principles was discussed in some reports, e.g. in sections under "Aid Effectiveness" (R6) or where TOR focused on ownership, donor coordination and accountability (R15). Similarly, in addressing the evaluation criteria of effectiveness and sustainability in the TORs, the "results" principle of the Paris Declaration was indirectly addressed, as it relates specifically to achieving outcomes of projects. So, while the Paris Declaration principles were not necessarily directly addressed in TORs, the principles were discussed comprehensibly through other related themes and criteria.

3.2 Factors related to quality reporting

TORs' facilitating or impeding role in achieving good quality reports is the key issue to be addressed in this section. For this purpose a comparative analysis was conducted in which the ten best scoring and the ten worst scoring reports on quality were assessed against their respective TORs to identify possible correlation.

When assessing the *ten best scoring reports* (Table 5) an immediate observation was that four out of ten TORs were based on multi-donor funded projects (R5, R13, R20, R35), while the number of multi-donor funded projects in the total sample of reports was 25%. Projects in Vietnam and Nepal constituted five out of the ten projects, 50%, while 20% of the total sample.

Eight out of the ten TORs provided clear intervention logic structures comprising relevant and concise questions to be assessed. The correlation between these features of the TORs and high quality reporting appears to be strong.

In comparison with the remaining 31 reports the ten best scoring reports were characterised by the following: their ability to address well evaluability, approach/methodology, intervention logic and questions, but also their level of involvement of stakeholders and their quality of summaries and context. As such, correlation exists between good quality TORs and good quality reports. This may indicate that the units responsible for Vietnam and Nepal seem to draft useful and good TORs leading to good quality reports. In addition, only one of the ten TORs was for an appraisal. However, it was also found that one poorly designed TOR resulted in a well written and good quality report (R8).

Table 5 Ranking of the ten best scoring reports on quality.

Report		Report Type	Scores of 33	Rank
R13	Education For All (EFA)-Fast Track Initiative (FTI) Country Case Study: Mozambique	MTR	25	1
R5	Gender and Governance Programme III in Kenya (2008-2011)	MTR	23	2-4
R19	Asian Development Bank/Core Environment Programme (CEP), Mekong	Evaluation	23	2-4
R30	Rural Development Programmes supported by Finland in Two provinces, Vietnam	Impact Evaluation	23	2-4
R20	Mekong Private Sector Development Facility (MPDF)	MTR	22	5-7
R21	Strengthening of Environmental Administration and Management at the Local Level in Nepal (SEAM-N) III	Appraisal	22	5-7
R26	Technical Assistance for Support to Management of P135 II, Vietnam	MTR	22	5-7
R22	Rural Water Supply and Sanitation Project in Western Nepal (RWSSP-WN)	MTR	21	8-9
R35	International Council of Human Rights Policy	Evaluation	21	8-9
R31	Trust Fund for Forests (TFF), Vietnam	Evaluation	20	10

Source: Meta-evaluation database.

When assessing the *ten worst scoring reports* (Table 6), it was observed that none of TORs covered multi-donor funded projects. Six out of the ten projects were geographically positioned in Africa of which 40% covered Zambia and Ethiopia. Two of the three Western Balkan reports were also represented.

The TORs were all well structured, including logically and systematically addressing evaluation criteria and questions. As such, there seemed not to be a correlation between the quality of the TORs and the quality of reporting.

Table 6 Ranking of the ten worst scoring reports on quality.

Report		Report Type	Score of 33	Rank
R39	Sustaining Rural Communities and their Traditional Landscapes through Strengthened Environmental Governance in Transboundary Protected Areas of the Dinaric Arc, Western Balkan	MTR	1	41
R2	The Information Society and ICT Development Project (TANZICT), Tanzania	Appraisal	2	40
R14	Appraisal of the Phase II of the Programme for Luapula Agricultural and Rural Development, Zambia	Appraisal	6	39
R9	Community Led Water, Sanitation and Hygiene Acceleration (COWASH), Ethiopia	Appraisal	7	38
R15	Environmental and Natural Resources Management and Mainstreaming Program (ENRMMP), Zambia	MTR	10	36-37
R12	Improving Food Security in East and West Africa through Co-operation in Research and Education	Appraisal	10	36-37
R11	Rural Water Supply, Sanitation and Hygiene Program in Beningshangul-Gumuz, Ethiopia	MTR	11	35
R28	Development of Management Information Systems for Forestry Sector (FORMIS), Vietnam	MTR	12	34
R1	United Nations DDC Programme of Catalytic Support to Implement the Convention to combat desertification in the Arab States Region, Phase IV 2009-2012, Middle East & North Africa	MTR	13	32-33
R38	Biodiversity and Ecosystem Services for Local Sustainable Development in the Western Balkans (South East Europe BAP)	MTR	13	32-33

Source: Meta-evaluation database.

In comparison with the remaining 31 reports the ten worst performing reports were characterised by their limitation in addressing rationale, objectives, scope and evaluability/intervention logic. Also, they scored less on addressing evaluation criteria and questions, the validation of information sources and overall quality in analysis. Finally, at least four of the TORs among the ten poorest performing reports were appraisals, contrary to one for the ten best performing reports.

Following the above analysis correlation between high quality TORs and high quality of reports was not confirmed. Therefore other factors than quality of TORs are likely to influence the quality of reporting. As such, additional data gathering and analysis are needed before factors that improve or reduce the quality of reporting can be identified. Some of the factors may include:

- the overall institutional context of the MFA, including staffing 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.

Since mid-2011 MFA has applied standard models of TORs for different types of assignments, e.g. appraisals, mid-term reviews and final evaluations. This development indicates further progress towards improvements in guiding consultants in their assignments and supporting MFA staff in designing useful and relevant TORs. The TORs analysed in this meta-evaluation were prepared prior to the applications of the new standards.

3.3 Summary of key findings

- Nearly all TORs provided clear intervention logics, comprising relevant and concise questions – this despite the report was of high or poor quality. As such, there was no correlation between high quality TOR and high quality reports. Factors that facilitate high quality reporting should be identified.
- In the majority of the TORs the rationale and purpose of the assignment were sufficiently described and reference made to DAC/EU evaluation criteria.
- In a majority of TORs the number of evaluation questions was often too many and not prioritised.
- In most TORs the CCOs comprised an integrated part of the content, but there was a lack of systematic application of CCOs as assessment criteria.
- Poverty reduction was mentioned in less than half of the TORs. Where poverty concerns were at the very core of the project itself, poverty reduction was often directly addressed in the TOR.
- The Paris Declaration principles were mentioned and addressed in less than half the TORs, that is, to a lesser degree than poverty reduction and CCOs.

4 QUALITY OF DEVELOPMENT COOPERATION

This section provides an analysis of the quality of Finnish development cooperation based on information provided in the reports. The analysis was done against seven DAC/EU evaluation criteria: relevance, efficiency, effectiveness, sustainability, impact, coherence and complementarity.

4.1 Overall assessment

A relatively clear picture emerged from the assessment of reports regarding their compliance with the DAC/EU evaluation criteria (Table 7). On a scale from 1 to 7, relevance achieved a high score, 5.5, significantly higher than second ranked score, coherence (4.6) and complementarity (4.0). Four criteria: efficiency, effectiveness, sustainability and impact, all clustered around an average score of 3.2 to 3.6, significantly lower than relevance.

Table 7 Ranking of DAC/EU evaluation criteria based on assessments of all reports.

Main criteria	Rank	Mean score
Relevance	1	5.5
Coherence	2	4.6
Complementarity	3	4.0
Effectiveness	4	3.6
Efficiency	5	3.3
Sustainability	6	3.3
Impact	7	3.2

Source: Meta-evaluation database.

The assessment indicated that there was a discrepancy between the relevance of development policies and beneficiary needs on the one side, and the implementation and results of project activities (efficiency and effectiveness) on the other. A relatively strong correlation was found between weak project implementation results (efficiency and effectiveness) and the lack of effects towards sustained change (sustainability and impact). This demonstrates the need for new, powerful and easily applicable structures and tools that would strengthen efficiency and effectiveness in project design and implementation. The relative strength of coherence showed that efforts were made to apply integrated policy and mainstreaming practices.

When looking at the evaluation criteria disaggregated by sector (Figure 6) and region (Figure 7) there was no correlation between high relevance on the one side and high

effectiveness, sustainability and impact on the other side. The same is evident for individual projects.

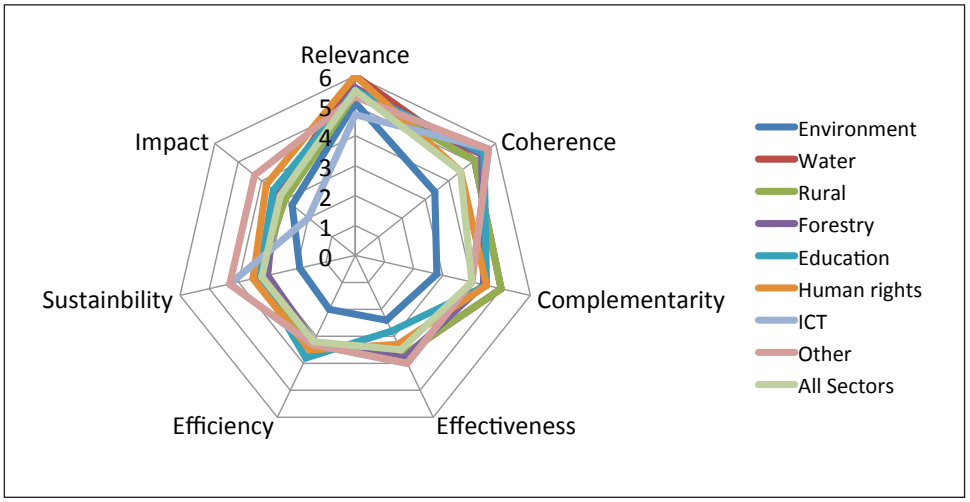


Figure 6 Evaluation criteria by sectors.
Source: Meta-evaluation database.

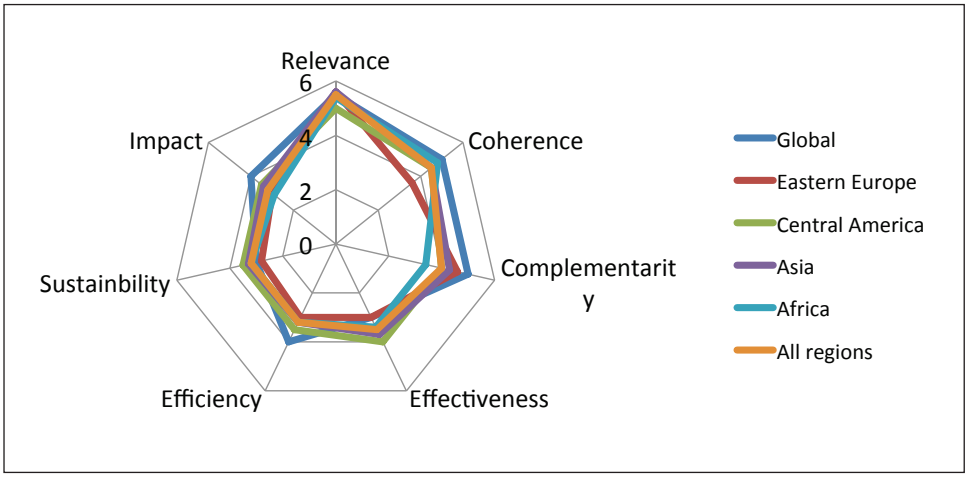


Figure 7 Evaluation criteria by regions.
Source: Meta-evaluation database.

4.2 Relevance

Relevance is defined in this meta-evaluation as the extent to which the aid activity is suited to the needs, priorities and policies of the target group, recipient and donor.

Relevance received a high score – more than 80% of the reports were rated in the Very good/Good category and five per cent of the reports in the Problems/Serious deficiency category (Table 8).

Relevance was primarily assessed for the entire project. Relevance was also assessed of project components covering different sectors, such as in the Luapula Agricultural and Rural Development project in Zambia (R14). While the overall relevance was considered high, the relevance of one component (agriculture) was considered low.

Table 8 Assessment of relevance by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	4	11	5	20
Good	3	5	6	14
Neutral	2	2	1	5
Problems	1	1	0	2
Serious deficiencies	0	0	0	0
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

4.2.1 Areas of strength

Assessment of the reports suggested that projects were generally aligned with Finnish development policy and policies and strategies of the partner country, first and foremost including the 2007 Development Cooperation Policy (MFA 2007a). Other Finnish development policies included civil society in development policies, security policies and policy guidelines for Western Balkans and Africa, and sector policy guidelines for forestry, water, and agriculture and food security. New policies and guidance on human rights-based approach to development was initiated in 2012 (MFA 2012f).

Some projects featured strong integrated relevance, e.g. where projects target the three dimensions of sustainable development, i.e. economic, social and environmental. The importance of addressing the three dimensions of sustainable development was emphasised in the 2007 Development Cooperation Policy as well as in the most recent policy (MFA 2012a), and also resonated in the conclusions of the outcome document of the Rio+20 Conference (UN 2012).

Projects assessed as having a high level of integration across sectors were mostly within the rural development, water and sanitation, forestry and environment, such as the Environmental Administration and Management at the Local Level project in Nepal (R21), the Rural Water Supply and Sanitation Programme in Western Nepal (R22) and the Water and Sanitation Programme for Small Towns in Vietnam (R29).

4.2.2 Areas of weakness

The degree to which beneficiaries' needs and priorities are met in a project is important when assessing the relevance of a project. Reports showed that a great majority of the projects did not explicitly assess these needs and priorities. This issue relates closely to the involvement of beneficiaries in project design and is addressed in details in the discussion of poverty needs and perceptions in Section 4.9.

Relevance is not static and it is therefore crucial – particularly during mid-term reviews – to address relevance and assess if drifts may have occurred. In some projects relevance either weakened or strengthened due to changes in the context or adjustments in project objectives, scope and activities. This was the case in the South East Asian Climate Change Network project (R34). The mid-term review found that focus and activities had diverted from the original objective, i.e. from supporting national climate change focal points, to supporting departments of energy, academic and research institutions working with energy issues, and private sector companies.

Another example was found in the Catalytic Support to Implement the Convention to Combat Desertification (R1). This project was generally assessed as highly relevant considering the overall strategies of the United Nations Convention to Combat Desertification as well as to the needs of the targeted countries. However, the project did not focus specifically on the reduction of land degradation but on environmental issues. The mid-term review therefore recommended a more focused project delivery against the priorities of the Convention to strengthen the relevance of the project.

Findings in the reports showed that projects tended to advocate for an integrated approach and address multiple dimensions of Finnish development policy. At the same time some projects were not realistic and balanced in their priorities against the resources available and the complexity of the context in which they operated. For example, the overall objective of the Development of Management Information Systems for Forestry Sector project in Vietnam (R28) did not achieve this balance. The project objectives focused on building a modern forest information system, but the project also emphasised other activity areas. This included gender equality and ethnic minorities through the creation of employment opportunities, reducing the number of poor households and completing allocation of forests land to organisations, communities, households and individuals. However, no activities in several of these areas were reported.

For projects that did not have an integrated approach there was a substantial risk of facing adverse effects. This was exemplified in the Support to the Mekong River Commission project (R18). The appraisal described how one component was designed to consider rivers and dams as producers of hydropower without considering issues of water supply, flood protection, protein production and environmental protection. By not recognising these "multiple use" and "water security" issues there was a substantial risk that the project was not meeting the needs of the people in the region. As such, project relevance was clearly reduced.

4.3 Coherence

Policy coherence for development entails the systematic application of mutually reinforcing policies across government departments to help promote the achievement of internationally agreed development goals along with other global and national policy objectives.

A majority of reports described policy coherence from basically two levels: (a) that of nationally adopted policies for a sector and/or cross cutting objectives, and (b) that of internationally and regionally adopted conventions, agreements and principles, e.g. MDG and Education for All.

Table 9 Assessment of coherence by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	2	8	2	12
Good	4	2	2	8
Neutral	3	3	5	11
Problems	0	5	2	7
Serious deficiencies	1	1	1	3
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

Almost half of the reports scored in the Very good/Good category on coherence while 25% scored in the Problems/Serious deficiency category. The Very good category itself stood out for more than 25% alone (Table 9).

4.3.1 Areas of strength

Coherence with national policies was assessed as strong in most of the reports, particularly in the water and sanitation sector (R8, R9, R11, R22). At the regional project level coherence was supported through both national and international policies and mainstreaming efforts. In the Mekong Private Sector Development Facility project (R20) policy support and advice were provided to national governments on policy innovations and development related to the private sector. In the Gender and Governance Programme in Kenya (R5) and the Special Needs Education project in Ethiopia (R10) international and regional instruments were promoted in national policy development processes.

4.3.2 Areas of weakness

In the regional Mekong River Water Dialogue project (R17) and in the Rural Development and Poverty Reduction project in Nicaragua (R33) national policies were ignored. When the latter project was initiated in 2004, the official rural development policy of the Nicaraguan government was directed towards increasing rural "productivity". The project, nevertheless, instituted a subsidised poverty reduction policy aiming at the poorest segments of the rural areas. The result was, combined with other factors, a conflicting policy framework that contributed significantly to divert project implementation from its very inception. It took several years before changes in the national policies made the project concept coherent with rural development policies. Such lack of policy coherence had a negative influence on implementation and mainstreaming of policies.

Other projects did not align with national or regional policies and priorities. As such, they were not able to achieve policy impact. This assessment was found in the Support to the Mekong River Commission project (R18) and the project on Catalytic Support to implement the Convention to Combat Desertification (R1).

Very few reports addressed or mentioned policy coordination across ministries/institutions in partner countries. This indicated that horizontal sector policies in partner countries were not applied or were insufficiently coordinated. For example, the Innovation Partnership Project (R27) showed how science and technology policies and strategies in Vietnam were spread across many ministries, without any coordination. This threatened the progress and sustainability of the National Innovation System. The Rural Development project in Vietnam (R30) was another example, where human rights and other CCOs were not adhered to, despite Finnish support over a period of more than 12 year.

4.4 Complementarity

Complementarity is defined as the optimal division of labour between various actors in order to achieve optimum use of human and financial resources. This implies that each actor focuses its assistance on areas where it can add most value, given what others are doing.

Complementarity was identified in the reports vertically between project entities and horizontally between different projects operating within the same or a related sector. Complementarity scored almost equal between the Very good/Good and the Problems/Serious deficiency categories, i.e. 41% to 37%, respectively. Complementarity scored relatively lower in appraisals than in mid-term reviews and evaluations (Table 10).

Table 10 Assessment of complementarity by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	0	3	1	4
Good	2	6	5	13
Neutral	2	4	3	9
Problems	3	4	2	9
Serious deficiencies	3	2	1	6
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

4.4.1 Areas of strength

An example of strong complementarity between two projects within the same sector was identified in the Mekong Water Dialogue project (R17) and the Support to the Mekong River Commission project (R18). The two regional projects focused on national stakeholders across the region through a bottom-up and a top-down approach, respectively. The latter provided its overall support to the basin countries on water governance with particular emphasis on management, modelling and technical data. The former was a "dialogue" based initiative addressing participation in water governance from a beneficiary perspective. This was a good example of a combined horizontal and vertical complementarity at regional level.

Complementarity was also strong in the Strengthening of Environmental Administration and Management at the Local Level project in Nepal (R21). The project did a solid analysis and clearly outlined division of labour among donors. This included urban governance (World Bank), urban environmental development, including sewages systems (Asian Development Bank) and solid waste management (Finland). Potential linkages with other on-going interventions were also identified and detailed areas of coordination and synergy spelled out.

The potential for horizontal complementarity between large projects was exemplified in the Rural Water Supply and Sanitation project in Nepal (R22). Finland supported both United Nations' Children's Fund's (UNICEF) Water Supply and Sanitation and Hygiene Sector Approach Reform Programme and the Rural Village Water Resource Management Project in Nepal. The overall policy environment was strong where stakeholders catered for durable coherence and complementarity – reinforced by UNICEF's overarching sector harmonisation project. However, evidence in the mid-term review suggested that effective complementarity was not achieved.

The International Council on Human Rights Policy project (R35) and the Institute for Human Rights and Business project (R36) were examples of projects which had the

capability to produce relevant and independent research that complemented on-going discussions and policy demands in the human rights area. The Institute of Human Rights and Business filled a niche that provided constructive and realistic dialogue between human rights concerns and business operations.

4.4.2 Areas of weakness

Insufficient analysis of horizontal complementarity was identified in several projects, including the Rural Development project in Kenya (R7), the Mama Misitu Forest Governance project in Tanzania (R4) and the Improved Food Security in Western and Eastern Africa project (R12). In the Luapula Agricultural and Rural Development project in Zambia (R14) and the Sustainable Forestry for Rural Development project in Lao PDR (R23) the potential for vertical complementarity was not sufficiently explored and the division of tasks, inputs or responsibilities among partners and stakeholders were not satisfactorily defined.

The East Asian Climate Change Network project (R34) did never fully take into account the multitude of climate change mitigation or adaptation projects financed and implemented by other donor countries and organisations. As a consequence bilateral and multilateral donors were not aware of the project's activities, and the sentiment was that the project worked in isolation with limited interfacing with other relevant bodies and potential partners. In the Gender and Governance project in Kenya (R5) the mid-term review noted that the same beneficiaries were benefitting from three similar activities by three different projects. Projects with insufficiently practiced complementarity were often those where formal cooperation with government programmes and donors were limited. This included the Special Needs Education Programme in Ethiopia (R10) and the Rural Development Programmes in Vietnam (R30).

Complementarity received a lower score in appraisals than in mid-term reviews and evaluations. This indicated that complementarity was inadequately addressed in project preparation and design. As such, avoiding duplication and complementing other initiatives in projects are therefore less likely to occur. Consequently, a complementarity analysis should be an important activity in the early stages of project preparation, i.e. identification and formulation.

4.5 Effectiveness

Effectiveness is defined as a measure of the extent to which an aid activity attains its objectives at outcome and impact levels. As such, it is measured at the end of the project life cycle.

Effectiveness of development activities scored near to 50% in the Problems/Serious deficiency category, while less than every fifth report would be considered in the

Very good/Good category. Whereas all report types had low scores on effectiveness, appraisals recorded seven out of ten in the Problems/Serious deficiency category (Table 11).

Table 11 Assessment of effectiveness by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	1	1	0	2
Good	1	2	3	6
Neutral	1	7	6	14
Problems	4	5	1	10
Serious deficiencies	3	4	2	9
Grand Total	10	19	12	41

Source: Meta-evaluation database.

4.5.1 Areas of strength

Tangible results at the outcome level were found in relatively few projects, and mostly in the water supply and sanitation projects, as evidenced in the Water Services Trust Fund's Community Project Cycle in Kenya (R8), the Rural Water Supply, Sanitation and Hygiene Programme in Nepal (R11) and the Water and Sanitation Programme for Small Towns in Vietnam (R29). While the latter project did measure towards outcome levels, the stated outcome indicators were not used, i.e. supporting regionally balanced economic development and decreased migration to large cities. Also, while acknowledging a relative strong effectiveness in the Finnish supported water supply and sanitation sector evidence also shows, according to Blue Planet Network, that up to half of these sector projects are not sustainable after five years (Brookland 2012). Other reports showing tangible results included R6, R21 and R31.

4.5.2 Areas of weakness

Measurement of effectiveness presupposes the existence of project objectives and a well designed results chain. However, most projects were characterised by an overall weak design, including the lack of well defined outcome indicators (e.g. R22, R36, R38). As a consequence the focus of the meta-evaluation was on outputs and the extent to which these seemed likely to contribute to project outcomes and objectives. Box 2 presents briefly selected deficiencies in the design of selected projects that limit measurement towards effectiveness.

Box 2 Selected deficiencies in the design of selected projects limiting measurement towards effectiveness.

Project		Project design deficiencies
R7	Rural Development project in Kenya	Weak project design with poorly formulated objectives and no measurable targets.
R9	Water, Sanitation and Hygiene project in Ethiopia	Outputs not balanced against resources.
R14	Luapula Agricultural and Rural Development project in Zambia	Serious flaws in the intervention logic as fundamental assumptions (of two out of three components) were contested by the appraisal team.
R17	Mekong Water Dialogue project	Insufficient logframe structure.
R34	South East Asian Climate Change Network Project	Insufficiently defined objectives of projects.
R35	International Council on Human Rights Policy project	Lack of strategy for medium and long term goals.
R38	Biodiversity and Ecosystem Services for Local Sustainable Development in Western Balkans	Inadequate project design with poor definitions of outcome and output indicators.

Source: Meta-evaluation database.

Lack of baseline data was a problem in most evaluations and mid-term reviews (R1, R10, R16, R26, R32). Lack of baseline data is further discussed in Section 5. Without baseline data outcome indicators cannot, if defined at all, be appropriately measured. An evaluation would therefore often need to "retrofit" the point of departure for the project. A typical approach, where a baseline is missing, is for evaluations to follow the results chain and focus on the extent to which outputs seemed likely to have contributed to project outcomes and objectives. Some evaluations, such as the Technical Assistance for Support to Management of P135II project in Vietnam (R26) lost sight of the overall project objectives and focused on what could be measured, e.g. workshops, trainings and the number of participants. As such, it failed to acknowledge that achievement of outputs is a measure of efficiency and not effectiveness. Finally, seven out of ten appraisals were in the Problems/Serious deficiency category. This indicated problems with drafting relevant intervention logic for projects in their initial phases, i.e. during identification and formulation.

4.6 Efficiency

Efficiency measures the outputs – qualitative and quantitative – in relation to the inputs. It is an economic term, which signifies the extent to which the project should use the least costly resources possible (funds, expertise, time, etc.) to achieve the desired results. The assessment of efficiency generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted. This approach was not applied in the reports which indicated a more subjective approach to assessing efficiency.

Efficiency is mainly assessed against a variety of factors. The most important factors include project implementation settings (institutionally and financially) and the actors engaged in shaping the implementation of the project. As such, efficiency assessment relates to a wide range of different project properties. They include at the least the following:

(a) project design and planning;
(b) implementation and management, including coordination and finances;
(c) monitoring and evaluation;
(d) the number, quality and capacity of technical assistance personnel;
(e) implementing partners and their staff;
(f) the engagement of donors;
(g) main beneficiaries of the project.

Almost two out of three reports (63%) received a score in the Problems/Serious deficiency category. Less than one in five reports received a score in the Very good/Good category. It was the same picture across types of reports (Table 12).

Table 12 Assessment of efficiency by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	1	0	1	2
Good	1	2	2	5
Neutral	1	4	3	8
Problems	6	7	2	15
Serious deficiencies	1	6	4	11
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

4.6.1 Areas of strength

Although overall assessment showed low efficiency some reports showed areas of high efficiency, for example through strategic partnership and prioritisation of activities. The Catalytic Support to Implement the Convention to Combat Desertification project (R1) made use of the comparative strength of a partner organisation by implementing activities through UNDP, which already had country and regional presence.

Experiences from the Rural Water Supply, Sanitation and Hygiene project in Ethiopia (R11) illustrated how a high degree of efficiency in a project could be achieved and sustained through the use of inexpensive technologies. The Nepali Rural Water and Sanitation project (R22) and the Mekong Private Sector Development Facility project (R20) exemplified how well structured monitoring and evaluation systems could be established to support the efficiency of the projects. In the Rural Development Programme in Two Provinces in Vietnam (R30) specific regulations in the project tendering procedures helped minimise the risks of corruption. This led to substantial savings in investment costs and cost effectiveness in the construction process.

4.6.2 Areas of weakness

The reports showed general efficiency problems in projects related to performance monitoring, administrative processes and human resources and technical assistance.

Performance monitoring
Operating unfitting performance monitoring systems or not executing appropriate corrections to well designed performance systems can have serious implications. For example, it can undermine the possibilities to take corrective measures, learn, replicate and further build a knowledge base, as found in the Environment and Security Initiative project (R41). Or it becomes an ad hoc activity, as found in the Innovation Partnership project in Vietnam (R27), or monitoring can become too cumbersome, e.g. if monitoring and evaluation templates are too detailed and comprehensive, as was found in the Finland Knowledge Partnership on ICT project (R16). Inefficiency was also observed where performance monitoring could not be executed in a satisfactory manner because data inputs expected from external institutions (e.g. the local government) were not delivered. This was the case in the Water and Sanitation Programme for Two Small Towns in Vietnam (R29).

Administrative processes
Efficiency problems also related to the administrative processes of the projects. A recurrent consequence of inadequate administrative processes was low disbursement rates as identified in several projects (R22, R26, R29, R32), which sometimes led to delays in programme implementation (R16, R33).

The administrative processes were particularly challenging in projects supporting innovative initiatives, where a "one size fits all"-approach in administration was not conducive to efficiency. This was observed in the Finland Knowledge Partnership on

ICT project (R16) and the Innovation Partnership project in Vietnam (R27). Also, while efficiency benefits could be achieved by applying well designed tendering procedures, the opposite was also evident. For example, contractors for the construction of water supply schemes in the Water and Sanitation Programme for Two Small Towns in Vietnam (R29) sub-contracted work to local firms with poor capacity and skills. As the contractors did not perform on-site supervision the delivery of project outputs was reduced.

Another administrative challenge was the management of joint efforts towards improved aid effectiveness (i.e. use of country systems), which is one of the key principles of the Paris Declaration. In the Fast Track Initiative Education for All project in Mozambique (R13) the World Bank required a special arrangement in order to use the government system for channelling funds. This demand reduced the efficiency of the pooled funding mechanism significantly. In the Technical Assistance for Support to Management of P135 II in Vietnam (R26) efforts were made by donors to apply joint technical assistance with the aim to strengthen and improve coordination of donor contributions to the national P135 II programme. The coordination mechanism did not succeed, had a negative effect on efficiency and was subsequently discontinued. In the Forests and Forest Management project (R32) in Central America a lack of a harmonised approach among key stakeholders resulted in increased administrative burdens.

Human resources and technical assistance
The reports also presented overall problems related to human resources in projects. They showed examples of projects where skills, knowledge and experience among partner institutions or other stakeholders had an impact on efficiency, positively as well as negatively. In a few water supply and sanitation projects (R8, R22) the involvement of community based organisations and non-governmental organisations as leaders and responsible for project implementation was problematic. Due to continuously poor performance and political interference they were not able to establish and manage the water supply schemes.

Technical assistance also appeared problematic in several projects. They included the following: lack of technical support in prioritised areas of the project (R4, R22); shortcomings in skills and experience (R30); confusion and disagreements over responsibilities and decision-making processes between technical assistance and the main implementing partner (R33); and massive overspending by the main implementing partner (R38).

4.7 Sustainability

Sustainability is concerned with measuring whether the benefits of an activity, output or outcome are likely to continue after donor funding has been withdrawn. Projects must be institutionally, environmentally and financially sustainable. Social and technical sustainability has been added in the *Manual for Bilateral Programmes* (MFA 2012b).

Almost two out of three reports were observed in the Problems/Serious deficiency category for sustainability, or 63%. Seven per cent scored in the Very good/Good category. The same picture was drawn across the three report types (Table 13).

Table 13 Assessment of sustainability by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	1	0	0	1
Good	0	2	0	2
Neutral	3	3	6	13
Problems	2	10	4	16
Serious deficiencies	4	4	2	10
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

4.7.1 Areas of strength

In several reports were participation and ownership described as a strength that catered for consolidating and sustaining projects results (R1, R7, R10, R30, R38). This included participation of parents and school councils in an ethnic primary education programme (R26). In some projects beneficiary groups set a real "demand-driven" agenda for sustaining benefits. An example included a strong commitment from water users to operation and maintenance of and willingness to pay for public water services (R22). Similar demand driven agendas were observed in the Water and Sanitation project in Vietnam (R29) and the Potato Sector Development Project in Tanzania (R3). While these demand-driven agendas often reflected the real needs and concerns of the beneficiaries, they were not addressed satisfactorily by the projects. See also Section 4.9.1 on poverty reduction.

Involving stakeholders with high-level prominence or political clout can help supporting project results. Such properties of a project can ensure long-term financial support, i.e. from businesses and donors as identified in case of the Institute for Human Rights and Business (R36). However, a change in the prominence or political power of stakeholders may also turn projects around as the political influence can have the opposite effect. An example was where the project became local partner-driven and not demand and beneficiary-driven. This was found with the Regional Environmental Center's management of the Education for Sustainable Development project in Western Balkan (R37).

4.7.2 Areas of weakness

The reports showed how assumptions comprised a critical factor in risk management and whether project benefits and results could be sustained or not. Examples includ-

ed the following assumptions: national recruitment policies would change to facilitate project achievements (R3); experienced and trained project staff would be absorbed into ministries following the completion of the project (R25); and data for project monitoring would be delivered from external institutions (R29). Similarly, it was assumed that institutional and programme support external to the project (through the "agriculture growth corridor") was required for the Potato Sector Development project to succeed (R3). It was also assumed that all stakeholders were willing and committed to engage in a "dialogue" process towards democratising water governance in the Mekong basin – and that the riparian governments would be willing to "change policies" based on this process (R17).

In a theory of change context identifying assumptions should allow for checking, debating and enriching project design and implementation (Vogel 2012, 4). However, in reality the reliance of assumptions as a part of risk management often leads to *complacency* among project stakeholders and eventually put a project increasingly at risk of failure. A brief discussion on risks and assumptions in projects is presented Box 6 in Section 6.6.

Institutionally compounded and financially large projects, covering sectors like rural development, water and sanitation and infrastructure, aimed toward integration with national structures, policies and strategies (R7, R8, R14, R29). However, the process was often slow as these projects were complex in nature, big in size and operated in cross-institutional contexts. The commercial viability of these projects was also limited. This often included poor cost recovery tariffs and revenue collection as well as high operation and maintenance costs – as observed clearly in at least two water supply and sanitation projects (R8, R29).

Without the Finnish support, project components of the Support to the Mekong River Commission project (R18) would collapse, despite long-term Finnish support to the Commission. Also, projects where subsidisation was an important aspect of project delivery often showed poor sustainability. Beneficiaries are often inclined to omit activities when subsidised services or products are terminated. This was the case in the Gender and Governance Programme in Kenya (R5) and the Strengthening of Rural Development and Poverty Reduction project in Nicaragua (R33). Poor environmental and livelihood sustainability was also observed in reports (R1, R7, R18).

A special feature of sustainability related to projects where imbalances of allocated resource were observed. While budgets were supposed to match the overall objectives of a project, budgets favouring one beneficiary group over another was observed in several projects, which tilted project objectives (R4, R17, R18, R29, R38). Particularly, the favouring of central stakeholders at the expense of important beneficiaries at local levels was common.

Increased participation and awareness in projects were frequently seen as properties of sustainability. However, often there was no evidence that linked them to meeting project outputs or outcomes (R17, R25). Therefore, participation and awareness

cannot and should not on their own merits be used as justifications for funding new projects or the continuation of existing projects.

4.8 Impact

Impact is defined as the positive and negative changes produced by a development activity, directly or indirectly, intended or unintended.

Impact scored 66% in the Problems/Serious deficiency category (or two out of three reports) and 12% scored in the Very good/Good category. While appraisals only reflected on impact as part of assessing the project's "feasibility", evaluations and mid-term reviews addressed impact directly. 15 out of 19 mid-term reviews were considered in the Problems/Serious deficiency category (Table 14).

Table 14 Assessment of impact by report type.

Score	Appraisal	MTR	Evaluation	Total
Very good	1	1	0	2
Good	1	0	2	3
Neutral	1	3	5	9
Problems	2	10	2	14
Serious deficiencies	5	5	3	13
Grand Total	**10**	**19**	**12**	**41**

Source: Meta-evaluation database.

4.8.1 Areas of strength

Impact was mainly identified in projects that related to the rural water supply and rural development sectors, and to policy development and mainstreaming. The reports also provided examples of how small projects can contribute to fast and positive impacts. While the assessment indicated overall weaknesses, several areas of strengths were also identified.

Support to policy development and efforts to mainstream policies were among the most tangible impacts observed in the reports. This included mainstreaming of the MDGs into national policies in the Gender and Governance Programme III in Kenya (R5); policy support to Universal Primary School Completion, teacher education and school construction in the Education for All Fast Track Initiative project in Mozambique (R13); and development and operations of a large number of policy and strategy tools in the Trust Fund for Forests Financing Mechanism project in Vietnam (R31). Concepts and policies developed in the two regional environmental projects in the Western Balkan (R37, R38) were tested for application and integration into national sector development policies in Serbia.

The two impact evaluations in the report sample (R8, R30) claimed high impacts of the projects as presented in Box 3. It should be noted that not all the stated impacts, of which some are not strictly impacts, were substantiated or verified. For example, there was no data showing enrolment of students related to the claim that increased access to education was achieved from infrastructure support (R30).

Small projects, like the Institute for Human Rights and Business project (R36) and the International Council on Human Rights Policy project (R35), demonstrated impact. This was identified in relation to the ability to set an agenda for international debates on corruption through effective networks and strategic partnerships. Codes of conduct of businesses in Kenya, including Unilever Tea and Kenya Flower Council were strengthened. Impact was achieved from the training provided to large tea growers and flower growers on how to apply the principles of human rights in their businesses.

Box 3 Impacts of projects claimed by two impact evaluation reports.

The Impact Evaluation of the Water Services Trust Fund project in Kenya (R8) assessed the project to have contributed to (a) reduced work load of women; (b) reduced disease burden and hence expenditure on medicine, (c) women participation and representation in leadership, (d) increased socio-economic activities; (e) reduced family conflicts; (f) improved hygiene and sanitation awareness; (g) improved accountability; and (h) improved water coverage.

Based on the Impact Evaluation of the Rural Development Programmes in Vietnam (R30) it was claimed that the project (a) increased access to education achieved from the infrastructure support together with (b) an increase in access to government services and (c) an increase in irrigated crop production. The project had a direct impact on (d) enhancing incomes and food security of participating households and, indirectly, on (e) increasing access to markets through the construction of roads, bridges and market facilities. The impact in one of the two provinces in Vietnam included contribution to (f) improved living conditions and (g) a reduction in the actual number of poor households in target districts compared with other districts.

Source: Meta-evaluation database.

The institutional framework within which a project is implemented should be conducive in order to turn project results into impacts. This approach was applied in the case of the Mekong Private Sector Development Facility project (R20), according to the mid-term review. This included a long-term strategy, the implementing partner's network and familiarity with the country context, and multi-year programmatic engagement. These were critical factors in achieving impact and continued to be a major source of competitive advantage for the project implementing partner.

4.8.2 Areas of weakness

Weaknesses of impact assessment related first and foremost to the lack of data and structures for measuring impact. Due to the poor intervention logic and logframe set-up, lack of baseline data and no clear theory of change descriptions in the reports, it was not possible to address impact in a satisfactory manner. Questions useful to evaluate impact could therefore not be answered: What has happened as a result of the project? What real long-term difference has the project made to the beneficiaries? How many people have been affected? The two impact evaluations (R8, R30) were not able to assess impact based on baseline data, but on available statistics, studies, surveys and interviews.

The most problematic impact issue in a project was found in the Support to the Mekong River Commission Plan for 2011-2014 (R18). One component of the project was to support sustainable hydropower, since the countries in the Mekong region plan for a massive introduction of new hydropower projects. The appraisal of the component showed that the project was inadequately designed and that without significant changes to the component, it would most likely contribute to the deterioration of the livelihoods of millions of people.

According to the appraisal a sustained impact could only be envisaged if "multiple use" and "water security" approaches were adopted through sustainable modelling and actively engaging national line agencies. With these recommendations the value of this appraisal has been significant. It challenged a project design that could have – over a period four years – wasted 11 million EUR of Finnish development funds.

4.9 Operationalisation of poverty reduction and cross-cutting objectives

4.9.1 Poverty reduction

Within the context of Finnish development cooperation, poverty reduction is understood as contextual and multi-dimensional and primarily achieved through activities that link economical, social and ecological sustainable development (MFA 2007a, 5; 12). Following this definition, the meta-evaluation found that poverty was addressed in less than half of the reports, while there was insignificant or no reference to poverty in the other half of the reports.

Perceptions of poverty and needs of beneficiaries
In some instances the differences in perception of what is meant by being "poor" can create a gap between what a project aims to deliver and what the beneficiaries perceive as their need. To better understand the needs of the beneficiaries an explorative analysis of perceptions of poverty was found valuable, because new dimensions or perspectives on poverty can emerge.

Examples of how poverty can be perceived differently were evident from several reports. In the Rural Development Programme in Kenya (R7) there was a strong focus on roads, market infrastructure and irrigation. Network creation between producers and buyers was not prioritised, though this was the key concern of the beneficiaries. In the Mama Misitu Forest Governance project in Tanzania (R4) it was noted that overall poverty would probably prevail without addressing land tenure and natural resources rights, which the project did not. In the Nepali Rural Water Supply and Sanitation project (R22) a successful mobilisation of resources from user groups and local bodies for water schemes was observed, but beneficiaries were more concerned with sanitation and hygiene needs. Also, several rural development, water supply and community forestry projects did not support women's requests for income generating activities to increase their family and the community's livelihood. Vulnerable groups are particularly exposed to poverty, but projects do not always address their problems. For example, while handicapped girls may receive services for special needs schooling, the girls' exposure to sexual harassment in rural Ethiopia refrain parents from sending their girls to school (R10). There were no indications of remedies being initiated to solve the problem.

In a poor locality in the Nicaraguan Rural Development and Poverty Reduction project (R33) the communities stated that next to water supply, road improvements for accessing towns were their main priority. This was not taking sufficiently into account by the implementation partner, i.e. the local authorities, when distributing funds for project activities. Another example of how poverty was perceived differently was the Sustainable Development in Education project in the Western Balkans (R37). This project focused on designing and operationalising environmental plans for local municipal-school partnerships. However, funds were used primarily on construction of basic school infrastructure. Therefore, school infrastructure may reflect the actual need of the local communities and not a "donor-driven" environmental approach to schooling.

These examples of how poverty and needs are perceived differently may not be exhaustive, but they indicate that project beneficiaries were not always sufficiently and actively included in the design of projects. This is an indication that is confirmed by the assessment of the criteria of relevance (4.2) and sustainability (4.7).

Methods like rapid appraisals (Chambers 1992) or classic grounded theory (Glaser 1992; Sørensen 2010) would identify more clearly the needs and concerns of the beneficiaries. This approach could have strong impacts on reducing poverty. At the same time this could strengthen project design and governance. This would also imply that poverty concerns and beneficiary needs were more realistically addressed. Furthermore, baseline data could be more focused by reflecting the actual needs of the beneficiaries, which could improve measurement of progress towards alleviating poverty.

Areas of strength

Findings in the reports indicated that a strong and more direct impact on poverty reduction was best achieved through bilateral funding to especially water and sanitation projects. Also, rural development, community forestry and environmental projects

showed strong impact on poverty reduction (R8, R11, R21, R22, R23, R29, R30). Examples are presented in section 4.8.1. Impact on poverty reduction was also achieved through large multi-donor projects, such as the Fast Track Initiative Education for All project in Mozambique (R13). Through this project affordable educational services were provided to the public.

Assessments made in appraisal reports suggested that a multi-dimensional poverty reduction approach was applied in several project proposals, such as the Improved Food Security project in Western and Eastern Africa (R12), the Environmental Administration and Management at the Local Level project in Nepal (R21) and the Sustainable Forestry project in the Lao PDR (R23). In some environmental projects the poverty-environmental nexus was well established, for example in R21 and R38 (Box 4).

Box 4 Two examples of the environment-poverty nexus.

Strengthening of Environmental Administration and management at Local Level in Nepal (R21):
The project design incorporated several results and activities that would contribute to increased well-being and reduced poverty of beneficiaries. For example, health and overall physical well-being of the beneficiaries would improve. Activities that would contribute to this included the model Village Development Committee (VDC) programme, emphasis on reducing waste water emissions from industries and addressing herbicide leakage from the tea estates. The project design specifically addressed the access of ultra-poor to services, particularly in the target VDCs. One important dimension the project addressed was the greater vulnerability of the poor to environmental hazards, including natural hazards, such as storms, floods and droughts, as well as human induced threats, such as air and water pollution. For instance, the poor people also suffered great loss of life and health from pollution and other environment-related causes. The project document explicitly recognised these concerns.

Biodiversity & Ecosystem Services for Local Sustainable Development in Western Balkans (R38):
Poverty reduction was a part of the overall objectives of the project. It related to employment and growth possibilities in connection with a "green economy" perspective. For example, project activities included eco-tourism and the cultivation of the Danube salmon.

Source: Meta-evaluation database.

Areas of weakness

No linkage was made between the project and poverty reduction in a few cases, despite an otherwise clear connection. Such a link seemed both possible and preferable in the Mekong Private Sector Development Facility project (R20), as the Facility oper-

ated a social and environmental business service that apparently did not include a job creation dimension, with particular focus on women entrepreneurs.

The Potato Sector Development project in Tanzania (R3) showed that a clear poverty reduction perspective was basically lacking, even though the project was oriented towards improved food security and poverty reduction. Another example was the Mekong Water Dialogue project (R17). This project's objectives were to improve livelihood security and human and ecosystem health through improving water governance. However, project activities did not sufficiently support the achievement of this objective. Appraisals also showed examples of projects where the purpose, objectives and activities of the project proposals were not sufficiently cohesive to prevent or reduce poverty (R3, R4, R7, R18).

Where the poorest and most vulnerable were targeted in project objectives, the results were often difficult to document. For example, in the Water and Sanitation Trust Fund project in Kenya (R8) approximately 10% of households reported that the project delivered positive impact on poverty reduction. However, due to lack of indicators and baseline data, it was not clear if this result was satisfactory. The Rural Water Supply project in Nepal (R22) targeted the poorest and most vulnerable in the project. Yet, because of an inadequate project design, including poor definition of output indicators, target achievements could not be documented.

4.9.2 Aid modalities for poverty reduction

Different aid instruments are used by Finland to achieve the poverty reduction goal of development cooperation, but only some of them are presented here. Different bilateral and multi-bilateral instruments and funding channels were used. Funding was channelled to international organisations, research institutes and development banks for project implementation. Finland also embarked on close collaboration with bilateral partners, including Denmark, Norway and United Kingdom. Bilateral funding was mainly provided through national organisations, including central ministries, local and provincial governments and national institutions (e.g. R4, R16). Funding to partner countries was mainly channelled through local authorities and trust funds for rural development, water and sanitation, community forestry and environment projects.

While the sample of reports showed a broad palette of modalities used by Finland, it was not possible to generalise on any correlation between modality and poverty reduction – with some reasonable exceptions made to the correlation between bilateral funding through local governments and decreased poverty.

4.9.3 Cross-cutting objectives

In accordance with the Development Policy Programme 2007 (MFA 2007a), a set of CCOs must be supported in all Finnish development policy. The CCOs are based on a human rights perspective and comprise:

- Promotion of the rights and the status of women and girls as well as gender equality and social equality;
- Promotion of the rights of easily marginalised groups, especially children, persons with disabilities, indigenous peoples and ethnic minorities, and the promotion of their equal opportunities of participation;
- Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome (HIV/AIDS) as a health and social challenge.

In addition to these CCOs, other cross-cutting objectives were subject to the meta-evaluation, including sustainable development and risk management. Several steps have been taken by the Finnish MFA to streamline CCOs in development cooperation. This includes the systematic integration of CCOs in all stages of the new case management system (AHA); generally strengthening attention to CCOs by the MFA Quality Assurance Group; and a team of sector advisors established as an expert resource to service the entire MFA organisation (MFA 2011c). Other steps taken include training and a web-site with open-access to tools for mainstreaming different CCOs.

Approximately 15% of the total report sample addressed one or more of the CCOs. If including the six focused environmental projects, almost one-third of the reports addressed one or more of the CCOs. Though efforts were made to integrate and address CCOs in the projects, it was not done systematically, which corresponded with findings of other recent 2009 Meta-analysis and 2010 Synthesis of evaluations. CCOs were considered only superficially in several reports (R3, R12, R13, R15, R28, R34, R41). For example, in the Development of Management Information Systems for Forestry Sector project in Vietnam (R28) the inception report linked the planned activities of the project to tasks with particular emphasis on gender equality and ethnic minorities. However, very limited progress was made in terms of monitoring CCOs.

Gender Equality

Gender equality issues were addressed in more than half of the projects. For example, the Environmental Management project in Nepal (R21) included gender equality, social inclusion and sustainable development as CCOs (Box 5). Other examples included the Lao Sustainable Forestry project (R23), where gender equality and ethnic minority activities were integrated elements in the overall project focus; the Rural Water Supply and Sanitation Programme in Nepal (R22), where a strategy on gender equality supported project implementation; the Catalytic Support to Implement the Convention to Combat Desertification project (R1), which showed disaggregation of data by sex; and the Special Needs Education project in Ethiopia (R10), which used male leaders as ambassadors for increasing the participation of women in leadership.

Box 5 Example of good practice of integrating CCOs in project design.

Strengthening of Environmental Administration and Management at the Local Level in Nepal (R21):
The appraisal of the project proposal for Local Level support to Environmental Administration and Management in Nepal (R21) showed strong integration of CCO features, e.g. relevant authorities, people and their representatives were identified with respect to gender equality and social inclusion objectives. The same applied to the climate change and disaster reduction component. While specific resources for gender equality and social inclusion objectives were not outlined, the theme was reflected in the responsibilities of local bodies in implementing the activities. Furthermore, gender equality, social inclusion and sustainable development were incorporated into the project document and reflected in the indicators and proposed activities. The appraisal found that the proposed strategies and activities were environmentally sound and cost-effective. Risks were identified and their probability of realisation analysed in sufficient detail. Finally the project completion phase addressed climate change issues by mainstreaming ecological sustainability through strengthening the resources of actors in the environmental sector. As such, this project provided a good example of how multiple CCOs intended to be operationalised in a relatively complex project context.

Source: Meta-evaluation database.

However, gender equality could have been more visible in the projects. As already mentioned the Mekong Private Sector Development Facility (R20) could have enhanced the business environment for women entrepreneurs, and projects in water supply and sanitation could more effectively have met women's requests for income generating activities. Similar aspects were relevant for the ethnic-women link in at least two other projects (R26, R30).

Democracy, human rights and climate change
Democratic development, human rights and climate change were rarely addressed explicitly. Only in a few projects were multiple CCOs addressed at the same time, such as in the Rural Water Supply and Sanitation Programme in Western Nepal (R22). Apart from the decentralised projects, i.e. rural development, water and sanitation and forest governance, good governance was particularly prevalent in the Mekong Water Dialogue project (R17), which aimed at strengthening democratic processes towards improved water governance in the region.

4.9.4 Paris Declaration principles

As a signatory to the Paris Declaration, the Accra Agenda for Action and the 2011 Busan Partnership for Effective Development Cooperation, Finland has strived to meet the commitments on ownership, harmonisation, alignment, managing for development results and mutual accountability. Over the years, Finland has been very active and forward looking in addressing challenges in Aid Effectiveness. This has in-

cluded an emphasis on effectiveness in policy guidance, operational planning, country-level programming and monitoring and in promoting effectiveness bilaterally and multilaterally. In practice Finland has prioritised the use of partner's public financial systems and strengthened the predictability of aid delivery. Overall Finland has put itself among the better performing half of the EU countries (OECD 2012c, 71-78).

Alignment

Looking more specifically at the individual principles, alignment stood out as the principle with highest visibility in the projects. In some projects, alignment was addressed explicitly, as in the Luapula Agricultural and Rural Development project in Zambia (R14). Here 32% of the total project budget was reserved for a specific component on alignment and harmonisation. The large budget allocation followed from the recognition that considerable resources were needed to invest in systems and capacity building to allow for effective project implementation under full alignment. In the regional Forest Management project in Central America (R32) alignment was internalised in the project outcomes and objectives, which were consistent with policies, laws and needs of the countries of the region.

If the capacity of the partner country is weak it can be tempting to avoid using the systems of the partner country. Such was the case in the project for Mainstreaming MDGs in Kenya's Development Process (R6), where the weak financial systems of the partner country meant that Finland used its own systems. Full alignment with institutions and systems of the partner country should nevertheless be an end target even in cases where existing structures are weak. At the same time, corruption practices prevail at the political and administrative levels in many countries that receive aid from Finland. This is critically counterproductive for establishing partnerships for effective development. Therefore, Finland is at the forefront of combating corruption in development cooperation – recently expressed in the production of the *Anti-Corruption Handbook for Development Practitioners* (MFA 2012d). Implementing the Paris Declaration principle of alignment as an important long-term development goal and effectively combating corruption and fraud are two major challenges in development cooperation. They can only effectively be addressed through common efforts made by the international aid community.

4.9.5 Millennium Development Goals and Finnish value added

The MDGs were not systematically addressed in the reports and were often not mentioned at all, even in projects with obvious connection to one or more of the MDGs. Despite not being addressed explicitly in a majority of the reports, projects contributed to the achievement of one or more of the MDGs. It would take little effort of projects to relate their interventions to these goals. Limited visibility of MDGs in evaluation reports was also found in 2009 Meta-analysis. Finnish value added was only visible in few reports (R14, R21, R23).

4.10 Summary of key findings

- The reports showed no correlation between high relevance on the one side and high effectiveness, sustainability and impact in the projects on the other side. Correlation was identified between weak efficiency and effectiveness and weak sustainability and impact. As such, the quality of the development cooperation was considered to be overall poor showing little effect towards sustained outcomes.
- Projects were generally aligned with Finnish development policy and policies and strategies of the partner country. Often the projects addressed multiple dimensions of Finnish development policy, but few reports explicitly assessed the degree to which the project was relevant for the main beneficiaries.
- Coherence with national policies was assessed as overall strong, particularly in relation to the water and sanitation sector. Horizontal sector policies in partner countries were not applied or were insufficiently coordinated.
- Evaluations and mid-term reviews were generally weak in terms of measuring impact, sustainability and effectiveness. This reflected that project objectives were poorly formulated, data at the outcome level were scarce, and baseline data lacking.
- Impact was relatively strong in integrated policy development. Also, a strong and more direct impact on poverty reduction was achieved through funding to projects in the water and sanitation sector, but also in the rural development, community forestry and environment sectors.
- It was not possible to generalise on any clear correlation between modality and poverty reduction.
- Needs and priorities of main beneficiaries were not sufficiently addressed in projects and not actively included in the design of projects.
- Reports show that CCOs were not addressed systematically in projects.
- Gender equality issues were addressed and/or included in more than half the projects. However, gender equality was lacking in several projects, where it should have been addressed.
- While the Paris Declaration principle of alignment is an important long-term development goal, effectively combating corruption and fraud is to be critically addressed in order to make alignment happen.

5 USE OF EVALUATIONS – ANALYSIS OF TEN SELECTED PROJECTS

Based on a comprehensive desk study of documents for ten selected projects, this section explores the extent to which project designs facilitated monitoring of progress, how monitoring was performed and if evaluation findings were used in the preparation of subsequent project phases.

The MFA selected the ten projects, which were different in terms of type, duration, geography and management. The sample included different types of interventions,

e.g. a rural development programme, a gender and governance programme, a knowledge partnership on ICT and an education programme on biodiversity (Table 15). Most of the projects were on-going between five and ten years and a few on-going since the late 1990s, while one project was still in its preparation phase. The projects were implemented in different regions and countries, including Kenya, South Africa, Lao PDR, Vietnam and the Western Balkan. Some projects were co-funded by other donors, some implemented through a multilateral organisation and others in partnerships between Finland and the partner country.

Table 15 List of the ten selected projects for the analysis of use of evaluation.

No.	Type	Project	Sector	Continent
R1	MTR	United Nations Programme of Catalytic Support to Implement the Convention to Combat Desertification in the Arab States Region, Phase IV 2009-2012, Middle East & North Africa	Environment	Africa
R5	MTR	Gender and Governance Programme III, Kenya (2008-2011)	Human rights	Africa
R12	Appraisal	Improving Food Security in East and West Africa through Co-operation in Research and Education	Rural Development	Africa
R16	Evaluation	Finland Knowledge Partnership on ICT (SAFIPA), South Africa	ICT	Africa
R17	Evaluation	Mekong Water Dialogues	Water	Asia
R23	Appraisal	Sustainable Forestry and Rural Development (SUFORD), Lao PDR	Rural Development	Asia
R25	MTR	International Law project, Phase III, Lao PDR	Other	Asia
R30	Evaluation	Rural Development Programmes supported by Finland in Two provinces, Vietnam	Rural Development	Asia
R34	MTR	Southeast Asian Climate Change Network Project, United Nations Centre Environmental Programme	Environment	Asia
R38	MTR	Biodiversity and Ecosystem Services for Local Sustainable Development in the Western Balkans (South East Europe BAP)	Environment	Europe

Source: Meta-evaluation database.

The documentation reviewed included project documents, evaluation and audit reports, progress reports and minutes from steering committee meetings (or similar). The set of project documents provided for each of the projects differed in terms of size and completeness, e.g. for some projects all annual and/or quarterly reports were available, in others they were not.

5.1 Strengths and weaknesses of use

The assessment showed that findings and recommendations from evaluations were generally disseminated among project stakeholders and addressed in subsequent project documents and/or annual work plans. The steering committee meetings were the most common venue for presentation and dissemination of evaluation findings. Minutes from these meetings indicated that there was a tendency to focus more on the positive findings and on issues pertaining to programmatic priorities than issues related to weaknesses of project design. The latter included lack of logic and consistency between results levels in the logical framework, weak indicators and lack of baseline data. This may have been one of the reasons why 71% of advisers considered responses to evaluations as being ineffective (Poate, Bartholomew, Rothmann & Palomäki 2011, 63).

5.1.1 Areas of strength

The approach taken to address evaluation recommendations following a discussion of these at a steering committee meeting (or similar) varied across projects. In two projects (R12, R34) a table was produced containing all recommendations, a response to the recommendation and the measures taken to address the recommendations. Such an approach was good practice and made it easy to track and monitor if recommendations were adequately addressed. For example, in the Food Security in East and West Africa project (R12) the following improvements were developed: a table format of tasks, outputs, outcomes and indicators with partners; a common template for reporting in collaboration with all partners; and an electronic self-evaluation form for use in connection with annual reporting.

Follow-up on evaluations and audit reports was also done through annual progress reports, e.g. in a specific section on follow-up on audit findings (R25, R30) and by including recommendations from the most recent evaluation in the project document for a subsequent phase (R34). This indicated that lessons were learned and also suggested that the project was responsive to recommendations, such as in the Mekong Water Dialogue project (R17). Recommendations from the evaluation of phase I and appraisal of phase II regarding the need to improve indicators resulted in a revised set of indicators of much better quality. They were presented at the Project Steering Committee meeting in September 2011.

5.1.2 Areas of weakness

However, all recommendations were often not addressed. In three of the ten projects it was not clear from the documents if and how recommendations were addressed (R23, R30, R38). For example, as a response to a recommendation to revise an indicator for the project objective to become more meaningful, the objective was removed from the logframe – without any justification (R12).

Timing was critical to ensure that evaluations added value to subsequent planning and implementation of programmes. In one project (R5) the planning of a subsequent project phase was undertaken two months before an evaluation was finalised. Findings from the evaluation were therefore not included in discussions and preparation of the subsequent phase.

5.2 Project design

The Guidelines for Programme Design, Monitoring and Evaluation (MFA 1999) and the *Manual for Bilateral Programmes* (MFA 2012b) describe the logframe as structured by the following four levels: development objective, overall purpose, results and activities. For each of these four levels, objectively verifiable indicators should be defined together with sources of verification and assumptions. The assessment of the ten projects showed a great variety in how logframes were structured. Some projects followed the structure recommended by MFA (R16, R30, R34), but a majority of projects made substantial modifications in the logframe structure. Table 16 provides some examples of how logframes were structured in five of the ten projects.

The different logframe structures reflected that the same terminology was interpreted differently across projects. What was considered a result in one project was considered an output in another project. While some projects had objectives, others had a goal and an expected impact and so forth. The confusion about terminology is common and can be frustrating for project teams and stakeholders. It complicates assessment of and reporting on progress across different projects. However, the logframes in the ten projects illustrated a bigger problem, namely, a general lack of causal relationships for all levels of results (output, outcome and impact) and their corresponding indicators.

Table 16 Five examples of logframe structures from ten selected projects.

R12: Food Security in East and West Africa	**R16:** Knowledge Partnership on ICT in South Africa	**R25:** International Law Project, phase III in Lao PDR	**R30:** Rural Development Project in Vietnam	**R38:** Education for Sustainable Development in Western Balkans
Task Output Deadline Outcome Outcome indicator User/Target audience	Overall objective Purpose Specific objective Results For each level: Indicators Means of verification Assumptions	Output targets Activities Responsible parties Inputs For each output: Baseline data Indicator Indicator target	Overall objective Purpose Results For each level: Indicators Means of verification Assumptions	Objectives Actions Outcomes Indicators Results

Sources: Meta-evaluation database.

5.3 Indicators

Looking across the ten projects there was a tendency towards improved quality of indicators in project documents from 2009/2010 and onwards (R5, R17, R25, R34). In six of the ten projects the indicators were not developed in line with the SMART criteria, i.e. Specific, Measurable, Attainable, Relevant and Timely. There was no consistency among the ten projects in terms of the results level for which indicators were defined. In five of the ten projects indicators were defined for all results levels but, as mentioned above, the logframes were not structured similarly. In the other projects indicators were defined only at the purpose and outcome levels (R1, R12, R38), or at the output level (R25). In the Sustainable Forestry and Rural Development project in Lao PDR (R23) no indicators were included in the project proposal for the period 2012-2016. Yet, the progress reports from 2009-2010 showed outcome/impact indicators formulated in line with the SMART principles.

Though indicators were defined for a certain results level, this did not ensure that indicators were meaningful proxies for the expected result. In one project indicators were simply a reflection of activities, which did not describe the actual progress towards purpose (R38). In another project indicators became irrelevant, as they were never modified despite strategic changes to the programme (R16). Few projects had defined target values for the indicators (R17, R25, R34) and only in three projects

were baseline data available (R1, R5, R25). In the Food Security in West and East Africa project (R12) no indicators were developed at the output level and indicators at the outcome level were of varied quality, i.e. downloads, citations, hits on webpages, dissemination of publications, etc. Good indicators are those that target behavioural change, e.g. number of farmers changing the dairy breed type. Outcome and output indicators are only useful when they measure something of importance and are developed in collaboration with key stakeholders.

5.4 Cross-cutting objectives

All of the ten reports addressed one or several of the CCOs, e.g. gender equality and reduction of inequality in their project documents. CCOs varied, from a section in a project document describing the general commitment of the project to mainstream, e.g. climate change or gender equality in activities, to activities where CCOs were targeted directly through project objectives, outputs, indicators and activities. In two instances a budget line was dedicated to gender equality (R1, R5) but in most projects CCOs were addressed at the overall level only, i.e. through project objectives (R12, R16, R23, R38, R30). The more visible CCOs were in outputs and indicators, the easier it was to monitor progress and evaluate results against these.

5.5 Baseline

A baseline study is described in the *Guidelines for Programme Design, Monitoring and Evaluation* (MFA 1999, 28) as a first step in the evaluation process, but it is not defined as a requirement when designing a project. This could explain why a baseline study was carried out in few projects only. This is problematic since any assessment of progress without knowledge of the point of departure often becomes a "qualified estimate".

A baseline study is of great importance for subsequent project management and evaluation, as it provides the basis for measuring whether a change has taken place or not. However, baseline studies are often either ignored or pushed forward. The reasons for this may relate to the lack of understanding of the inter-linkage between outcomes, indicators and baseline data. Baseline studies can only be useful if objectives, purpose and outputs of the project are properly described and important indicators developed.

Baseline data was identified in two of the ten projects (R1, R25). In two other projects the project documents included a commitment to develop baseline data as part of the inception or implementation phase (R12, R16) and in one project (R5) a baseline study was carried out half way through the implementation of the third phase of the project. In the latter case, the baseline study was undertaken as a response to a recommendation from the evaluation of phase II. Though the project document from November 2008 stated that a baseline study was in progress, it was not finalised until September 2010.

Where baseline data were defined in the project document, the type of baseline data differed. In one of the projects (R25) the baseline data focused on the current capacity in terms of qualifications among different groups of staff, e.g. 80% of instructors lack formal qualifications in international law. Such data allowed for measuring progress as the total number of qualified staff increased. In the other project (R1) baseline data stated qualitative descriptions of a situation, e.g. limited investment in the drylands or advocacy efforts needed. Such statements could not be used for monitoring progress and did not qualify as baseline data. In neither of the two projects (R1, R25) was baseline data disaggregated by sex or CCOs. However, in case of the former project the indicator targets were disaggregated by sex where relevant in its annual work plan for 2009.

In two projects plans were made to carry out baseline studies as part of the inception phase (R16) or as part of delivering the specific project components (R12). In case of the former, the baseline study would have qualified strategic priorities as these were all reflected in indicators in the project logframe. In case of the latter a baseline survey was an output under each component, but the exact data to be produced was unclear and stated in general terms, e.g. household survey.

5.6 Monitoring

Progress was reported through quarterly (R16, R23, R30), semi-annual (R1, R5, R17, R38) and annual progress reports (R24, R25, R35). With the exception of two projects (R23, R25), progress reports were mainly narrative descriptions of activities with no reporting on progress towards project outcomes or objectives.

In the Sustainable Forestry and Rural Development project in Lao PDR (R23) quarterly progress reports were comprehensive and included descriptions of both activities and progress made against outcome/impact indicators. The same counted for the International Law project in Lao PDR (R25) regarding the project's annual progress reports and quarterly board meetings (semi-annual from 2010 and onwards). Results were described against indicators and explanations provided if results were not achieved. Also, in addition to reporting on progress against the project logframe, a description was provided of activities performed to support ownership, a partnership approach, gender mainstreaming, etc. Table 17 shows an overview of the various types of performance monitoring mechanisms that were used in the ten projects.

Table 17 Overview of performance monitoring systems in ten selected projects.

Project	Performance monitoring
R1	Semi-annual progress reports. Primarily activity based reporting with long narratives of activities and few tangible results. Furthermore, progress was not reported against any clear target.
R5	Semi-annual progress reports were comprehensive with long narratives on activities performed in relation to the project objectives.
R12	Project implementation had not begun.
R16	Quarterly progress reports focused on activities of projects and a visual (smiley) was applied to indicate progress made in activities, budget and (towards) outcomes. It was not clear what progress was measured against since no targets were included. One reason for this was that the matrices developed for the project reporting focused on outputs, saying little about progress towards impact.
R17	Semi-annual progress reports. Primarily activity based reporting with long narratives for each of the results, but reporting was not done against the revised set of indicators.
R23	Quarterly progress reports which were comprehensive and included descriptions of both activities and progress made against outcome/ impact indicators using the logframe.
R25	Annual progress reports and quarterly board meetings (semi-annual from 2010 and onwards). Not clear if quarterly progress reports as such were produced. Results were described against indicators and explanations provided when results were not achieved. A particularly helpful feature of the report format was that progress or activities supportive of e.g. ownership, a partnership approach, gender mainstreaming etc. were reported on, in addition to the progress made against the logframe targets.
R30	Quarterly and annual progress reports. Both types of reports were comprehensive and focused on financial aspects, activities, outputs and lessons learned.
R34	Annual progress reports were narrative descriptions of activities but showed little in terms of results since targets were unclear.
R38	Semi-annual progress reporting. However, only minutes from one Steering Committee meeting (in Quarter 4) were available. The results presented were activities.

Source: Meta-evaluation database and documentation from ten selected project.

5.7 Summary of key findings

- At the project level, evaluations were generally disseminated among stakeholders and addressed in subsequent project documentation. However, there was no common approach to follow-up on evaluation recommendations at the project or corporate level.
- Logframes in the ten projects were designed differently and only three projects follow existing guidelines;
- There was a tendency towards improved quality of indicators in project documents from 2009/2010 and onwards, but in six of the ten projects indicators were not developed in line with SMART criteria.
- All of the ten selected projects addressed one or several of the CCOs, but varied in how it was done.
- Baseline data was identified in two of the ten selected projects and in neither of the two was baseline data disaggregated by sex or CCOs.
- Progress reports were mainly narrative descriptions of activities with no reporting on progress towards project outcomes or objectives.

6 TRENDS

This section assesses the trends related to the qualities of the reports, their TORs and the development cooperation as well as trends related to CCOs, the Paris Declaration principles and the goal of poverty reduction. To identify trends the analysis of the reports in this meta-evaluation was compared against the findings and recommendations of the four previous studies mentioned in Section 1.5.

Comparison across reports has limitations. Particularly, the different evaluations applied different assessment tools. Furthermore, the reports and TORs for each of the meta-analyses differed in numbers, sectors, geographical spread and in the distribution of types of reports (appraisals, mid-term reviews and evaluations). For example, the 2009 Meta-analysis (p 40) covered Fragile States projects with a total percentage of 42, while they represented only 12% in this meta-evaluation. The 2010 Synthesis of evaluations report only included assessment of evaluation reports commissioned by EVA-11. With these limitations in mind trends have been identified.

6.1 Quality of reports

When comparing the trend in quality of reporting from the two 2007 and 2009 meta-analyses and this meta-evaluation the picture shows the following: a relatively positive assessment in 2007; a much more negative assessment in 2009; followed by a slight positive trend towards improvements in 2012 – compared with 2009 Meta-analysis (Figure 8).

The quality assessment in the 2007 Meta-analysis was measured against the European Commission's evaluation guidelines at that time, and focused primarily on the product, i.e. the report itself. The 2009 Meta-analysis and this meta-evaluation assessed the quality of the reports as well as the entire evaluation processes, using the *DAC Evaluation Quality Standards (for test phase application)* (OECD 2006) and the DAC/EU Quality Standards (MFA 2011a). The reliability of comparison between the 2009 Meta-analysis and this meta-evaluation is therefore higher and the trend towards improved quality of reporting validated.

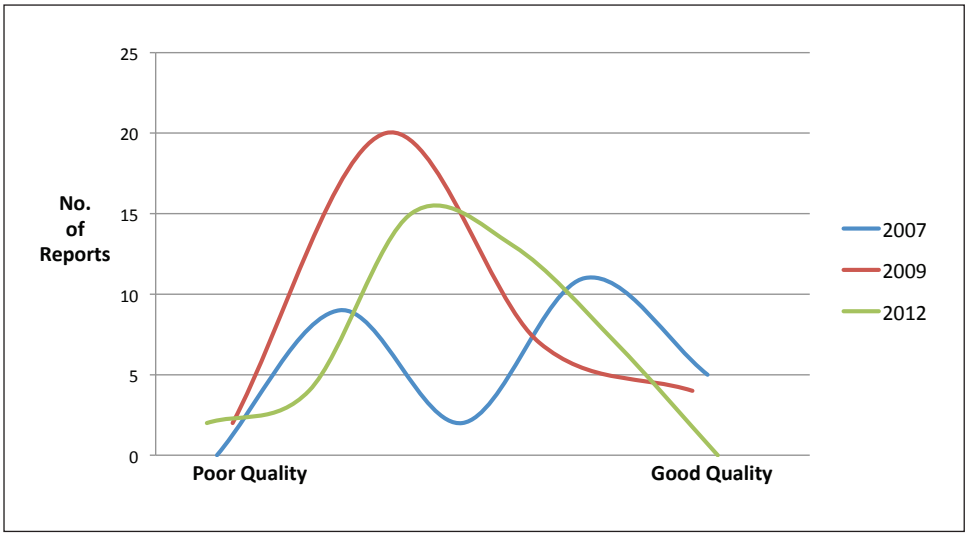


Figure 8 Quality of reports across the 2007 and 2009 Meta-analyses and this meta-evaluation – data normalised.

Source: Meta-evaluation database; White & Stenbäck 2007; Williams & Seppänen 2009.

6.2 Terms of Reference

The 2007 and 2009 meta-analyses had difficulties in accessing TORs, which impacted the use of the sample and their final analysis. It was not the case with this meta-evaluation. All relevant documentation was accessible from the Dropbox server.

In this meta-evaluation it was found that TORs were satisfactorily designed with some flaws related to prioritisation and numbers of evaluation questions. Similar findings were identified in the 2009 Meta-analysis. Also, as in the 2007 and 2009 meta-analyses, reports responded overall well to the content and requests of the TORs.

Important development cooperation issues and policy priorities were reasonably addressed in TORs in this meta-evaluation, including poverty reduction, CCOs and Paris Declaration principles. However, they could have been addressed more systematically. These findings represented a significant improvement from the 2007 Meta-analysis (p 56) in which there was hardly any requirement included in the TORs to assess

human rights, democracy or other CCOs in projects. The 2009 Meta-analysis did not address development policy priorities in relation to the TORs.

The meta-analyses of 2007 (p 45) and 2009 (p 42) stressed the importance of high quality TORs as a precondition for achieving a high quality evaluation. This meta-evaluation did not find correlation between high quality TORs and high quality reporting. Rather, there is a need for identifying other factors than the quality of TORs that influence the quality of reporting. The difference in findings between 2007 and 2009 meta-analyses and this meta-evaluation could to some extent be explained by difference in the use of methodology.

The need for balancing time, budget and scope of work in TORs was an important concern in the 2007 and 2009 meta-analyses, but was not possible to detect sufficiently in this meta-evaluation, due to lack of data. In the 2009 Meta-analysis the most highly rated reports were based on TORs prepared in collaboration with other partners, which was also the case in this meta-evaluation. Overall TORs have improved in quality since 2007.

6.3 Quality of development cooperation

2007 Meta-analysis did not provide scoring of results on the quality of development cooperation and comparison can only be done against narratives. The 2009 Meta-analysis provided a scoring system based on a 1-3 rating and the 2010 Synthesis of evaluations applied a 1-7 score, which was similar to the one used in this meta-evaluation. As such, the latter two can be directly compared and the result is as follows:

This meta-evaluation showed a somewhat more negative assessment for all evaluation criteria (3.4 against 3.8 average score). There was, however, a clear trend for the data in the two analyses: relevance scored highest followed by coherence obtaining the second highest score. Effectiveness, efficiency and sustainability were all identified within the same range of scoring and at a lower level than coherence, and significantly lower than relevance. Complementarity was however lower in the 2010 Synthesis of evaluations than in this meta-evaluation (3.8 against 4.0), and impact much higher (4.3 against 3.2). Impact's high score could be caused by the fact that the data in the Synthesis of evaluations were based on evaluation reports commissioned by EVA-11 only, not including decentralised evaluations, appraisals and mid-term reviews.

The narratives of the 2007 Meta-analysis, the scorings in the 2009 Meta-analysis and the 2010 Synthesis of evaluations report, all confirm the results of this meta-evaluation's assessment of the quality of Finnish development cooperation: poor performance of projects with little effect towards sustained outcomes.

6.4 Cross-cutting objectives

The 2008 CCO evaluation showed that insufficient priority was given to the project design phase, which is crucial for mainstreaming of CCOs. Furthermore, the evaluation showed that the TORs for project preparations did not adequately address CCOs. Projects seldom included specific human or financial resources allocated for CCO implementation. Similarly, projects did not always monitor CCOs, the embassies reported little on them, and the MFA did not demand reporting on them.

To a large extent the findings in the 2007 and 2009 meta-analyses were supportive of the 2008 CCO evaluation findings. It was generally found that only superficial attention was given to CCOs – if they were mentioned at all. CCOs were only given substantial attention in projects where they represented the key priority of the project.

This meta-evaluation found that projects did not systematically integrate and address CCOs. In general the visibility of CCOs was more pronounced in projects targeting poverty reduction, namely in rural development, water and sanitation, forestry management and environment. These projects included a broad spectre of CCOs, e.g. sustainable development (one or more of the three sustainability dimensions of Finnish development cooperation priorities were included), gender equality (different aspects of rights, participation and wealth creation), and democratic governance (decision-making at local and community levels). These projects also often addressed the poorest and the most vulnerable.

The MFA has tried to strengthen the mainstreaming of CCOs, including the issuance of the 2009 Instruction (MFA 2009). Findings in this meta-evaluation indicate that progress has been made towards addressing gender equality in projects. Compared with the findings of the 2009 Meta-analysis progress has been achieved. Allocation of funds to gender focused activities confirms the positive trend (OECD 2012c, 52).

6.5 Paris Declaration principles

Considering the degree by which Finland has been engaged in the overall aid effectiveness process internationally and in the EU, the visibility of the Paris Declaration principles in the reports is low – particularly for managing development results, ownership and mutual accountability. This finding coincides with findings in the 2009 Meta-analysis. It mentions on page 54 that "consideration of the Paris Declaration, or efforts of a programme or project to harmonize and align with government procedures and work together with other donors, was not much stressed in the reports of the projects".

6.6 Use of evaluations, results-based approach and risk management

The 2007 and 2009 meta-analyses found that there was room for improvement in the use of evaluations. They suggested that a standard process for sharing results and follow-up on evaluations, mid-term reviews and appraisals should be established. This meta-evaluation did not find any clear evidence of any systematic follow-up on these evaluation recommendations, not in the reports themselves nor in the additional documents provided for ten of the 41 reports.

Weak project designs and poor result-based management set-ups were recurrent problems described in the reports. This was also found to be a critical issue in the 2010 Synthesis of evaluations (p 60) where project design was rated lowest of 14 criteria applied. The reports assessed in the 2007 and 2009 meta-analyses showed a lack of baseline data, weak indicators and monitoring practice in projects. The 2011 analysis of the result-based approach in Finnish development cooperation found similar

Box 6 Brief considerations on addressing risks and assumptions in projects.

Addressing risks is crucial for development cooperation to be successful. Identifying and managing risks of any kind (developmental, fiduciary or reputational) must be targeted at every level of the project – from policies, strategies to project outcomes, outputs and activities and during the course of the entire project cycle. At the project design stage mitigating risks should be given a particular high priority and risks identified should be incorporated, where possible, into project activities, outputs and/or outcomes. Risk mitigation is the most important activity to be undertaken on a continuous basis throughout the project management cycle to ensure accomplishment of project objectives. Simplified methods and tools for effectively mitigating risks must therefore be tested, developed and applied in project design to ensure project success.

Assumptions – the other side of the coin – are the key elements in establishing a solid theory of change for a project. This should have several advantages, according to a recent study. They include activating critical reflection in response to dynamic contexts, and encouraging on-going questioning of what might influence change in the context. Basically, assumptions of the project theory are to be regularly tested and monitored. Other advantages include drawing on evidence and learning during implementation. At the same time it is acknowledged that the realities of funding and performance management systems in international development make this very challenging to achieve, at all levels (Vogel 2012, 4-5).

Combining this challenge with the *complacency* that stakeholders often display towards assumptions, projects are at risk of failure. The good intentions of emphasising an assumption-based theory of change for effective development performance may therefore not be easily applicable in the real world.

weaknesses (Poate *et al* 2011, 17-19). These weaknesses are unfortunately very common in development cooperation across development agencies. One recent example is evidenced in a report from the British Independent Commission for Aid Impact on EU's aid to low-income countries (ICAI 2012).

Several projects showed inadequate risk analysis. It was also observed in the 2009 Meta-analysis (p 19), where only a few reports referred to focused risk assessments undertaken by projects. Some brief considerations on addressing risks and assumptions in projects are presented in Box 6.

6.7 Summary of key findings

- Development cooperation issues and policy priorities, including poverty reduction and the Paris Declaration principles, were better addressed in the TORs for this meta-evaluation than in 2007 Meta-analysis, though not systematically.
- Current TORs address CCOs significantly better than the TORs in 2007 Meta-analysis, where CCOs were rarely required or mentioned in TORs. However, CCOs are still addressed superficially.
- There was a slight improvement in the quality of reports compared to findings of the two previous meta-analyses.
- When comparing with the 2007 and 2009 meta-analyses and the 2010 Synthesis of evaluations, the trend in the quality of the development cooperation showed poor performance of projects with little effect towards sustained outcomes.
- The visibility of the Paris Declaration principles was low, particularly for managing development results, ownership and mutual accountability. This finding coincided with findings in the 2009 Meta-analysis.
- Weak project designs were a recurrent problem identified in the reports. This was also found in the 2007 and 2009 meta-analyses and in the 2010 Synthesis of evaluations report.
- No evidence was found of overall progress towards increased use of results-based management practices.
- Risk analysis in projects was inadequately addressed, which was also observed in the 2009 Meta-analysis.

7 CONCLUSIONS AND LESSONS LEARNED

7.1 Project design and results-based management

This meta-evaluation showed that results-based approach was either inadequately applied or not used. Projects included weak results frameworks, inadequate definition of result targets at outcome and output levels, indicators not being measurable and

baseline studies not being performed. This meta-evaluation has identified at least two overall factors behind these inadequacies: (a) the experience and skills of development actors engaged in project design, development, implementation and evaluation, i.e. MFA staff, consultants, technical advisers, national staff, etc., and (b) the degree by which national stakeholders and beneficiaries are involved in the project design.

It is important on a continuous basis to strengthen structures and tools for improved project design, implementation and results for all development actors engaged in development cooperation. Only in this way can measurement of effectiveness, sustainability and impact take place and ensure sufficient information for decision-makers to make better judgements and correct activities.

Simplification of the results chain/logframe could provide an effective fast-track mechanism towards improvements in result-based management – if applied in a smart and innovative fashion. The use of innovative tools should not be underestimated in making performance measurement substantially more effective than it is today. Simplification of the results-based tools will strengthen consultants' and MFA/project staff's overall skills in project monitoring and evaluation.

Intervening at an early stage would allow for targeting better goals of poverty reduction and CCOs. This will raise the quality of appraisals, the project document and the monitoring of progress and measurement of achievements.

7.3 Quality of evaluation reports and TORs

Reports assessed in this meta-evaluation lacked information of the evaluation process and were incomplete with regards to the content of the reports. As such, there exists a gap between the present evaluation practice of decentralised reporting and what is perceived as best evaluation practices (=DAC/EU Quality Standards). This gap in performance also confirms that there is an inadequate approval process and no clear quality assurance mechanism in place for the decentralised evaluation reports, which is crucial for their usability. TORs are however satisfactorily designed despite some flaws related to prioritisation and numbers of evaluation questions. Also, while important development cooperation issues and policy priorities are addressed they could be addressed more systematically.

There was no clear correlation identified between quality of reports and quality of TORs. Therefore other factors that influence high as well as low quality reporting should be identified. The inappropriate approval process of decentralised evaluation reporting stands out as an obvious factor. This is confirmed in the *Manual for Bilateral Programmes* which shows no clear or persuasive procedures for effective report approval mechanism for appraisals (MFA 2012b, 54-57). While some factors may very well relate to the institutional and human resources challenges and obstacles in the MFA, other factors could, for example, include the role of the evaluation team and its technical and/or reporting skills; the relationship between the significantly high

number of TORs for appraisals and poor reporting; and the relationship between high quality TORs and the number of reports that are based on joint multi-donor evaluations.

7.4 Quality of development cooperation

The quality of the development cooperation over time shows the same picture: relevance has the highest score, coherence second highest, followed closely by complementarity and then a group ranked somewhat lower comprising effectiveness, efficiency, sustainability and impact. This indicated poor performance of project design and implementation with little effects towards sustained outcomes. The relatively strong correlation between weak project implementation and results, and the lack of effects towards change confirms the need for improved project design and results framework. The meta-evaluation concluded the following:

Relevance and Poverty issues: If real needs and priorities of main beneficiaries are not adequately targeted poverty reduction cannot be realistically tackled. This reduces project relevance and therefore important development cooperation objectives, e.g. poverty reduction, become indistinct. In assessing relevance particular attention should also be given to detect any changes or drifting in the project objectives and purpose.

Coherence: The issue of mainstreaming policies, especially cross-cutting policies (gender equality, human rights or environment) makes sense only if horizontal policy coherence is addressed in projects. As mainstreaming important policies is a key development cooperation focus, the particular complexities of coherence for development should be properly addressed.

Complementarity: The lack of attention paid to complementarity in appraisals is likely to backfire during project implementation. Parallel project structures and lack of coordination in development cooperation have previously demonstrated unsustained development.

Sustainability: Sustainability of project results is weakened by insufficient risk analysis. Participation, awareness and commitment should not on their own merits justify the initiation of a project or the continuation of on-going projects into a new phase.

Effectiveness and Impact: A strong and more direct impact on poverty reduction is best achieved through funding to projects in water and sanitation, rural development, community forestry and environment. Lessons learned and best practices from these sectors should be identified.

Cross-cutting objectives: If CCOs are not addressed through specific outputs they are less likely to be given priority. Considering the emergence of CCOs in recent years as significant drivers of change towards poverty reduction and sustainable de-

velopment in Finnish development cooperation, commitment should be confirmed in project designs and budgets.

Use of evaluations: Recommendations from mid-term reviews and evaluations should be addressed systematically in projects. Only in this manner is value for money guaranteed and learning for future projects or phases enhanced significantly.

7.5 Lessons learned

Findings from this meta-evaluation should be understood and applied within a broader framework of institutional and human resources patterns across organisations (MFA, consultants and partner country organisations). In this way the interconnectedness between the quality of reports, TORs and the use of evaluations for supporting effective development cooperation could be identified better and corrective measures more effectively applied.

Development cooperation results will remain poorly measured in years to come if findings from this meta-evaluation are not seriously addressed. The Finnish MFA should strive to create an enabling environment for a results-oriented approach among staff to planning, implementation and measurement of results. This will include identification of new and innovative tools, building of staff capacity, initiate structural changes and adjustments, e.g. in the organisational environment, incentives and resources. A prerequisite for this to be successful will, first and foremost, depend on willingness and ability of the MFA leadership to facilitate such a process.

8 RECOMMENDATIONS

Recommendations have been consolidated across the report structure and organised according to key areas of concern for the future quality of Finnish development cooperation.

Strategic considerations

1. Establishment and maintenance of a **coherent and system-wide quality assurance** framework in the Ministry for Foreign Affairs (MFA) for improving quality in development cooperation – with a special focus on simplified project design, results-based approach, risks and reporting.

2. **Willingness of MFA leadership** is required to develop and maintain quality assurance. This must include testing and applying new, innovative approaches and tools – supported by institutional changes, human resource management and staff training.

Project design and results-based approach

3. Existing tools and guides related to result-based management should be reviewed and improved, e.g. *Manual for Bilateral Programmes*.

4. **Simplification** of the results chain/logframe should be targeted on a **pilot** basis. This will enable MFA/project staff and consultants to test new, easy and time-saving tools for project monitoring and results achievements. Simplifications should include: phrase output narratives in generic terms, apply easy Quantity, Quality and Time indicators, strengthen risk mitigation by incorporate risk issues into project activities, etc.

5. MFA staff and consultants must strengthen their project **design skills** and MFA its **approval procedures**.

6. The decentralised evaluations under the responsibility of the regional departments and embassies should effectively deal with project design discrepancies, particularly during the **early stages** of the project cycle. The Quality Assurance Group must provide effective supervision of adherence to quality and approval procedures.

7. **Project beneficiaries**, particularly the poor and vulnerable and their institutions, should have absolute priority in project design. They should constitute the basis upon which policy priorities and country strategies are developed.

8. **Specific tools**, e.g. classic grounded theory and participatory appraisals, should be used during project identification and formulation to identify the **main concerns** and real needs of beneficiaries.

9. TORs for mid-term evaluations should always include an assessment of **changes** to objectives and the consequences hereof to project results chain/logframe.

10. **Horizontal policy coherence** must be strengthened in project design to support improved mainstreaming of policies. The OECD guidelines for policy coherence for development should be applied in this process.

11. Complementarity must be thoroughly addressed at the **early stages of the project design** to support harmonisation and avoid duplication.

12. MFA/project staff and consultants must undertake thorough **risk analysis** and avoid complacency when dealing with assumptions. Effective risks mitigation measures must be applied. Influential and determining factors that support sustainability of outputs and outcomes must be identified.

13. When justified for continuation of a project, participation, awareness and similar processes should be **clearly linked to the output and outcome levels**.

14. MFA should consider increasing its financial support to those sectors that prove to achieve **high effectiveness on reducing poverty**.

15. Prioritised CCOs should be incorporated into all projects, at outcome and output levels, with a designated **budget line**.

16. Improved management tools for the use of evaluations should be developed. **Standardised** management response, follow-up on recommendations and back-reporting on the decentralised evaluations should be designed.

Quality of evaluation reporting

17. **Capacity building support** should be provided to MFA/project staff and consultants on how to comply with the DAC/EU Quality Standards.

18. MFA must review its **approval and quality assurance procedures** for decentralised evaluation reporting.

19. **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.

Terms of reference

20. Current TOR guidelines should be reviewed and clear instructions prepared on how to use evaluation criteria and questions. Depending on the scope and budget allotted for an appraisal, a mid-term review or an evaluation, focus should be on **prioritising** development and evaluation criteria and questions.

Meta-evaluation recommendations

21. Recommendations from this meta-evaluation should be integrated into the MFA **case management system** (AHA), country programming, country strategy papers and feed into **policy guidelines and tools**.

REFERENCES

References contain also references used in the annexes.

Brookland J 2012 *A wave of support for WASH*. 21 August 2012. Available: https://www.devex.com/en/news/a-wave-of-support-for-wash/79034 (accessed 09.12.2012).

Caldecott J, Halonen M, Sørensen SE, Dugersuren S, Tommila P & Pathan A 2010 *Evaluation of the Sustainability Dimension in Addressing Poverty Reduction: Synthesis of Evaluations*. Evaluation Report 2010:4. The Ministry for Foreign Affairs of Finland, Kopijyvä Oy, Jyväskylä, 138 p. ISBN 978-951-724-874-7 (printed).

Chambers R 1992 *Rural appraisal: rapid, relaxed and participatory*. IDS Discussion Paper 311, October, 68 p. Available: http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/774/Dp311.pdf?sequence=1. (accessed 27.11.2012).

EuropeAid 2012 *ROM Handbook. Results-oriented Monitoring*. European Commission (EuropeAid Co-operation Office) April 2012, EuropeAid EVA/129-719, 112 p. Available: http://ec.europa.eu/europeaid/how/ensure-aid-effectiveness/documents/rom_handbook2011_en.pdf (access 27.11.2012).

EU 2007 *Evaluation Standards*. Available:
http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/standards_c_2002_5267_final_en.pdf (accessed 28.11.2012).

EU 2012 *Code of Conduct on Complementarity and the Division of Labour in Development Policy*. Available: http://europa.eu/legislation_summaries/development/general_development_framework/r13003_en.htm. (accessed 27.11.2012).

Glaser BG 1992 *Emergence vs. Forcing. Basics of Grounded Theory Analysis*, Sociology Press, Mill Valley, CA, 1992, 128 p.

ICAI (Independent Commission for Aid Impact) 2012, *DFID's Oversight of the EU's Aid to Low-Income Countries*. Report 17, December 2012, 42 p. Available: http://icai.independent.gov.uk/wp-content/uploads/2012/12/ICAI-EU-report-061212-FINAL.pdf (accessed 11.12.2012).

Kääriä T, Poutianen P, Santisteban R, Pineda C, Munive A, Pehu-Voima S, Singh K & Vuorensola-Barnes S 2008 *The Cross-cutting Themes in the Finnish Development Cooperation*. Evaluation report 2008:6. Ministry for Foreign Affairs of Finland, Hakapaino Oy, Helsinki, 93 p. ISBN 978-951-224-714-6 (printed).

MFA 1999 *Guidelines for Programme Design, Monitoring and Evaluation*. The Ministry for Foreign Affairs of Finland, Hakapaino Oy, Helsinki, 69 p. ISBN 978-951-724-624-8.

MFA 2007a *Development Policy Programme 2007: Towards a Sustainable and Just World Community*. Government Decision-in-Principle. Ministry for Foreign Affairs of Finland, Erweko Painotuote Oy, Helsinki, 39 p.

MFA 2007b *Evaluation Guidelines. Between Past and Future*. Ministry for Foreign Affairs of Finland, Helsinki, Hakapaino Oy, 2007, 84 p. ISBN 987-951-724-624-8 (printed).

MFA 2009 *Instruction; Cross-cutting themes in Development Policy Programme of the Government of Finland*. An unofficial translation from Finnish to English from the original Finnish document [MFA 2009 *Kehityspolitiikan läpileikkaavat tavoitteet; Toimintaohje*. Ministry for Foreign Affairs of Finland, HEL7251-63, 18.6.2009, 4 p.]

MFA 2011a *DAC/EU Evaluation Quality Standards and EU Quality Assessment of an Evaluation Report*. Development Evaluation (EVA-11), a working tool, dated 20.01.2011.

MFA 2011b Ministry for Foreign Affairs standing order for development cooperation evaluations. By-law, Norm 4/2011, HEL8328-35, 4 p.

MFA 2011c *KYT Evaluation; Back reporting on CC-themes*. An unofficial translation from Finnish to English from the original Finnish document. [MFA 2011 *KYT Evaluointi; Seurantaraportti: Läpileikkaavat teemat (LLT) Suomen kehitysyhteistyössä, 2008:6*. Ministry for Foreign Affairs of Finland, HELM120-15, 30.08.2011, 6 p.]

MFA 2012a *Finland's Development Policy Programme*. Government Decision-in-Principle, 16 February 2012. Ministry for Foreign Affairs of Finland, Erweko Painotuote Oy, Helsinki, 48 p.

MFA 2012b *Manual for Bilateral Programmes*. Ministry for Foreign Affairs of Finland, 98 p. Available: http://formin.finland.fi/public/default.aspx?contentid=259204&nodeid=34606&contentlan=2&culture=en-US (accessed 26.12.2012).

MFA 2012c *Good Governance and Anti-Corruption*. Available: http://formin.finland.fi/public/default.aspx?nodeid=43937&contentlan=2&culture=en-US (accessed 27.11.2012).

MFA 2012d *Anti-Corruption Handbook for Development Practitioners*. Ministry for Foreign Affairs of Finland, Department for Development Policy, Erweko Oy, 2012, Helsinki, 214 p. ISBN 978-952-281-026-7.

MFA 2012e *Development Evaluation. Bi-Annual Report 2011-2012*. Ministry for Foreign Affairs of Finland, Kopijyvä Oy, Jyväskylä, 48 p. ISBN 978-952-281-050-2 (printed).

MFA 2012f *Human rights principles and normative criteria*. Available: http://formin.finland.fi/public/default.aspx?nodeid=46440&contentlan=2&culture=en-US (accessed 24.12.2012).

OECD 2002 *OECD Glossary of Key Terms in Evaluation and Results Based Management*. Paris 2002/2010, 38 p. Available: http://www.oecd.org/dataoecd/29/21/2754804.pdf (accessed 27.11.2012).

OECD 2006 *DAC Evaluation Quality Standards (for test phase application)*. Organisation for Economic Co-operation and Development. Development Assistance Committee, 8 p. Available: http://www.mfcr.cz/cps/rde/xbcr/mfcr/DAC_Evaluation_Quality_2006.pdf (accessed 27.11.2012).

OECD 2012a *Policy Framework for Policy Coherence for Development*. Working Paper no. 1, 2012. OECD Office of the Secretary General. Unit for Policy Coherence for Development, 46 p.

OECD 2012b *DAC Criteria for Evaluating Development Assistance*. Available: http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html (accessed 27.11.2012).

OECD 2012c *Finland. Development Assistance Committee (DAC)*. Peer Review 2012. Available: http://www.oecd.org/dac/peerreviewsofdacmembers/PRFINLAND2012.pdf (accessed 27.11.2012).

Poate D, Bartholomew A, Rothmann I & Palomäki A 2011 *Result-Based Approach in Finnish Development Cooperation*. Evaluation Report 2011:2. Ministry for Foreign Affairs of Finland, Kopijyvä Oy, Jyväskylä, 93 p. ISBN 978-951-724-941-6 (printed).

Sørensen SE 2010 *Poverty Assessment and Vulnerability in Greenland* (in Danish). Greenland Self-Rule Authority. Available: http://mipi.nanoq.gl/~/media/2ecd6fbd202f4ffd94a7f0ec3ed43571.ashx (accessed 27.11.2012).

UN 2012 *The future we want*. Draft resolution submitted by the President of the General Assembly, United Nations A/66/L.56, General Assembly Distr.: Limited, 24 July 2012, 12-43688 (E) 250712, **1243688*,* Sixty-sixth session, Agenda item 19, 54 p. Available: http://www.un.org/en/ga/president/66/Letters/PDF/Rio+20%20Draft%20Resolution-24%20July.pdf (accessed 27.11.2012).

Vogel I 2012 *Review of the use of "Theory of Change" in international development*. Review Report. UK Department for International Development (DFID), April 2012, 82 p. Available: http://www.dfid.gov.uk/r4d/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf (accessed 24.12.2012).

White P & Stenbäck T 2007 *Meta-Analysis of Development Evaluations in 2006*. Evaluation report 2007:2. Ministry for Foreign Affairs of Finland, Hakapaino Oy, Helsinki, 88 p. ISBN 978-951-724-632-3 (printed).

Williams PC & Seppänen M 2009 *Meta-Analysis of Development Evaluations in 2007 and 2008*. Evaluation report 2009:9. Ministry for Foreign Affairs of Finland, Hakapaino Oy, Helsinki, 85 p. ISBN 978-951-724-809-9 (printed).

Wood B, Etta F, Gayfer J, Kabell D, Ngwira N, Sagasti F & Samaranayake M 2011 *The Evaluation of the Paris Declaration, Phase 2*. Final Report. IOD/PARC, Aarhus, May 2011, 231 p. ISBN 978-87-7605-436-6.

THE EVALUATION TEAM

Svend Erik Sørensen (Team-leader) is an economic historian and sociologist. He is an experienced evaluator, having led major reviews and evaluations. He led the Evaluation of United Nations Development Programme's (UNDP) 2007-2013 Country Programme in Croatia (to be published in March 2013); in 2007 he led the Evaluation of Urban Development and Environmental Support to the Western Balkans (2000-2006), for Swedish International Development Cooperation Agency (Sida); and in 2006 the Outcome Evaluation of UNDP's Good Governance and Justice Programme in Asia and the Pacific. He is well familiar with Finnish development cooperation, has performed peer reviews and was a team member of the *Synthesis Evaluation of Sustainability in Poverty Reduction* in 2010.

His evaluation expertise is based on his long practical experience in development cooperation. From 2000 to 2004 he was Process Consultant on project design, implementation and evaluation for regional youth, media and human rights networks in the Danish supported Peace and Stability through Transboundary Civil Society Collaboration in South East Europe; Survey Manager for the 2004 Vulnerability and Poverty Assessment Survey of all 200 inhabited islands of Maldives for the World Bank; Project Manager of United Nations Environment Programme's (UNEP) tsunami recovery programmes in Indonesia and Maldives in 2005; Institutional and Human Resources Development Specialist on risk analysis for EU's Programme for Improvement of Border Relations between Ukraine and Moldova (EUBAM) in 2007-2008, and the Principal Researcher for a 9-months comprehensive poverty study in Greenland in 2009-2010.

Casper Thulstrup (Team member) is a political scientist. For the past 10 years he has been working with quality assurance of project and programme design, monitoring and evaluation, policy and sector analysis and performance management. He is a results-based management specialist and an experienced project manager with excellent skills in the use of qualitative and quantitative techniques such as interview, observation, case study, logical framework, questionnaires and web-based surveys, data-analysis, etc. His main areas of expertise include environmental governance, climate change, resource efficiency, education, policy coherence, institution and capacity building and democratisation. He has worked for UNEP and the Danish Ministry of Foreign Affairs. In his work he has carried out a large number of evaluations and analyses, provided analytical support to senior management, supported the project and programme managers in design of projects and reporting on results, developed programme performance frameworks and provided progress reports to programme stakeholders.

Elisabeth Lewin (Quality Assurance team) has a Master of Business Administration (MBA) from the Stockholm School of Economics and post-graduate courses in Evaluation Methodology, at Carleton University, Ottawa, Canada. She has over 20

years experience in leading positions at the Swedish International Development Cooperation Agency (Sida). Her most senior management position was that of Director of the Department for Latin American and member of Sida's Management Committee. She has also served as Head of the Evaluation Section, and Director of Multilateral Coordination, i.e. Sida's coordination with the United Nations, EU, and the World Bank. Throughout her career much of her work has focused on policy frameworks, the formulation of development strategies, and the project cycle. She is the sole author of Sida´s evaluation manual which was used by the agency worldwide 1993 – 2004. She has frequently conducted training in evaluation methodology. She is now an independent consultant undertaking evaluation assignments for international development agencies.

Stefanie Bitengo (Quality Assurance team) is a trained social worker with an advanced degree in Development Studies. She has worked in the field of community development for over 20 years, garnering a wealth of experience through undertaking consultancies in both governmental and donor funded projects in the Eastern Africa region, including Kenya, Rwanda and Ethiopia. Her work expertise covers the entire project management cycle, i.e. planning, monitoring, evaluations and reporting. She has progressively offered leadership and quality assurance in areas of participatory planning, organization development, project/programme targeting, and overall efficiency and effectiveness in project implementation. She has held senior positions both in government and donor funded projects, in which she has sharpened her skills in overall project management, through development of project implementation manuals, monitoring and evaluation tools, reporting frameworks as well as quality assurance tools. She is an experienced facilitator and the co-author of the Group Development Section for International Fund for Agricultural Development (IFAD) supported Central Kenya Dry Area Smallholder and Community Services Development Project (CKDAP) Publication on project community development initiatives.

Dorte Kabell (Quality Assurance team) is an economist with a social science interest and more than 25 years' experience in international development. She has worked in UNDP and OECD and held senior positions in the Asian and African Development Banks respectively, including senior adviser to the President of the African Development Bank, and later as Director of Cabinet. She was instrumental in pioneering the Quality Assurance Unit in Danish International Development Agency (Danida), including being responsible for the development of the multilateral performance management framework. She specializes in methods and processes that generate results-oriented and evaluative knowledge for strategic purposes. Assignments include the development of a monitoring and evaluation framework for leadership development for the World Bank, various policy and organisational evaluations for the African Development Bank, and *The Evaluation of the Paris Declaration*, the latter granted the American Evaluation Association's Outstanding Evaluation Award in 2012.

ANNEX 1: TERMS OF REFERENCE

Developing countries, Meta-analysis of Decentralised Development Evaluations in 2010 and 2011

1. Background to the evaluation

The evaluation function of the Ministry for Foreign Affairs of Finland (MFA) is divided into the central evaluations, meaning those carried out by the Development Evaluation (EVA-11) attached to the office of the Under-Secretary of State for development, and into the decentralised evaluations, which are done by the regional and other departments and units and in some cases also by the embassies. Thus, in addition to the 5-7 evaluations performed by the central system, MFA commissions a large number of evaluations, mid-term evaluations and other assessments of aid, which target specific projects, programmes or restricted topics.

Since autumn 2010 EVA-11 has offered help-desk services in evaluations and organized regular trainings. The objective is to provide basic knowledge of and understanding on evaluation of development cooperation. This includes the introduction of basic concepts and quality standards of development evaluation by the OECD/DAC and the EU. Through the training the participants gain practical skills for preparation of high quality evaluations and use of evaluation results for learning and accountability.

The primary goal of Finnish development policy of 2012 is the eradication of extreme poverty in accordance with the UN Millennium Development Goals. In addition, the key objective is wide-ranging and value-based cooperation that ensures meeting the needs of the most vulnerable. The priorities in the activity include democratic and responsible society promoting human rights; green economy that empowers people and provides employment; sustainable management of natural resources and environmental protection; and humane development. The three cross-cutting objectives to be observed in all activities are promotion of gender equality, climate sustainability, and reduction of inequality. These issues were also central in the 2007 and earlier development policies.

Results of previous evaluations

There have been four meta-analyses of evaluations done in: 1) 1991, 2) 1996 (covering the years 1988-1995, 150 reports), 3) 2007 (covering the year 2006, 29 reports) and 4) 2009 (covering the years 2007 and 2008, 33 reports). When the conclusions of first three meta-analyses were compared in the 2009 meta-analysis, fairly little improvement in terms of sustainability and impact were observed.

The synthesis evaluation (Sustainability in Poverty Reduction: Synthesis 2010:4) summarizing the findings of 22 wider evaluations commissioned in 2008-2010 by EVA-11 recommended to evaluate not only the quality of development cooperation and evaluations, but also how the quality has developed and how the lessons learnt and improvements can continue to be made; review regularly the implementation and recommendations from evaluations and systematically disseminate lessons learnt and best practices; identify indicators that explicitly link with partner country priority MDGs and poverty reduction strategies; and ensure that the proxies of sustainability are routinely assessed during appraisals and monitoring. The evaluation also recommended that the cross-cutting objectives should be defined as *critical enabling factors*.

The evaluation of cross-cutting objectives (including human rights in all of its aspects with special attention to the rights of the most vulnerable groups like children and indigenous peoples and minorities; women's rights and gender equality; and democracy, good governance and rule of law) was completed in 2008. A management decision on the implementation of the recommendations of the evaluation was affectuated in 2008. According to it the relation between the cross-cutting objectives and other regulating principles and how to promote cross-cutting objectives in practice should be clarified. It also recommended trainings for operationalizing the cross-cutting objectives for everyone in the ministry. The Instruction on how to take into account the cross-cutting themes in implementation of the 2007 Development Policy was issued and different trainings were organized as a follow-up of the evaluation.

The back-reporting on the materialization of the results was done in 2011. It showed that progress has been achieved, for example the cross-cutting objectives have been integrated into the new AHA information management system and it gave credit for wider evaluations having cross-cutting objectives always in agenda. To achieve more progress, directions of the responsibilities and systematic accountability how to promote the cross-cutting objectives at all levels of the ministry are required, as well as it is important to agree and have instructions how to integrate cross-cutting objectives to mandates of country negotiations and memorandums of objectives, etc. Keeping this in mind, it is of atmost interest to assess, how these central issues have been adreressed in the 2010-2011 decentralized evaluations.

2. Scope of the Meta-Analysis

This meta-analysis will be a desk study. It shall assess and collate information contained in the decentralized evaluation reports from years 2010 and 2011. The sample contains appraisals (about 1/3), mid term reviews (almost 1/2) and final and *ex-poste* evaluations (about 1/4). A preliminary categorization shows that the sectors evaluations cover most are rural development, environment and water 18% each, forest 15 % and other sectors less than 10 % each. The best presented countries in this selected material are Vietnam, Mekong (regional), Kenia, Tanzania and Lao PDR.

The evaluations for this analysis have been selected so that different units and sectors will be presented to the extent possible. (Tentative list of selected evaluation reports

for this meta-analysis Annex 1). The total number of evaluation reports to be analyzed will be no more than 45. The evaluation material should be analyzed both 1) by grouping by types of evaluation and 2) as an overall synthesis including all reports.

In addition, there will be about ten evaluation reports, selected from the reports included in this study, with other documentation, including the project documents, annual reports, minutes of the steering committee and supervisory board meetings and possible previous evaluation reports (Tentative list Annex 2). The majority of this documental material will be provided to the evaluation team. Nevertheless, the team may still need some additional documentation that needs to be retrieved from the archives or internet. Retrieval of this additional material is the responsibility of the evaluation team. The evaluation team should use their own judgment, knowledge base, expertise and experience to ensure the achievement of the objectives and purpose of this meta-analysis.

The cross-cutting objectives of the Finnish Development Policy of 2007 are the following: the promotion of the rights and the status of women and girls, gender and social equality; promotion of the rights of easily marginalised groups, especially children, persons with disabilities, indigenous peoples and ethnic minorities, and the promotion of their equal opportunities of participation, and HIV/AIDS as a health and social challenge. Moreover, the promotion of sustainable development, rights-based approach, democracy, good governance (including corruption), rule of law and human rights, climate change: adaptation and related disaster risk reduction, and risk management are defined as cross-cutting objectives of this meta-analysis.

3. Rationale, purpose and objectives of the evaluation

Rationale

The meta-analysis of evaluations is an excellent means to bring together the otherwise scattered knowledge and lessons learnt from the decentralised and centralised evaluation systems. At the time of emergence of the 2012 development policy and the imminent commitment of the strategic planning of the development programmes, it is important that lessons from the past experience are assessed, made available and utilized.

Purpose

The overall purpose of this meta-analysis is to draw lessons from the decentralised evaluations to benefit development cooperation practices, and widen the scope of evaluation utilities and use of results in the institutional learning.

Moreover, evaluation *per se* is a tool for accountability and increased transparency towards general public, parliamentarians, academia and development professionals outside the immediate sphere of the decision-makers in development policy.

The users of the results of this meta-analysis will be the desk-officers, advisers and decision-makers of the ministry and the wider constituencies in the implementation of development cooperation.

Specific objectives

The specific objectives of the meta-analysis are

· to build a comprehensive overall independent analysis of the quality of Finnish development cooperation, decentralized evaluation reports from 2010 and 2011 and their terms of reference assessed against 13 development criteria;

· to assess the operationalization of the poverty reduction and the cross-cutting objectives as defined in the section 2 of these TORs; and

· to compare the results of the first two objectives to the selection of previous evaluations: Sustainability in Poverty Reduction: Synthesis 2010:4, Cross-Cutting Themes in the Finnish Development Evaluation 2008:6 and Meta-analyses of 2007:2 and 2009:9 to identify possible trends of change and address trends related to best practises, challenges and obstacles.

· Another task of this meta-analysis is to study the ten selected interventions and effects of the evaluations in the practical level.

4. Evaluation criteria

The meta-analysis will use DAC and EU evaluation criteria: *relevance, effectiveness, efficiency, impact, sustainability, coherence, complementary;* the Paris Declaration principles as criteria: *ownership, harmonisation, alignment, managing for results, mutual accountability,* and the additional criteria *the cross-cutting objectives* as spelled out above in section 2.The criterion *impact* has been included, however understanding that in many cases it is not possible to analyze, but whenever possible, the evaluation team may want to display towards possible impact trends. Patterns, trends and best practices should always be pointed out and justifications given for them.

The major applied principle of any of the criteria will be their presumed significance in the relationship with poverty reduction and g objectives.

An evaluation matrix should be constructed and included in the inception report which will attribute the criteria to the major evaluation questions opened up into specific research questions (Section 5).

5. Evaluation questions

The following evaluation questions have been designed to achieve the fulfilment of the purpose and the specific objectives. The list is not exhaustive and the evaluation team may suggest other questions in the inception report or later in the process to achieve the purpose and objectives of this meta-analysis in the best possible way.

The questions of 5.1 Quality of Development Cooperation and 5.2 Quality of Evaluation TORs and Reports are for all the evaluation reports and will be answered based on the evaluation reports and their TORs (except in case of the ten sample interventions all available material should be used to answer all the questions).

Comparing the results of the selection of previous evaluations as defined in Specific objectives should be used with all the questions of 5.3 Trends. 5.4 Additional questions related to ten sample interventions are only for the ten sample interventions with additional documentation.

5.1 Quality of Development Cooperation

The following evaluation questions will be examined within the scope of this meta-analysis defined in section 2.

1) How is the overall quality of cooperation assessed against the DAC and EU evaluation criteria? Do the selected development cooperation modalities enable and facilitate effective implementation of Paris Declaration principles and the Millenium Development Goals? How is the goal of poverty reduction addressed in general and in specific with the most vulnerable? Whose poverty has been reduced and through which measures of modality?

2) Are there any best mechanisms of enhancing, programming and implementing the cross-cutting objectives? What are the major achievements and challenges in a concrete way? Do the evaluations show that the interventions have had any positive or negative effects on cross-cutting objectives?

3) Are concrete results identified and discernible? At what level (output, outcome and impact)? Have adequate indicators been used? Were they specific, measurable, achievable, relevant and time-bound (SMART)? How have the risks been identified and managed?

4) Have the participatory and partnership approaches of beneficiaries and different stakeholders been addressed in projects?

5.2 Quality of evaluation TORs and reports

TORs

5) How does the quality of TORs compare against the DAC and EU evaluation quality standards? Are the TORs conducive to good evaluation reports in addressing the right issues? Do the TORs cover poverty reduction and the cross-cutting objectives? In which way the Paris Declaration is addressed in the TORs?

6) Are the tasks assigned in the TORs reasonable as compared with the time, budget and other resources allocation described in the TORs. Do the TORs give adequate guidance for the performance of the evaluation task?

7) What are the factors of the TORs which facilitate the achievement of the good quality of evaluation reports? Give justifications for factors affecting the quality of analysis of the evaluations. Are there any good examples within the material of this meta-analysis? Are there any unsatisfactory factors?

Evaluation reports

8) Do the evaluations comply with the TORs and fulfill the purposes and objectives of the evaluations? How does the quality of evaluation reports compare against the DAC and EU evaluation standards?

9) Is the overall description of the methodologies of the evaluation reports clear and exact enough for readers to understand how the evaluation analyses have been done? Are the chosen methodologies suitable for the type of tasks commissioned and for seeking answers to the evaluation questions and issues spelled out in the TOR? Is the scope of analyses and assesments performed adequate to fulfill the purpose and objectives of the evaluations?

10) Are the cross-cutting objectives systematically and explicitly mainstreamed into evaluation findings, conclusions and recommendations? Do the evaluations confirm the use of participatory and partnership approaches towards the beneficiaries and different stakeholder groups?

11) Do the evaluations address the Finnish value-added, and how is it defined? What have been the best way and the results of implementation to attain best Finnish value added, if any?

12) Can best practise examples of any kind be identified? If any, justify why or why not?

3.3 Trends

13) Are there any trends discernible in regard to the cross-cuting objectives, compliance with Paris Declaration and the goal of poverty reduction, in terms of the quality of the development cooperation, evaluation reports and their TORs? How these issues are addressed in the evaluation reports and their TORs? Can any lessons be drawn on good practices or unsatisfactory practices? Can some patterns, for example obstacles or achievements be found?

14) Are there any changes towards results-oriented planning and management, inclusion of risks and their management, and regular results oriented monitoring? Are there any trends to be found and are these trends likely to continue? Has there been any discernible changes, for instance in the quality and contests of the TORs and evaluation reports?

5.4 Additional questions related to the ten sample projects (Annex 2)

Here it should be analyzed what has happened to the results of the evaluations, the analysis being based on the evalution reports and other documentation, including the project documents, annual reports, minutes of the steering committee and supervisory board meetings and possible previous evaluation reports. In the following, there are also questions whether the poverty reduction of the most vulnerable and the promotion of cross-cutting objectives were taken into account in baseline data, project document and monitoring.

15) What has been learnt and how the results of the evaluations have been used in the ten sample projects and programmes, as shown by the documentation of the projects' management records? Have the lessons learnt been taken into account in the next phase of projects? What have been the challenges and achievements in implementation of evaluation results? How the results of evaluations have been used, what works, what not and why? How the results of evaluations could have been utilized better?

16) Have adequate and appropriately dissagregated baseline data been available on the poverty reduction and cross-cutting objectives for projects? Are the data disaggregated by gender, age, income or other qualities? Have clear cross-cutting objectives been factored in budget, objectives, activities and monitoring indicators?

17) Are the indicators SMART in the project documents? How have the monitoring and data on progress been recorded?

6. Methodology

The meta-analysis shall use multiple methods with defined tools to draw findings and conclusions and formulate recommendations. The more precise methodology of this meta-analysis shall be defined in the inception report at the outset of the evaluation. The methodology shall contain the description of analytical methodologies, including the methods, indicators, judgement points and tools of analyses, scoring or rating systems and alike. Validation of results must be done. No single statements should be taken as a general outcome. All results must be factually based with a verifiable trail of evidence.

7. The evaluation process

There will be an initial start-up meeting to which the evaluators will prepare a short note where they summarize their understanding of the TORs and present their initial approach to the evaluation task. Thereafter, the evaluation team shall prepare an inception report within three weeks from the start-up meeting. Within five months later the draft report shall be ready for comments. After receiving the comments, that will take about 10-14 days, the team has four weeks to submit the final report.

8. Reporting and deliverables

The following deliverables will be submitted. Each deliverable is subject to written approval by EVA-11. The evaluation process will advance only after the approval has been granted. Instructions to the Authors of the Evaluation Reports of the Ministry will be made available to the evaluation team in the start-up meeting. The deliverable reports included in this evaluation shall follow those instructions.

– **Start-up note for the initial meeting**
  The meeting shall be organised as soon as the team has been identified and will be available. The objective of this meeting is to discuss through the TORs, the start-up note and the initial approach, the evaluation process, reporting and administrative matters. The evaluation team will also have an opportunity to pose questions on this assignment. The start-up note will be further elaborated into the inception report.

– **Inception report**
  The Inception report will be submitted as specified in section 7. The inception report shall suggest an outline for the final report and contain the precise methodology as specified in section 6, as well as detailed division of labour within the team, any critical issues of the evaluation and a time-table, with an indication of the dates of deliverables. The inception report should be kept short and concise, no more than 20-25 pages, annexes excluded.

– **Draft report**
  The draft report should be ready no later than five months after the inception report. The MFA and the relevant stakeholders will submit comments on the draft report within about three weeks. The draft report must be of high quality, so that it can be approved by EVA-11 after only one round of comments. The commentary round is only to correct misunderstandings and possible mistakes, not to rewrite any part of the report. The results presented must be supported by evidence. The draft report shall discuss all evaluation criteria and questions and feature the methodology used and define the limitations to the evaluation, what benchmarking has been deployed to arrive at the results described. If the results are only a perception of the team, it should be said so. The draft report shall follow the format of the final report, with abstracts, summaries, references and annexes included. The references, abbreviations and acronyms and other details must be already carefully checked.

– **Final report**
  The final report must be submitted within four weeks after receiving the comments. The final report must follow the Instructions to Evaluation Report Authors. It shall be clear and concise, with carefully checked language that is easily comprehensible by informed ordinary readers and include illustrations, figures, tables, boxes, and clear messages to the potential users of the results of this meta-analy-

sis. In the compilation of the captions to the tables, figures and boxes, the clarity should be kept in mind. The captions must be informative enough that the reader may understand the messages contained in the table, figure or box without reading the report.

The text body should not exceed 60 pages, abstracts, summaries and annexes excluded. Annex 1 is the terms of reference, Annex 2 the people interviewed, other annexes can be added as need arises. The report will be written in English, and include an abstract of no more than 250 words in Finnish, Swedish and English and also a summary in Finnish, Swedish and English. The abstract will be followed by a maximum of five key words to describe the subject of the report.

The report must be proof-read, copy-edited and ready for printing. If the report does not comply with these requirements, it will be returned to the authors for correction. Careless finalization of the report may also cause penalties to the contractor.

– **Individual assessments of each evaluation report**
  Individual assessments of each evaluation report as a separate document will be surrendered to the Ministry at the same time of submitting the final report.

– **Quality Assurance Grid**
  The quality grid as a separate document will be surrendered to the Ministry at the same time of submitting the final report.

– **Oral presentation on the evaluation findings**
  The evaluation team is expected to give a PowerPoint supported presentation on the evaluation findings in a public seminar organized by EVA-11, and the possibility of web based presentation for the wider audience will be considered and confirmed later.

9. Expertise required

The Framework agreement contractor should suggest to the Ministry two (2) international experts with multidisciplinary senior expertise and one (1) junior assistant for practical and organizational matters. None of the experts proposed for this task have had anything to do with the projects, appraisals, mid-term evaluations or final evaluations concerned. A statement of non-bias is required from each of the experts.

The senior evaluation team members should preferably represent gender balance, one male and one female. Preference will be given to the senior consultants from the South, preferably from Finland's long term partner countries. The senior evaluation team must have long-term (more than three years) of residence in a developing country.

The junior assistant to the evaluation team should be resident in Finland, and be available personally to perform the retrieval of archived documents from the archives of MFA in Helsinki and other daily tasks. There can be no additional accommodation or per diem charges for the junior assistant. He/she shall have full command of oral and written Finnish as some of the material is not available in English. He/she should have some experience in evaluations.

The competencies of the team members shall be complementary, but the team leader should have proven experience in four team leaderships of wider development evaluations and track record of analytical skills.

The overall requirements towards the senior evaluation team include relevant long-term experience (minimum of 15 years) in development, including experience in performing different types of development evaluations, including meta-analysis of highly heterogenic material (a minimum of five evaluations each). Proven ability to tease out the essence and conclude on the results in a compact and clear manner is a must.

The evaluation requires senior expertise particularly in the following areas of development:

- ·Results-based planning, monitoring and evaluation, including meta-evaluation.
- Poverty reduction, cross-cutting objectives in general and incompliance with these TORs.
- Operational experience in a multitude of sectors and themes relevant to this meta-analysis in different countries (Annex 3).
- Overall international development issues, development instruments, capacity building and institution building.

Familiarity with Finnish Development Policies is an asset.
The entire team shall have fluency in oral and written English. Spanish or Portuguese is an asset.

Quality Assurance
In addition to the evaluation team, the service provider will nominate two persons, external to the team and the service provider, who are responsible for the quality of the evaluation planning (inception), process and the deliverables. The quality control experts are not members of the team, but their CVs must be presented and their roles explained. They must have earlier proven experience in quality assurance tasks, and be senior of their stature. The quality assurers will fill in the EU Commission's evaluation report quality grid, with inclusion of their justification for their rating of the nine categories. The quality grid as a separate document will be surrendered to the Ministry at the same time of submitting the final report. The grid will also be made available to the peer reviewers of the final product of this meta-analysis, the final report.

10. Budget

The total tentative budget of meta-analysis is 200 000 Euros, VAT excluded, which cannot be exceeded. The service provider will submit the budget with the limited tender against these TORs.

11. Time-table

The evaluation will be carried out between 15.3.2010–15.10.2012.

12. Mandate

The evaluation team has no immaterial rights to any of the material collected in the course of the evaluation or to any draft or final reports produced as a result of this assignment.

The consultants are not authorised to make any statements, commitments or act on behalf of the Government of Finland.

13. Authorisation

Helsinki, March 1, 2012

Aira Päivöke
Director
Evaluation of Development Cooperation
Ministry for Foreign Affairs of Finland

Annexes:
Annex 1: Tentative list of the selected evaluation reports
Annex 2: Tentative list of the ten sample projects
Annex 3: Most represented sectors and countries
Annex 4: List of annexes that will be provided to the team as a package

ANNEX 2: LIST OF REPORTS

No	Type	Project	Sector	Continent	Country/Region	Completion
R1	MTR	United Nations Programme of Catalytic Support to Implement the Convention to combat desertification in the Arab States Region, Phase IV 2009-2012, Middle East & North Africa	Environment	Africa	Algeria, Jordan, Lebanon, Morocco, Syria, Tunisia and Yemen	9/2011
R2	Appraisal	The Information Society and ICT Development Project (TANZICT), Tanzania	ICT	Africa	Tanzania	22.10.2010
R3	Appraisal	Potato Sector Development Project, Tanzania	Rural Development	Africa	Tanzania	22.12.2010
R4	Appraisal	Mama Misitu: Addressing Forest Governance in Tanzania	Forestry	Africa	Tanzania	11.1.2011
R5	MTR	Gender and Governance Programme III in Kenya (2008-2011)	Human rights	Africa	Kenya	12/2011
R6	Evaluation	Mainstreaming MDGs in Kenya's Development Process: MDGs Program Phase I	Other	Africa	Kenya	7/2011
R7	Appraisal	Programme for Agriculture and Livelihoods in Western Communities (PALWECO), Kenya	Rural Development	Africa	Kenya	30.4.2010

R8	Evaluation	Water Supply Trust Fund/Community Project Cycle (CPC), Kenya	Water	Africa	Kenya	9.12.2011
R9	Appraisal	Community Led Water, Sanitation and Hygiene Acceleration (COWASH), Ethiopia	Water	Africa	Ethiopia	20.10.2010
R10	MTR	Special Needs Education programme (2008-2012), Ethiopia	Education	Africa	Ethiopia	12/2010
R11	MTR	Rural Water Supply, Sanitation and Hygiene Program in Beningshangul-Gumuz, Ethiopia	Water	Africa	Ethiopia	9/2011
R12	Appraisal	Improving Food Security in East and West Africa through Co-operation in Research and Education	Rural Development	Africa		16.1.2011
R13	MTR	Education For All (EFA)-Fast Track Initiative (FTI) Country Case Study: Mozambique	Education	Africa	Mozambique	2/2010
R14	Appraisal	Appraisal of the Phase II of the Programme for Luapula Agricultural and Rural Development, Zambia	Rural Development	Africa	Zambia	5/2010
R15	MTR	Environmental and Natural Resources Management and Mainstreaming Program (ENRMMP), Zambia	Environment	Africa	Zambia	2011
R16	Evaluation	Finland Knowledge Partnership on ICT (SAFIPA), South Africa	ICT	Africa	South Africa	2011

R17	Evaluation	Mekong Water Dialogues	Water	Asia	Cambodia, Lao PDR, Vietnam and Thailand	8/2010
R18	Appraisal	Funding Application of Mekong River Commission 2011-2014	Water	Asia	Cambodia, Lao PDR, Vietnam and Thailand	11/2010
R19	Evaluation	Asian Development Bank/Core Environment Programme (CEP), Mekong	Environment	Asia	Greater Mekong Sub-region	2011
R20	MTR	Mekong Private Sector Development Facility (MPDF)	Other	Asia	Cambodia, Lao PDR, Vietnam	3/2011
R21	Appraisal	Strengthening of Environmental Administration and Management at the Local Level in Nepal (SEAM-N) III	Environment	Asia	Nepal	9/2011
R22	MTR	Rural Water Supply and Sanitation Project in Western Nepal (RWSSP-WN)	Water	Asia	Nepal	2011
R23	Appraisal	Sustainable Forestry and Rural Development (SUFORD), Lao PDR	Rural Development	Asia	Lao PDR	30.8.2011
R24	MTR	Sustainable Forestry and Rural Development (SUFORD), Lao PDR	Forestry	Asia	Lao PDR	12/2010
R25	MTR	International Law project, Phase III, Lao PDR	Other	Asia	Lao PDR	3.8.2011
R26	MTR	Technical Assistance for Support to Management of P135 II, Vietnam	Human rights	Asia	Vietnam	11/2010

R27	MTR	Innovation Partnership programme (IPP), Vietnam	ICT	Asia	Vietnam	9/2011
R28	MTR	Development of Management Information Systems for Forestry Sector (FORMIS), Vietnam	Forestry	Asia	Vietnam	9/2011
R29	MTR	Water and Sanitation Programme for Small Towns, Phase II, Vietnam	Water	Asia	Vietnam	11/2011
R30	Evaluation	Rural Development Programmes supported by Finland in Two provinces, Vietnam	Rural Development	Asia	Vietnam	11.1.2010
R31	Evaluation	Trust Fund for Forests (TFF), Vietnam	Forestry	Asia	Vietnam	23.10.2010
R32	MTR	Forests and Forest Management, MAP-FINNFOR, Central America	Forestry	Americas	Central America	2011
R33	Evaluation	Program for Strengthening of Rural Development and Poverty Reduction (FOMEVIDAS I), Nicaragua	Rural Development	Latin Ameria	Nicaragua	2011
R34	MTR	Southeast Asian Climate Change Network Project, United Nations Centre Environmental Programme	Environment	Asia	South East Asia	6/2011
R35	Evaluation	International Council of Human Rights Policy	Human rights	Global	Global	2/2010
R36	Evaluation	Institute for Human Rights and Business	Human rights	Global	Global	26.7.2011

R37	MTR	Programme of Education for Sustainable Development in Western Balkans	Education	Europe	Bosnia & Herzegovina, Montenegro, Serbia and broader Western Balkans	11/2011
R38	MTR	Biodiversity and Ecosystem Services for Local Sustainable Development in the Western Balkans (South East Europe BAP)	Environment	Europe	Bosnia & Herzegovina, Montenegro, Serbia, Albania, Kosovo and Macedonia (FYROM)	2011
R39	MTR	Sustaining Rural Communities and their Traditional Landscapes through Strengthened Environmental Governance in Transboundary Protected Areas of the Dinaric Arc, Western Balkan	Environment	Europe	Montenegro, Albania, Croatia	2011
R40	Evaluation	Organisation for Security and Cooperation in Europe (OSCE), Bishkek Academy's Master of Art Program in Politics and Security, Central Asia	Education	Asia	Kyrgystan, Uzbekistan Tajikistan Kazakhstan, Turkmenistan, Afghanistan	8.12.2011
R41	Evaluation	Environment and Security Initiative (ENVSEC)	Environment	Europe	Southern Caucasus, Central Asia and Eastern Europe (Belarus, Moldova, and Ukraine)	17.11.2010

ANNEX 3: DETAILED METHODOLOGY

1 Evaluation Criteria

The main criteria applied in this meta-evaluation are:

- The DAC/EU evaluation criteria;
- The DAC/EU Quality Standards for high quality evaluation reporting (MFA 2011a).

The DAC/EU evaluation criteria include: relevance, coherence, complementarity, effectiveness, efficiency, sustainability and impact. The definition of each criterion is presented in Table 1.

Table 1 Definition of evaluation criteria as per OECD/DAC and the European Commission.

Criterion	Definition
Relevance	The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor.
Coherence	Coherence entails the systematic application of mutually reinforcing policies across government departments and integration of development concerns to help promote the achievement of the internationally agreed development goals along with other global and national policy objectives.
Complementarity	Complementarity is the optimal division of labour between various actors in order to achieve optimum use of human and financial resources. This implies that each actor focuses its assistance on areas where it can add most value, given what others are doing.
Effectiveness	A measure of the extent to which an aid activity attains its 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.

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, institutionally as well as financially sustainable.
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 […] include the positive and negative impact of external factors.

Sources: OECD 2012a, 3; OECD 2012b; EU 2012.

The DAC/EU Quality Standards comprise one overall assessment and three different phases of the evaluation process:

- Overarching considerations;
- Purpose, planning and design;
- Implementation and reporting;
- Follow-up, use and learning.

Each of those includes several sub-items (criteria) that characterises the requirements and factors to observe. The DAC/EU Quality Standards are applied and discussed in detail in Section 2.

The DAC/EU evaluation criteria provide the internationally recognised guidance for evaluating development cooperation and policies, the DAC/EU Quality Standards the same for evaluating the quality of evaluation reports (including appraisals and mid-term reviews). They both constitute tools that Finland is obliged to comply with as a member state of the OECD and the EU. Additional criteria applied to the meta-evaluation include Finnish development cooperation's policy priorities of cross-cutting objectives (CCOs) and poverty reduction, as well as the Paris Declaration principles, the Millennium Development Goals (MDGs) and Finnish value added.

2 Analytic Framework

The specific objectives of the meta-evaluation cover four major, interrelated themes:

Theme 1: Quality of TORs and evaluation reports;
Theme 2: Quality of Finnish development cooperation;
Theme 3: The use of evaluation findings in projects;
Theme 4: Trends that can be identified from the assessment of decentralised evaluation reports, TORs, and findings in decentralised evaluation reports on Finnish development cooperation.

Theme 1: The DAC/EU Quality Standards define the framework of what constitutes high quality evaluation reports. Each of the 41 reports has been assessed against these criteria, to identify 1) *if* the criteria have been addressed and 2) *how* the criteria have been addressed. Furthermore, the three main types of reports, i.e. appraisals, mid-term reviews and evaluations are all expected to consider strategic priorities in Finnish development cooperation, i.e. policies of poverty reduction and CCOs. The TORs have been assessed against the DAC/EU Quality Standards, the quality of reports as well as selected additional criteria, i.e. poverty reduction, CCOs and the Paris Declaration principles.

Theme 2: The assessment of quality of Finnish development cooperation was assessed against the DAC/EU evaluation criteria. As part of this assessment the following additional criteria were also addressed (a) the strategic priorities in Finnish development cooperation, i.e. policies of poverty reduction and CCOs, and (b) the Paris Declaration principles, the MDGs and Finnish value added.

Theme 3: The DAC/EU Quality Standards also set out the standard for good practice in regards to use of evaluations, dissemination and follow up on recommendations made in reports. Based on an in-depth study of ten selected projects the meta-evaluation attempted to identify patterns and examples of good as well as poor practice (including data handling, project design and monitoring) for the use, dissemination and follow-up activities to the sample reports.

Theme 4: Possible trends have been investigated with the outset in the findings and results from the assessment of the first three themes of the meta-evaluation. This included trends identified from comparing the findings and results with previous meta-analyses and evaluations, assessing any important linkages, possible best practices, and main challenges and obstacles.

As such the four themes provided answers to the evaluation questions presented in the TOR (Annex 1).

3 Approach

The approach of the meta-evaluation included four phases:

Phase 1: Evaluation matrix
To guide the meta-evaluation an evaluation matrix was developed with an outset in the evaluation questions defined in the TOR for the meta-evaluation. Research questions and judgment criteria were defined to outline the key dimension(s) of the evaluation questions. Also, one or more indicators were defined for each judgment criteria to clarify the basis upon which the judgment criteria and research question would be assessed.

Phase 2: Screening of sample reports
Two assessment tools were developed to allow for a screening of the reports and collection of data: Assessment tool 1 for assessing the quality of reports and the quality

of development cooperation, and Assessment tool 2 for assessing the use of findings and the quality of performance measurement data.

Assessment tool 1 was designed to capture information on three issues:

- General background data on the reports and the development activity;
- An assessment of whether or not the reports addressed criteria in the DAC/EU Quality Standards for evaluations;
- The performance of the development activity compared to the DAC/EU evaluation criteria.

The assessment was done in narrative form and supported by a score for each of the DAC/EU evaluation criteria and the DAC/EU Quality Standards. This allowed general comparison across reports. Each of the 41 sample reports was screened individually, using an assessment template. Assessment of the quality of TORs was performed in narrative form and their quality assessed against the narratives and scoring of the reports.

All reports were also screened against the additional criteria (poverty reduction, CCOs, etc.), which were all related to one or more of the seven main criteria, e.g. relevance, coherence, complementarity, efficiency, effectiveness, sustainability and impact. The additional criteria were not given a numerical rating but a degree of "visibility" to which they were considered in the project, i.e. addressed explicitly or implicitly through project objectives, outcomes and/or outputs.

Assessment tool 2 was developed for the analysis of use of evaluation findings and performance measurement data in documents related to ten of the 41 sample reports. The tool included questions addressing five issues related to design, baseline data, CCOs, performance monitoring and the use of the reports. The latter included to which extent actions were taken to dissiminate evaluation results, and if there was any uptake of recommendations in project implementation.

Phase 3: Population of database and consolidation of data
All data from the screening exercise were added to a database to facilitate analysis across the reports. The database was developed in an Excel spreadsheet. With all data contained in one repository, scores and narrative assessments were consolidated across the reports. The database included the following:

Basic facts about each report, its background and presentation of main purpose and objectives, type of evaluation report (appraisal, mid-term-review, evaluation), title of the evaluation report, geographic position of the project, country(ies) involved, sector(s), name of main partner(s), commissioner of the evaluation, project budget, evaluation budget, report date for submission, project start and end. Not all data were accessible.

Each of the 41 sample reports included narratives and scoring data for each of the seven DAC/EU evaluation criteria, the five additional criteria and the 33 criteria of the DAC/EU Quality Standards for assessing quality of evaluation reporting. While efforts were made to assess as many criteria as possible, not all reports qualified to make this possible.

Narrative and scoring data for each of the questions included in the assessment of the ten additional development interventions – Assessment tool 2.

The specific assessment method and scoring systems applied for each of the different set of criteria are presented in their respective sections of the report.

Phase 4: Data analysis

The analysis of data was done in four steps basically following the four main themes outlined in the analytic framework: (a) analysis of quality of TOR and evaluation reports, (b) analysis of the quality of development cooperation, (c) analysis of use of evaluation findings and performance measurement data, and (d) analysis of trends across sample reports and between findings from this meta-evaluation with findings in previous studies.

Steps (a) and (b) included an exploratory and a descriptive analysis for each of the evaluation criteria (i.e. DAC/EU quality standards, DAC/EU evaluation criteria and the additional criteria) based on the narrative assessments across all reports. The purpose of the exploratory approach was to allow findings to emerge from the narrative data in the reports. The purpose of the descriptive analysis was to identify patterns across the different types of reports and within each report type. The results of both analyses were discussed and analysed against the specific evaluation questions relevant for theme 1 and theme 2.

In step (c) the analysis of use of evaluation findings and performance measurement data was done using a case study approach. This meant that an individual analysis was performed for each of the ten selected reports. The case study included an analysis of the supplementary documentation. For each of the ten selected reports an analysis was prepared and findings across the case studies summarised.

The analysis of trends in step (d) was based on findings from the previous three steps (a-c) and an analysis of trends identified in this meta-evaluation compared with findings of four previous studies.

4 DAC/EU Quality Standards

The Quality Standards applied for this analysis comprised 33 criteria: (a) *Overarching considerations* (6 criteria), (b) *Purpose, planning and design* (12 criteria), and (c) *Implementation and reporting* (15 criteria). The fourth phase of the Quality Standards – Follow-up, use and learning – was omitted from the general analysis of the 41 reports and assessed against the ten selected reports (Section 5).

Despite the differences between the evaluation types included in this meta-evaluation (appraisals, mid-term reviews and evaluations), *all* criteria were applied to *all* sample reports to assess whether or not they addressed these criteria. As a guiding principle in the assessment of the reports, it was assumed that all sub-criteria were ***not*** fulfilled from the outset of the assessment. It means that reports had to show narratives of clear descriptions or strong indications of no deficiencies in meeting the stated criteria, in order for it to switch from a pre-set "not addressed" to "addressed". Each of the reports was given a numerical score based on the total number of criteria "addressed" in the Quality Standards. Table 2 defines the qualitative results of the scoring and describes in brief the consequences for the commissioner of the report for each score.

Draft OECD Quality Standards have been in use since 2006 (OECD 2006) and for EU since 2007 (EU 2007). EVA-11converted the OECD/DAC criteria into a matrix tool in 2008 and updated it in 2011 with the inclusion of EU quality standards for evaluation reports (MFA 2011a). Finland has prepared instructions in preparing high quality evaluation reporting in the *Guidelines for Programme Design, Monitoring and Evaluation* (MFA 1999), the 2007 *Evaluation Guidelines* (MFA 2007b) and in different by-laws (norms) issued by the MFA (e.g. MFA 2011b). They all provide a comprehensive and useful framework for understanding the entire evaluation process and the requirements for producing high quality reports. The DAC/EU Quality Standards guiding this meta-evaluation has therefore been based on reference materials used in centralised evaluation assignments and distributed to the decentralised wider circles. As such, the DAC/EU Quality Standards have been parts and parcel of the Finnish evaluation structure during the evaluation period covered by the reports (from late 1990s to 2011).

The meta-evaluation applied the DAC/EU Quality Standards as defined in the MFA matrix tool. It should be noted that the design of the Quality Standards tool included obvious overlaps in issues to be assessed – for example with regard to stakeholders involvement (criteria 1.4, 2.5 and 3.3) and free and open evaluation process (criteria 1.2 and 3.2). This overlap is inevitable as the first part of the Quality Standards relates to all phases of the evaluation process, while the subsequent phases relate to specific stages of the evaluation process. A redesign of the matrix should be considered to avoid overlapping.

Table 2 Scoring system for the quality of evaluation reports applying the DAC/EU Quality Standards.

Number of items "addressed"	Numeric score	Qualitative result	Brief description
28-33	6	Very Good Quality	The report is covering all important issues and be of good value to the commissioner of the report.
22-27	5	Good Quality	The report is sufficiently covering important issues and is valuable to the commissioner of the report.
17-21	4	Adequate Quality	The report is acceptable but need some improvements to be of value to the commissioner of the report.
11-16	3	Inadequate Quality	The report needs significant improvements to be considered of value to the commissioner of the report.
6-10	2	Very Poor Quality	The report is to be re-written to be considered of any value to the commissioner of the report.
0-5	1	Unacceptable Quality	The report should not be considered of any value to commissioner of the report.

Source: Meta-evaluation team.

5 Quality of Development Cooperation

Each of the main evaluation criteria has been given a numerical rating in the assessment of the quality of Finnish development cooperation. The rating represents a summary of an assessment of the extent to which a given project has addressed particular evaluation criteria. The scoring system is based on the European Commission's monitoring practice (EuropeAid 2012, 64). In this system, a score of 'a' means *very good*; 'b' means *good* 'c' means *problematic*; and 'd' means *serious deficiencies*. Allowing for intermediates, seven possible scores can be given using this system. It is transformed into numerical equivalents for analytical purposes (d = 1, c/d = 2, c = 3, b/c = 4, b = 5, a/b = 6, a = 7). This allows mean scores to be calculated across all seven main evaluation criteria for each project and across all projects for each criterion. The scoring system is presented in Table 3.

The total distribution of scoring from a-d was as follows: a = 5%, a/b = 11%, b = 18%, b/c = 24%, c = 28%, c/d = 12% and d = 5%. The meta-evaluation assessment concentrated on addressing the four "outer" scores (a-b and c-d, excluding b/c):

- 'a' and 'a/b' categorised as being "Very good"
- 'b' being "Good"
- 'b/c' being "Neutral"
- 'c' being "Problematic"
- 'c/d' and 'd' categorised as being "Seriously deficient"

Table 3 Scoring system for the quality of Finnish development cooperation.

Rating	Score	Qualitative	Description
a	7	Very good	Considered highly satisfactory, largely above average and potentially a reference for good practice. Recommendations focus on the need to adopt these good practices in other operations.
a-b	6		
b	5	Good	The situation is considered satisfactory, but there is room for improvements. Recommendations are useful, but not vital for the operation.
b/c	4		
c	3	Problems	There are issues which need to be addressed; otherwise the global performance of the operation may be negatively affected. Necessary improvements however do not require major revisions of the operations' strategy.
c/d	2		
d	1	Serious deficiencies	There are deficiencies which are so serious that, if not addressed, they can lead to failure of the operation. Major adjustments and revision of the strategy are needed.

Source: EuropeAid 2012, 64.

REPORT 2010:2 Development Cooperation with Ethiopia 2000–2008
ISBN: 978-951-724-839-6 (printed), ISBN: 978-951-724-840-2 (pdf), ISSN: 1235-7618

REPORT 2010:1 The Transition Towards a New Partnership with Egypt
ISBN: 978-951-724-837-2 (printed), ISBN: 978-951-724-838-9 (pdf), ISSN: 1235-7618

REPORT 2009:9 Meta-Analysis of Development Evaluations in 2007 and 2008
ISBN: 978-951-724-809-9 (printed), ISBN: 978-951-724-810-5 (pdf), ISSN: 1235-7618

REPORT 2009:8 Natural Disasters, Climate Change and Poverty
ISBN: 978-951-724-807-5 (printed), ISBN: 978-951-724-808-2 (pdf), ISSN: 1235-7618

REPORT 2009:7 The North-South-South Higher Education Network Programme
ISBN: 978-951-724-790-0 (printed), ISBN: 978-951-724-791-7 (pdf), ISSN: 1235-7618

REPORT 2009:6 DEMO Finland Development Programme
ISBN: 978-951-724-784 9 (printed), ISBN: 978-951-724-785 0 (pdf), ISSN: 1235-7618

REPORT 2009:5 Finnish Aid in Western Kenya; Impact and Lessons Learned
ISBN: 978-951-724-783-2 (printed), ISBN: 978-951-724-786-3 (pdf), ISSN: 1235-7618

REPORT 2009:4 Meta-analysis of Development Cooperation on HIV /AIDS
ISBN: 978-951-724-769-6 (printed), ISBN: 978-951-724-770 2 (pdf), ISSN: 1235-7618

REPORT 2009:3 Support to Development Research
ISBN: 978-951-724-755-9 (printed), ISBN: 978-951-724-756 6 (pdf), ISSN: 1235-7618

REPORT 2009:2 Agriculture and Rural Development. A Preliminary Study
ISBN: 978-951-724-746 7-(printed), ISBN: 978-951-724-747 4 (pdf), ISSN: 1235-7618

REPORT 2009:1 Finland´s Development Cooperation in Central Asia and South Caucasus
ISBN: 978-951-724-728 3-(printed), ISBN: 978-951-724-729 0 (pdf), ISSN: 1235-7618

REPORT 2008:7 Kosovo Country Programme
ISBN: 978-951-724-716-0 (printed), ISBN: 978-951-724-717-7 (pdf), ISSN: 1235-7618

REPORT 2008:6 The Cross-cutting Themes in the Finnish Development Cooperation
ISBN: 978-951-224-714-6 (printed), ISBN: 978-951-224-715-3 (pdf), ISSN: 1235-7618

REPORT 2008:5 Finnish NGO Foundations
ISBN: 978-951-724-709-2 (printed), ISBN: 978-951-724-710-8 (pdf), ISSN: 1235-7618

REPORT 2008:4 FIDIDA: An Examle of Outsourced Service 2004–2008
ISBN: 978-951-724-690-3 (printed), ISBN: 978-951-724-691-0 (pdf), ISSN: 1235-7618

REPORT 2008:3 Evolving New Partnerships between Finland and Namibia
ISBN: 978-951-724-701-6 (printed), ISBN: 978-951-724-702-3 (pdf), ISSN: 1235-7618

REPORT 2008:2 Local Cooperation Funds – Role in Institution Building of Civil Society Organizations
ISBN: 978-951-724-701-6 (printed), ISBN: 978-951-724-702-3 (pdf), ISSN: 1235-7618

REPORT 2008:1 Finnish Partnership Agreement Scheme
ISBN: 978-951-724-672-9 (printed), ISBN: 978-951-724-673-6 (pdf), ISSN: 1235-7618

SPECIAL EDITION 2008:1 (SWE) FAO: Utmaning till förnyelse. Sammanfattning
ISBN: 978-951-724-670-5 (print), ISBN: 978-951-724-671-2 (pdf), ISSN: 1235-7618

SPECIAL EDITION 2008:1 (FI) FAO: Haasteena uudistuminen. Lyhennelmä
ISBN: 978-951-724-655-2 (painettu), ISBN: 978-951-724-659-0 (pdf), ISSN: 1235-7618

SPECIAL EDITION 2008:1 (ENG) FAO: The Challenge of Renewal. Summary
ISBN: 978-951-724-657-6 (printed), ISBN: 978-951-724-661-3 (pdf), ISSN: 1235-7618

Development evaluation

P.O. Box 451
00023 GOVERNMENT

Telefax: (+358 9) 1605 5987
Operator: (+358 9) 16005
http://formin.finland.fi
Email: eva-11@formin.fi



MINISTRY FOR FOREIGN
AFFAIRS OF FINLAND