

Discussion Paper

No. 132

June 2012

Learn to walk before you run?

Review of methodological approaches
for evaluating coherence in the field
of international cooperation

Niels Keijzer and Jorrit Oppewal

www.ecdpm.org/dp132

Learn to walk before you run?

A review of methodological approaches for
evaluating coherence in the field of
international cooperation

Niels Keijzer and Jorrit Oppewal

Table of contents

- Table of contents i
- List of acronyms ii
- Executive Summary iii
- 1. About this study 1
- 2. Defining and operationalising coherence for evaluation purposes 3
 - 2.1. Defining (levels of) coherence 3
 - 2.2. Evaluating the practice and results of efforts to promote coherence 5
 - 2.3. Relating the concept of coherence to the five DAC evaluation criteria 6
 - 2.4. Situating coherence in the results framework 8
- 3. Methodology 11
 - 3.1. Constructing the inventory of studies 11
 - 3.2. Analysing the studies 12
- 4. Study findings 15
 - 4.1. Overview of the studies 15
 - 4.2. Main findings in relation to the four key questions 15
 - 4.3. Other findings and observations from the studies 25
- 5. Analysis of findings, conclusions and recommendations 30
 - 5.1. Introduction 30
 - 5.2. Analysis of findings in relation to the four research questions 30
 - 5.3. Conclusion and recommendations 33
- Bibliography 35
- Annex 1: definition of research methods listed in table 3 36
- Annex 2: Search terms used and studies analysed 38
- Annex 3: Studies and evaluations analysed 40
- Annex 4: Template used for analysing the studies 42

List of acronyms

CSOs	Civil Society Organisations
DAC	Development Assistance Committee
EU	European Union
IOB	Policy and Operations Evaluation Department, Netherlands Ministry of Foreign Affairs
NGO	Non-governmental Organisation
ODA	Official Development Assistance
OECD	Organisation for Economic Cooperation and Development
PCD	Policy Coherence for Development
RPE	Netherlands government regulation on evaluation and policy information (<i>Regeling periodiek evaluatieonderzoek en beleids-informatie</i>)

Executive Summary

Objectives and methodology

1. This study looks into how coherence has been evaluated in different policy fields and, on that basis, puts forward recommendations on how to improve the measurement of coherence in the field of international cooperation. It does so on the basis of a systematic review of past studies that examine coherence inside and across public policies, paying special attention to the applied research methods. The study addressed four research questions:
 - a. In what ways and to what extent can ‘coherence’ be defined and operationalised for evaluation purposes?
 - b. To what extent can the relation between coherence and effectiveness/efficiency be evaluated (i.e. is coherence additional or complementary)?
 - c. What methods have been used in past studies and evaluations that look into coherence inside or between policies, at what levels (micro, meso, macro), and what are their respective strengths and weaknesses?
 - d. Based on the answers to the first three questions, what practical and methodological dilemmas can be observed with regard to improving the evaluation of policy coherence in the specific field of policies on international cooperation?
2. Based on a structured search of academic journals as well as evaluation reports, 22 studies were identified as the basis for the analysis under research questions 2 and 3. These could be grouped under three separate policy ‘fields’:
 - a. Development cooperation (n=3),
 - b. Environmental safety and improvement of the environment (n=15),
 - c. External action towards fragile states and the 3D approach (n=4).

Main findings

3. In response to the **first research question**, the study finds that there is presently **no widely accepted definition of coherence** in the development evaluation community. This is related to the absence of a well-developed practice in this area, as well as by differences in overall evaluation policies and definitions of key concepts between OECD members. Recent political discussions on Policy Coherence for Development (PCD) clarify that the focus should be on the contribution of different public policies to development outcomes, but at the same time give rise to a multitude of conceptual and methodological questions including the definition of the desired development outcomes.
4. The analysis of 22 studies identified three ‘schools of thought’ on evaluating coherence:
 - a. A **side-effects perspective**: this perspective refers to a general felt need to ensure that a policy does not undermine the achievement of the objectives of other policies. In terms of the operationalisation of coherence for evaluation purposes, from this perspective evaluation of coherence does not seem fundamentally different from regular evaluation of one single policy.
 - b. The **horizontal objective perspective**: here coherence is defined as the extent to which policies alone or together contribute to the achievement of one horizontal objective. The focus is then on how policies relate to each other. Do certain policies contradict each other when it comes to that horizontal objective; can certain policies reinforce each other with respect to the horizontal objective? For evaluation purposes,

this type of definition would require the analysis of inter-linkages and interdependencies between several different policies.

- c. The **trade-off perspective**: this approach places the focus not only on different policies, but also on different objectives. The main question asked by studies approaching coherence in this way is whether and to what extent different objectives are compatible with one another. This approach requires investigating how a policy or combination of policies affects the relationship between objectives.
5. A related finding is that in order to operationalise coherence for evaluation purposes, the most important thing is to clearly **define all the terms in the equation**. Thus, it is not only crucial to define coherence itself, but also to define the different policies and especially their objectives. Leaving any ambiguity in this respect can reduce the possibility to properly assess coherence.
 6. This study could not satisfactorily answer the **second research question** on the extent the relation between coherence and effectiveness/efficiency could be evaluated. The structured analysis of the 22 studies included an assessment of the overall 'level of analysis' of each study, based on Palenberg 2011. Level 2 analysis would correspond to relating observed outcomes to coherent (or incoherent) policies, and being able to say what the outcomes would have been in the absence (or presence) of coherent policies. It was found that none of the reviewed studies was effectively able to reach such conclusions, so no studies were classified as 'level 2 analysis'. A total of 13 studies were classified as 'level 1 analysis' (a basis for improving the interventions studied), while the other nine were classified as 'descriptive analysis'.
 7. A reliable and valid assessment of the relation between coherence on the one hand and effectiveness and efficiency on the other hand requires extensive data on many policies and variables combined with a valid and reliable research design. Furthermore, ideally one would have access to such data through time and across space, in order to be able to use baselines and counterfactuals.
 8. In relation to the **third research question** on research methods, it was found that the 22 studies most frequently used **interviews, document analysis and descriptive statistics**. A separate body of studies principally relied on a quantitative approach, using modelling techniques and regression analysis. Analysing the respective strengths and weaknesses of different methods was challenging due to two reasons. First of all, few if any studies offered much reflection on the benefits and limitations of their methodological approaches. Secondly, methods are not intrinsically useful as the usefulness depends on the way in which the method was applied.
 9. The comparison of the studies showed that studies with a qualitative, interview-based approach generally were able to present a wealth of information, but were not very good at systematically analysing and presenting those data. Conversely, studies using modelling and regression techniques displayed more analytical rigor, but were constrained in terms of the assumptions required for these models, so that possibly important aspects of the equation were not considered at all. In other words, qualitative studies seemed to prioritise validity at the expense of reliability, while quantitative studies seemed to prioritise reliability at the expense of validity.
 10. The studies were relatively evenly divided over the micro-, meso- and macro-levels. Further, it was found that the studies each used two methods on average. Few studies used more than three different methods. Few studies combined quantitative and qualitative approaches; studies were either predominantly quantitative or qualitative.

11. The study's findings in relation to the **fourth and last research question** pointed to a major challenge for improving the evaluation of coherence in international development in terms of setting an **appropriate level of ambition** and determine **appropriate levels of investment** in relation to that ambition. All policies affect the behaviour of societal actors in some way, which in turn always has certain knock-on effects. However, it is simply not feasible to evaluate all inter-linkages, even including the smallest ones, between all policies. One inevitably has to adopt a certain focus and draw a line.
12. A large number of the studies reviewed adopt a **trade-off perspective** and as such examine the relations between different policy objectives. If adopting this perspective in evaluating coherence in international development, evaluations would need to look not only at effects of policies in developing countries, but also consider the effect and purpose of donor country policies in their own domestic economy and society. This could then be combined with the analysis of the effects of these same policies on developing countries. Only then could the trade-off between the different policy objectives (domestic and international development) really be evaluated.
13. However, one could also decide to take the **horizontal objective** of international development as a starting point and then evaluate the coherence of policies within that framework. On a practical level, this could be a superior approach, since it can bring to light policies that are at odds with each other. When such contradicting policies actually share the same overall objective, it is clear that the incoherence needs to be resolved. If the contradicting policies have very different objectives, it is important to properly evaluate their respective impacts as a basis for trade-off decision at the political level.

Conclusions and recommendations

14. The study findings as presented above indicate that the evaluation of coherence is still in an **early and nascent stage**, especially when compared with 'main stream' development evaluation in relation to other criteria such as effectiveness, efficiency and impact. This is not surprising considering the limited amount of investment that has been made to evaluate coherence in international cooperation.
15. The findings of the study provide no basis to conclude that rigorous evaluation of coherence is not possible. The evidence on how coherence is evaluated in other policy areas instead confirms that with additional investments this practice can be shaped in the field of international cooperation. Slow progress has been made in evaluating the impact of donor country 'non-aid' policies on developing countries, and when taking this further helpful clues can be taken from other policy fields. In line with this overall conclusion and in view of existing political commitments to take this further, this study puts forward four recommendations on how this could be done.
16. **Recommendation 1 – Manage expectations and identify feasible steps forward:** in view of the limited investments and evidence base in the field of international cooperation, the study's findings call for further 'management of expectations' in this area and for a more focused discussion on how the increasing call for evaluating coherence can be translated into a feasible path to further developing this emerging field of work. It would not be realistic to assume that evaluations of entire whole-of-government approaches to development will soon approach the quality and reliability of current evaluations of the impact of single development cooperation interventions. Evaluations that measure the trade-offs between domestic and international development objectives at macro-level and from there determine whether policies deliver win-win

situations or affect the trade-offs do not appear a realistic prospect for the time being.

17. **Recommendation 2 – Support and invest in pilot studies to reduce data and methodological deficits:** one element of an incremental approach could be to undertake more exploratory studies from a ‘side-effect’ perspective, which could gain further insights on whether assumptions on (in)coherence can be confirmed at the level of development outcomes, including by support discussions in the DAC to undertake pilot studies in this field. In such studies, it will be important to take account of some of the substance-related findings in this paper. As such, they will need to consider country heterogeneity, as well as heterogeneity within countries.
18. **Recommendation 3 – Explore joint action while seeking flexibility in dealing with mandate limitations:** in order to stimulate such evaluations, it will be important to provide evaluation offices of development ministries the flexibility to invest in them. In many EU member states, development evaluation units have retained a rather narrow mandate up until today, which means that they can only really invest in evaluations of the development policies implemented by their own ministry. In view of international commitments to take further the evaluation of coherence it seems important to look for creative ways to agree to undertake such efforts collectively at the DAC or EU, despite the fact that not all members’ individual mandates would allow for this.
19. **Recommendation 4 – Challenge the international community of evaluation experts:** although evaluating coherence requires a strong political mandate, also to make intended coherence results more explicit as a basis for improved intervention logics, a stronger base for giving shape to such a mandate can be created through involving the community of independent evaluation specialists. Similar efforts were made in making the shift from project evaluations to evaluations of programme-based approaches, including through discussion papers and specialist workshops. Such efforts may be needed again now that a stronger focus is emerging on evaluating development cooperation in relation to other policies.

1. About this study

This paper presents the results of a study commissioned by the Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs (IOB). The study looks into how coherence has been evaluated in different policy fields and, on that basis, to put forward recommendations on how to improve the measurement of coherence in the field of international cooperation. It does so on the basis of a systematic review of past studies that examine coherence inside and across public policies, paying special attention to the applied research methods. The study focuses on the approach to conceptualising and examining coherence in three thematic fields: (1) development cooperation, (2) environmental safety and improvement of the environment and (3) the link between diplomacy, defence and development.

The study's Terms of References identify the following four key questions that are to be answered:

- In what ways and to what extent can 'coherence' be defined and operationalised for evaluation purposes?
- To what extent can the relation between coherence and effectiveness/efficiency be evaluated (i.e. is coherence additional or complementary)?
- What methods have been used in past studies and evaluations that look into coherence inside or between policies, at what levels (micro, meso, macro), and what are their respective strengths and weaknesses?
- Based on the answers to the first three questions, what practical and methodological dilemmas can be observed with regard to improving the evaluation of policy coherence in the specific field of policies on international cooperation?

The study seeks to help evaluators compare the strengths and weaknesses of different research methods that have been used to study coherence and on that basis select methodologies for their future work. Although primarily having a technical and operational purpose, the findings of this study may also inform further policy discussions on the desirability and necessity of mandating and resourcing development evaluation functions to assess the coherence of policies.

The authors would like to thank Paul Engel, Ruerd Ruben and Henri Jorritsma for their comments on earlier versions of this report. The views expressed in this paper are those of the authors only, and should not be attributed to any other person or institution.

The structure of this report follows the order of the four main study questions:

- **Chapter 2** presents the results of a desk study of evaluation policy documents and additional research that analyse the extent to which the concept of coherence can be defined and operationalised for the use in evaluations of international cooperation. Given important differences in the conceptualisation and definition of key evaluation concepts among members of the DAC Development Evaluation Resource Network, this analysis includes efforts to relate coherence to a standard results framework.
- Based on this conceptual framework, **Chapter 3** describes the study's methodology and the main steps made in terms of identifying and analysing relevant studies based on the second and third study questions. These two questions were further operationalised in a standard template filled in for each study (see annex 4) which revolved around four main questions:
 - How was coherence defined and operationalised in the study concerned?
 - What methods were used to collect and analyse the information to evaluate progress in relation to the objective as defined under question 1?
 - At what level(s) did the evaluation focus, i.e. at micro (households), meso (district/provinces) or macro (countries or regions)?

- What is the quality of the analysis, and what are the strengths and weaknesses of the applied methods?
- **Chapter 4** presents the main findings of the analysis of studies, following the order of these four questions. A key aspect of this overview is a matrix of methods that summarises observed strengths and weaknesses in the reports that were identified.
- **Chapter 5** analyses the main findings and on that basis puts forward conclusions and recommendations on how the evaluation of coherence could be improved in international cooperation.

2. Defining and operationalising coherence for evaluation purposes

2.1. Defining (levels of) coherence

The term ‘coherence’ is neither included in the OECD/DAC’s Glossary of Key Terms in Evaluation and Results Based Management, nor is it part of the five key criteria for evaluation: relevance, effectiveness, efficiency, impact and sustainability.¹ Although the term has been increasingly used in development cooperation debates and statements of differing legal standing², **there is no accepted definition and operationalisation of coherence in development evaluation methodology**. This chapter examines this subject to inform further discussion in the development evaluation community, and will do so by approaching the issue in five steps:

1. Exploring the difference between the terms consistency and coherence, in order to come to a more informed definition of the latter concept
2. Analysing different levels at which coherence can be promoted
3. Conceptualising the types of ‘coherence results’ that can be evaluated at two levels, policy formulation and implementation, and how this relates to evaluations covering policy formulations and/or implementation processes
4. Relating the concept of coherence to the five DAC evaluation criteria, and a further analysis of how this relates to efficiency and effectiveness
5. Situating coherence in a theory of change model

A recent paper (Den Hertog & Stross 2011) investigating available definitions of coherence concluded that scholars differ on whether the terms ‘consistency’ and ‘coherence’ are synonyms, and suggest the following distinction:

Consistency	the absence of contradictions within and between individual policies
Coherence	the synergic and systematic support towards the achievement of common objectives within and across individual policies

Coherence is observed through the presence of synergies, while consistency refers to the absence of contradictions.³ Although past documents and policy discussions have treated the two as synonyms⁴, more recently the debate has moved and the OECD now argues that Policy Coherence for Development (PCD), “(...) *goes well beyond minimizing the adverse impact that public policies can have on developing countries*” (i.e. preventing contradictions), and also “(...) *entails the systematic application of mutually reinforcing policies across government departments*” (i.e. creating synergies)⁵. The preference in development policy discussions for the definition of coherence as set out is also reflected by the addition ‘for development’ as part of the PCD acronym.⁶ Efforts to strengthen Policy

¹ See: <http://www.oecd.org/dataoecd/29/21/2754804.pdf>

² E.g. the 2005 European Consensus on Development puts forward the following definition: “We reaffirm our commitment to promoting policy coherence for development, based upon ensuring that the EU shall take account of the objectives of development cooperation in all policies that it implements which are likely to affect developing countries, and that these policies support development objectives.” http://ec.europa.eu/development/icenter/repository/european_consensus_2005_en.pdf

³ The same conclusion was reached in the joint evaluation of EU mechanisms promoting Policy Coherence for Development (Mackie et al 2007).

⁴ Part of the conceptual confusion is also due to the different official language versions of the EU treaties, e.g. the French and German versions referred to coherence, while the English version referred to consistency.

⁵ http://www.oecd.org/faq/0,3433,en_2649_18532957_48787762_1_1_1_1,00.html

⁶ One could also consider the meanings of their negatives, i.e. inconsistency and incoherence. The meaning of ‘inconsistency’ appears to be the exact opposite of the meaning of ‘consistency’. While ‘inconsistency’ could be

Coherence for Development however predate the existence of this specific term, and refers to a felt need to represent the interests of developing countries in the process of revising existing policies and formulating new policies, in order to improve the contribution of these policies to the achievement of international development objectives. While finding its origin in influential NGO campaigns during the 1990s and in international policy discussions in forums such as the UNCTAD, the specific term Policy Coherence for Development only came into vogue among OECD members after 2000, and in EU policy discussions after the adoption of the European Consensus on Development in 2005.⁷

The practice of promoting coherence should be seen as part of the regular process of policy formulation, refinement and change, and as part of a broader goal to improve the effectiveness of policy. As such, the practice of promoting coherence can be distinguished into two types of activities: (1) **Strengthening coherence** when formulating new policies; (2) **Resolving incoherence** when assessing and revising existing policies. In situations where it is not possible to decide which policy prevails, and where full coherence is not possible, this second activity may also include managing political **trade-offs** (Mackie et al 2007).

Coherence as defined above could be confused with the term '**mainstreaming**'. While the exact definition of mainstreaming is debatable, it is typically only concerned with the first activity of strengthening coherence. Mainstreaming efforts generally revolve around 'horizontal objectives' that are given a standing, which should lead to their inclusion in all areas of an organisation's work. The 'chain of command' for promoting coherence in countries such as the Netherlands is similar to promoting mainstreaming, in the case of the Netherlands through (a) an overall objective stated in a policy, (b) a specialist unit without implementation responsibilities in a ministry with a catalysing function, (c) the entire administration expected to contribute to achieving the overall objective and (d) implementation (Uggla 2007). However past political discussions in the European Union point to the need for coherence to maintain the credibility of the EU as a global actor⁸, which underlines the overt political nature and the competing claims associated to promoting coherence that due to its additional focus on resolving incoherencies goes beyond mainstreaming. Mainstreaming is certainly related to coherence in the sense of working towards the achievement of a horizontal objective. The crucial difference between the two concepts is that mainstreaming does not involve the managing of political trade-offs, while that is a crucial element in promoting coherence.

Given the existing EU policy definition provided by the 2005 EU Consensus on Development (and Art. 208 of the EU Treaty), the definition of coherence provided in table 1 above seems most appropriate to further conceptualise and operationalise in an evaluation context. This further operationalisation depends on the context in which it is used, given that coherence in development can be promoted at five different levels (ECDPM and ICEI 2005):

1. **Internal coherence.** Coherence in the policy field itself, which should achieve consistency between its goals and objectives, modalities and protocols.
2. **Intra-governmental coherence.** Coherence across all of the policies and actions of an OECD country in terms of their contributions to development.
3. **Inter-governmental coherence.** Policies and actions should be consistent across different OECD countries (as well as with policies adopted at the EU or in regional organisations) in terms of their

defined as the presence of contradictions between policies, one would use 'incoherent' when it is clear that some policies contribute negatively to the achievement of development outcomes. However in policy discussions the terms are again used as synonyms because often the data is missing to ascertain whether a policy is simply inconsistent or incoherent.

⁷ In view of its purpose this study does not take a detailed look at past and ongoing policy discussions and statements in relation to promoting coherence in international cooperation, which are for instance discussed in more detail in Keijzer 2010.

⁸ See http://ue.eu.int/uedocs/cms_data/docs/pressdata/EN/foraff/130225.pdf

contributions to development, to prevent one from unnecessarily interfering with, or failing to reinforce, the others.

4. **Multilateral coherence.** Coherence of the policies and actions of bilateral donors and multilateral organisations, and to ensure that policies adopted in multilateral fora contribute to development objectives.
5. **Developing country coherence.** Developing countries should be encouraged to set up policies that allow them to take full advantage of the international climate to enhance their development.

2.2. Evaluating the practice and results of efforts to promote coherence

At these different levels, evaluations looking into coherence can investigate three types of results:

- **process results;** what institutional mechanisms and other process elements are in place that can either hinder or facilitate the promotion of coherence?
- **outputs** of policies; what changes to the legislative texts of existing policies or new policies are anticipated to result in an increase of coherence in terms of desired outcomes?
- **outcomes;** what eventual outcomes, or impacts, can be attributed or otherwise associated to the effects produced by these process results and policy outputs.

Assessing and analysing results at any of these levels requires an analysis of the *inputs* provided.

In relation to the process results, the synthesis report on the seven joint-evaluations related to the Treaty on European Community's precepts coherence, complementarity and coordination concluded that moving towards the realisation of horizontal objectives in such a complex environment requires some form of institutional coordination, formal or informal, between the relevant parties at different levels. The synthesis report concluded that institutional coordination needs two additional, mutually reinforcing efforts to be successful (ECDPM 2008):

- **Political commitment** on the part of the relevant stakeholders; with leadership and clearly defined policy objectives, priorities and criteria for assessing progress.
- **Adequate analytical capacity;** effective systems for monitoring, evaluating impact; adequate capacity for generating, sharing and processing relevant information and for developing and implementing common standards.

These are among the essential conditions for delivering *inputs* that can help produce coherence results 'higher' up in the result framework.

The conceptual framework for this study is based on the **result framework** (or 'logical framework'), as it currently is the most widely accepted tool used for planning interventions. Consequently it also remains a dominant approach in operationalising evaluation methodology by the IOB, which commissioned this study, and which in turns bases its practices on the government-wide evaluation regulation. Given the complex and multiple players and factors involved in determining how policies contribute to development outcomes, it is nonetheless important to highlight that the IOB and other evaluation offices have also experimented with alternative analytical frameworks to guide the operationalisation of the evaluation methodology, including a broad group of studies referred to as **systems thinking** (e.g. De Lange et al 2011).

Coherence at the level of policy **outputs**, and linked to this at the **outcome** level, can be defined as the result of good institutional coordination practices (ECDPM 2008). This leads to the conclusion that it would be important for this study to not only look at results but also examine how past evaluations looked at the process through which results were achieved in relation to coherence and horizontal objectives. Assessing both also allows evaluators to see to what extent investing in achieving good process results (e.g. inter-ministerial coordination groups, changed screening procedures of new EU policy proposals) contributes to policy outputs, and to what extent these affect **development**

outcomes, thus justifying the investments made. Section 2 of this study makes a more comprehensive attempt at operationalising coherence in the results framework.

Whereas the actual practice of promoting coherence in international cooperation (e.g. by diplomats in MFAs, experts in sectoral ministries, non-government organisations, politicians, researchers) is mostly focused on the policy formulation phase, development evaluations can cover both the policy and implementation phase with the overall objective to establish what results have been achieved and to improve future formulation and implementation. Evaluations can therefore address one or both phases, which in precise terms would entail the following:

1. in the **policy formulation phase** it is the actual content of the policies that is the objective, based on assumptions of what policy actions benefit developing countries' needs and priorities (i.e. the text of the policy document as an end),
2. in the **implementation phase** it is the actual effects in developing countries that have to confirm the positive or negative contribution of policies of an OECD country (i.e. the policy as a means).⁹

2.3. Relating the concept of coherence to the five DAC evaluation criteria

Further ideas to inform the operationalisation of coherence for the use in evaluations can be gained by associating this concept to the existing five evaluation criteria:

OECD/DAC criteria¹⁰	Link with coherence
Relevance: The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor	Formulation: extent to which an aid project can make a difference given the presence of other policies affecting development
Effectiveness: A measure of the extent to which an aid activity attains its objectives	Formulation: degree of coherence in policy outputs Implementation: extent to which development intervention achieves its goals and is affected in doing so by other policy actions.
Efficiency: Efficiency measures the outputs -- qualitative and quantitative -- in relation to the inputs	Formulation: efficiency of process-investments to achieve policy outputs.
Impact: The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended	Implementation: extent to which other policies affect the achievement of development impact.
Sustainability: Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn	Formulation and implementation: extent to which sustainability of development results are affected by other policies

Given the five different levels at which coherence can be promoted, there are various relationships and attribution questions to deal with when tackling the question of the relation between coherence on the one hand and effectiveness and efficiency on the other. When tackling the question only with one dimension in mind (e.g. does a coherent trade policy of one country make 'aid for trade' interventions financed by that same country through ODA more effective) the answer would be that there can be a strong relationship, but at the same time that due to all the other factors at play good results could even be achieved with an incoherent trade policy.

⁹ Low investments in evaluation and research of the effect of OECD policies in developing countries have led to a situation where there is most emphasis on the process in the policy formulation stage (e.g. the DAC Peer Reviews that look into institutional mechanisms), and insufficient evidence to confirm or reject assumptions on which policies are 'development-friendly' and under what conditions this is the case.

¹⁰ <http://www.oecd.org/dataoecd/15/21/39119068.pdf>

Unfortunately, different donors and organizations use different definitions of some of the evaluation criteria set out above. This complicates the above exercise of relating these criteria to the concept of coherence. Research question 2 was therefore further looked into by comparing different organisations on how they define concepts of **efficiency and effectiveness**, which initial studies identified as the most relevant criteria to examine for the purpose of evaluating coherence.

To the **OECD**, effectiveness is a measure of the extent to which an activity attains its objectives, while efficiency measures the attained results (which could be outputs, outcomes or impacts) in relation to the inputs used. The main difference between efficiency and effectiveness then is that efficiency considers costs and efforts made (i.e. inputs), while effectiveness does not. From this perspective, coherence appears more closely related to efficiency than to effectiveness. An aid intervention could attain its objectives even if non-aid policies are highly incoherent. However, incoherent non-aid policies could affect the efficiency of aid interventions, since the same results could have been achieved at fewer costs if non-aid policies had been more coherent.

The **IOB** (2009) uses definitions of efficiency and effectiveness that are different from those agreed at the OECD. In their guidelines, efficiency measures how economically inputs are converted into outputs. Meanwhile, effectiveness relates to the extent to which outputs contribute to outcomes. Thus, they “refer to two successive levels in the results chain” and are consequently independent of each other. To IOB, “an intervention can be efficient without having an effect, and an effective intervention can be inefficiently implemented”. From this perspective, coherence relates more to effectiveness than to efficiency. Incoherent non-aid policies do not affect an aid intervention’s generation of outputs (i.e. efficiency), but rather undermine the translation of outputs into outcomes (i.e. effectiveness).

A third approach is to consider **different types of efficiency**, as proposed by Palenberg (2011). Palenberg distinguishes between ‘**transformation efficiency**’ and ‘**optimization efficiency**’:

1. Similar to the definitions of efficiency discussed above, transformation efficiency refers explicitly to the relationships between inputs and results. In other words, it measures efficiency by examining the transformation of costs into benefits.
2. Optimization efficiency relates more to the efficiency of the entire socio-economic system, without referring to specific inputs. The concepts of Pareto-efficiency and Kaldor-Hicks efficiency relate to optimization efficiency, since they deal with the question whether any changes can be made which would leave society better off as a whole.

Palenberg further argues that efficiency dominates all other evaluation criteria, including effectiveness. This implies that an intervention cannot be efficient without having any effect. When defined this way, this study argues that rigorous efficiency-analysis suffices for welfare-optimizing decision-making. Other evaluation criteria in isolation, such as effectiveness, relevance or sustainability, are merely required because in practice, efficiency analysis may be too complex and therefore lack the required reliability or scope.

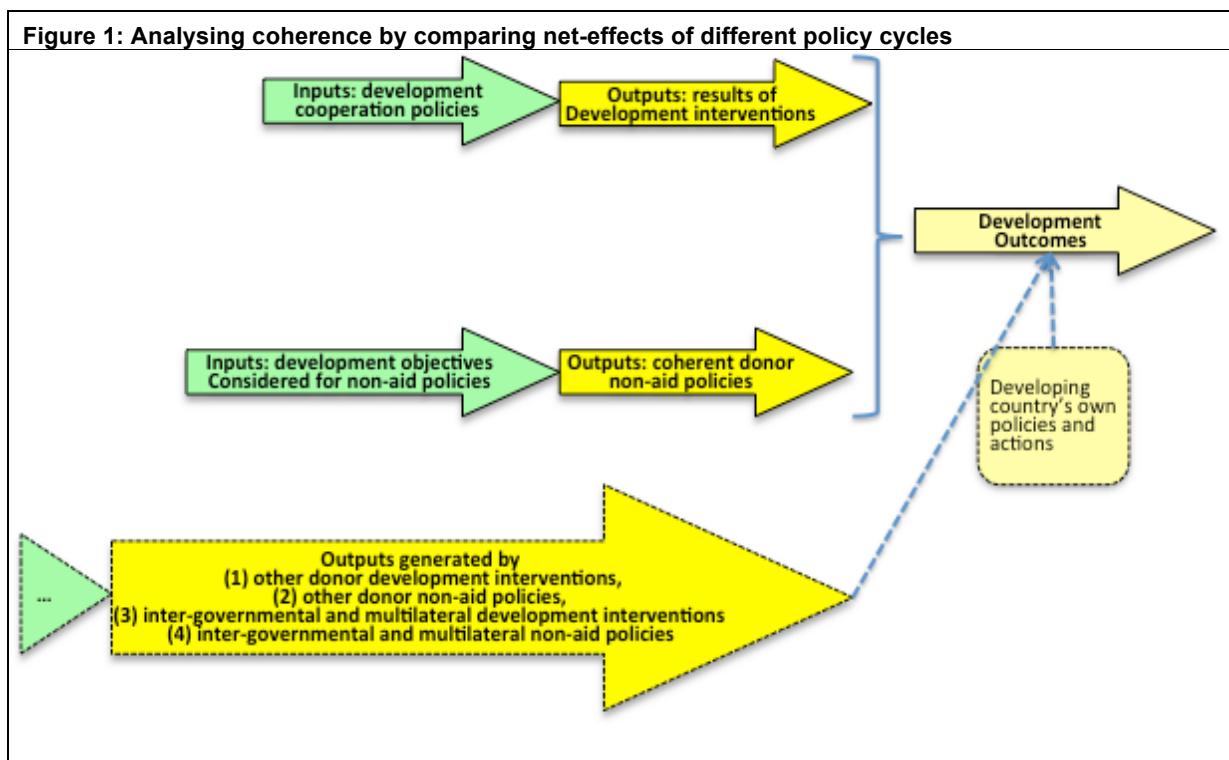
How could we relate these two types of efficiency back to coherence? Transformation efficiency could be said to relate coherence to aid effectiveness. If non-aid policies work at cross-purpose with aid policies and undermine the achievement of aid policy objectives, then this incoherence impairs the translation of aid policy inputs into aid policy results, thus reducing its transformation efficiency. When instead relating coherence to optimization efficiency one could move the conceptualization of coherence beyond the effectiveness of aid interventions (which is desirable according to the National Focal Points for PCD (OECD 2009)). Rather, the question then becomes how non-aid policies affect the efficiency of the global socio-economic system. In doing so, however, it is important to adopt a

dynamic rather than static approach.¹¹

The above comparison of studies and guidelines shows how slight differences in definitions of key terms can lead to different conclusions on how and where to situate ‘coherence’ in the evaluation criteria framework. Both given the past policy debates (OECD 2009) as well as the compatibility with the IOB’s guidelines (derived from overall government policy) it would seem best to further explore conceptualisation of coherence in terms of how policies are made more coherent with development objectives and how they both individually and collectively affect developing countries. The next section will explore this further and present what can be learnt from existing approaches to operationalise and evaluate policy contributions to other horizontal policy objectives.

2.4. Situating coherence in the results framework

Similar to the definitions of coherence and the five DAC evaluation criteria, there is **no international consensus** on defining key terms in the result chain¹², which remain the dominant approach to planning in development cooperation. When using the terms defined by the IOB (2009), the following results framework can be presented with some ideas on the role of coherence as derived from the discussion in the previous section. Impact does not feature in this overview because of the choice made in this publication to define it as the ‘net-effects’ of the intervention (analogous to the IOB’s guidelines), i.e. the part of the outcome achieved that can be related to the outputs achieved by the intervention.



¹¹ For instance, complete free trade may maximize static welfare gains, but may be sub-optimal from a dynamic perspective, since developing countries may need temporary strategic trade policy to build up infant industries.

¹² E.g. this page describes differences in terms used by the EC and OECD: http://ec.europa.eu/europeaid/evaluation/methodology/glossary/glo_en.htm

Definition of terms used in the result framework (adapted from IOB 2009 and RPE 2006)¹³:			
Input	Output	Gross-Effect (Outcome)	Net-effect (policy effect)
The physical and human resources used to produce the actions/interventions and the policies describing these	The interventions or actions made in the context of the policy that is evaluated.	The societal results as observed.	The effects that can be attributed or otherwise related to the policy itself.

In this understanding coherence can be assessed by comparing the impact (or net-effects) on development outcomes of at least two of the following six result chains¹⁴:

1. The effects of a donor's bilateral development cooperation on development outcomes
2. The effects of a donor's non-aid policies on development outcomes
3. The effects of other donors' bilateral development cooperation on development outcomes.¹⁵
4. The effects of other donor's non-aid policies on development outcomes
5. The effects of intergovernmental (including EU) and multilateral development cooperation on development outcomes
6. The effects of intergovernmental (including EU) and multilateral non-aid policies and interventions on development outcomes

It should be emphasised that while it is important to be aware of what different kinds of actors and 'sources' of the results can affect development outcomes, as represented in the 'left to right' logic in the result chain, the actual measurement of coherence often has to reason back from the results as identified from an agreed result framework guiding the evaluation process (in the case of a theory based evaluation).

When analysed in the above fashion, coherence can be judged as a 'horizontal policy objective' which can be evaluated both in relation to (1) the extent to which this objective is reflected in policies defined in different contexts and levels and (2) the extent to which these policies, by themselves, in different combinations or collectively, contribute to the outcome to which the horizontal objective relates (in this case development).

Analysing it in this manner allows the development evaluation field, which has not made strong investments into assessing coherence, to draw on existing experiences in evaluating horizontal objectives in other policy areas. Horizontal objectives where it is assumed that most progress has been made include:

- Environmental safety and improvement of the environment (as per EU Treaty Art. 37, a 'high level' of environmental protection and improvement of environment is to be promoted by all policies of the EU). In the environment field, issues of coherence are often discussed in terms of 'environmental policy integration'.

¹³ It should be noted that these definitions relate to evaluations of the implementation phase, not that of the policy formulation phase. In the latter case a separate result chain should be drawn up, since some actors (e.g. advocacy groups, officials tasked with promoting coherence in policy making) are only held accountable for their contribution to the content of policies, not the extent to which this leads to development objectives. In view of the main focus of development evaluations on development outcomes, this separate result chain is however not developed here.

¹⁴ The coherence inside development policy (e.g. between economic and environmental interventions) is not included here, while intergovernmental and multilateral coherence can be seen as two or more separate result chains but have been collapsed into one here. The use of 'policies' here should be understood as both the overall strategies and their operationalisation in specific interventions and actions.

¹⁵ N.B. the interrelations between the different donors' interventions and efforts to optimise positive synergies during the planning and implementation phase is also referred to as 'complementarity' or 'division of labour'

- Ensuring consistency in external action towards fragile states, as shown by the increased attention to '3D' (Diplomacy, Defence, Development) approaches (as per EU Treaty Art. 3, The Union shall in particular ensure the consistency of its external activities as a whole in the context of its external relations, security, economic and development policies).¹⁶ In the literature on 3D approaches to fragile states, issues related to coherence between the different activities are often discussed in terms of a 'Whole-of-Government' approach.

Other examples of horizontal objectives include subsidiarity (at what level of governance are decisions best to be made), gender equality, resource efficiency and respect for human rights. However investments made in evaluating the progress towards these horizontal objectives are expected to be lower than the three highlighted here. Furthermore, some of these issues, gender equality for instance, could be said to not so much involve competing objectives and consequent political trade-offs, thus rendering them more 'mainstreaming' than 'coherence' issues.

As discussed, most evaluations that have been carried out on coherence in the field of international cooperation so far have focused on the policy process by assessing the outputs (i.e. the actual policies as formulated). To gain further understanding of how these policy outputs contribute to outcomes in developing countries, studies need to further invest in exploring these linkages. One reason why this is important can be drawn from Picciotto's (2005) distinction between necessary and unnecessary incoherence. Some incoherence is necessary, because perfect coherence is impossible in a pluralistic political system and could only be achieved by absolute and competent dictatorships. Any policy produces winners and losers, and interest groups must find an overall array of policies that ensures a stable political settlement through principled negotiations. Incoherencies are a very likely side-effect. However, Picciotto emphasizes that incoherencies can only be necessary if they are the outcomes of such a principled negotiation. Evidence-based evaluations of the actual outcomes of incoherencies on the ground can inform and influence this negotiation process and may bring to light that what was perceived to be a necessary incoherence is in fact an unnecessary incoherence that needs to be resolved (a similar conclusion was reached in ECDPM 2008).

With regard to the development field, it has been posited that 'the tool kit of the evaluation profession is well stocked to deal with PCD' (Picciotto 2005, p.326). However, these tools have still hardly been applied in practice in relation to PDC, so much remains unknown. Therefore, this study seeks to make a systematic review of the various ways in which coherence has been assessed in evaluations covering international cooperation but more importantly in other domains of policy making that have to promote horizontal policy objectives.

¹⁶ Technically speaking this is more a consistency than a coherence objective, although different studies reviewed do point to horizontal objectives such as state-building or conflict prevention that are used to vet the contribution of each of the three Ds.

3. Methodology

3.1. Constructing the inventory of studies

In view of this study's objectives, an inventory was produced of a number of evaluations and academic journal articles that assess progress towards one or more of the following three policy fields¹⁷: (1) development cooperation¹⁸, (2) environmental safety and improvement of the environment and (3) the link between diplomacy, defence and development.

After this initial investigation of available databases and overviews and discussions with the IOB¹⁹, it was agreed to use the following **three-pronged approach** to identifying useful reports to assess for this study:

1. First of all, a limited number of ex-post evaluations commissioned by the Netherlands government and the EU were studied, which were identified as relevant on the basis of the evaluation object (i.e. it had to involve the assessment of a horizontal objective).²⁰ In addition, a number of evaluative field studies by independent organisations that focused specifically on coherence for development were also assessed. The search process was complex because neither the Netherlands government, nor the European Commission, nor the OECD organise their evaluations in a searchable database.²¹
2. Secondly, a structured search was made of peer review databases. Annex 2 presents the search terms used, which led to a body of studies from which specific studies for analysis were selected based on a selection process presented in box 1 below. Further studies were selected through 'snowballing sampling', i.e. selected references of the papers identified through the searchable database that were deemed relevant to the study were also subjected to the same selection process described in box 1.
3. Finally, subject matter specialists on environment and security studies were contacted to suggest particular evaluations or studies in their respective fields of work.

Box 1: Selection process for studies

From the large number of articles generated by the search engines using the search terms, which are provided in Annex 2, the study team narrowed down the number of articles.

As a first step, this was done by screening the titles, which led to the elimination of a large number of articles that were deemed not to be relevant for the purpose of this study.

The remaining articles were narrowed down further based on a quick scan of the text and main sections, in order to further determine the article's suitability for this study. Four main criteria were used in determining this suitability, of which the first two broadly relate to the relevance of the article's subject ('the what') and the other

¹⁷ The study first also looked into available material on evaluating the promotion of gender equality but did not pursue this further as the material identified did not include sufficient information on the results achieved.

¹⁸ This concerns a broad category of policies and policy actions that are assessed for their positive or negative contribution to development outcomes, also referred to as 'Policy Coherence for Development'.

¹⁹ The study team also reviewed the structured approach to identifying relevant literature adopted by the recent IOB food security study Subtitled: a systematic review of the impact of interventions in agricultural production, value chains, market regulation, and land security. <http://www.minbuza.nl/bijlagen/producten-en-diensten/evaluatie/afgeronde-onderzoeken/2011/improving-food-security/rapport.html>

²⁰ Some ex-ante evaluations in the field of environment were looked into, given that they play an important role in this policy field and provided some interesting lessons.

²¹ The Netherlands government was most extreme with, apart from the IOB, no relevant ministries having a dedicated webpage giving access to evaluations, which instead had to be sought among the full written correspondence with the Parliament. The exception was this page with policy reviews (*beleidsdoorlichtingen*) which however had not been updated since 2010: <http://www.rijksbegroting.nl/algemeen/overzichten,Beleidsdoorlichtingen.html>

two relate more to the nature of the article ('the how'):

1. Explicit reference to the interrelations between different public policies, for instance between development and environment policies.
2. The level of the analysis and the presence of a wider policy discussion. For the purpose of this study, articles that only explored the outcomes of the interrelations between policies exclusively at the smallest micro-level (e.g. community or village) were deemed less relevant.
3. The use of data. For this study, only those studies that actually process, or at least analyse data, whether primary or secondary, were included. Thus, articles that are mostly limited to a review of previous studies without analysis of new data or new analysis of existing data are excluded.
4. The transparency of the methodology. Since this study is mostly interested in how the interrelations between different policies can be assessed, articles that do not clearly set out the employed methodology to explain how the data was collected are less relevant.

Annex 3 presents the 22 studies that were selected through the above three means and were covered by this study. The flexible methodology used for this study and the informed choice to look into three particular domains was deemed appropriate given its exploratory and 'fact-finding' nature as well as the fact that only academic journal articles were 'searchable'. In addition to the selection criteria outlined in box 1, it should also be remarked that the study sought to target studies examining coherence at the outcome level, but this was not adopted as a 'hard' criteria in the sense of necessarily excluding studies that did not focus at the outcome level. Finally, it is clear that none of the four criteria used to judge a study's suitability were 'hard' in the sense of having unambiguous threshold levels determining in- or exclusion, although criteria 1 and 3 were given slightly more weight than the other two.

As a consequence of these methodological choices, it should be emphasised that the studies as identified concern a relevant group but by no means a 'representative sample'. This relative 'selection bias' should thus be kept in mind when interpreting the study's results in terms of strengths and weaknesses of the methods used in the different studies, which essentially results from a desire to look into both evaluations and relevant journal articles.

3.2. Analysing the studies

For each study that was identified, the following main questions were looked into:

1. How was coherence defined and operationalised in the study concerned?
2. What methods were used to collect and analyse the information to evaluate progress in relation to the objective as defined under question 1?
3. At what level(s) did the evaluation focus, i.e. at micro (households), meso (district/provinces) or macro (countries or regions)?
4. What is the quality of the analysis, and what are the strengths and weaknesses of the applied methods?

These four main questions were operationalised into a more detailed data grid, which in turn provided the basis for the inventory of studies that will be one of the outputs of this study. The data grid is shown in annex 4.

Under the first question, we looked at how the various studies defined and approached the concept of coherence. Some may focus on disentangling competing objectives and looking for ways to quantify the involved trade-offs or search for win-win outcomes. Others may approach coherence more from a perspective focused on horizontal objectives, examining whether integrated policy-making can lead to better results.

In relation to the second question, the evaluations were distinguished along various dimensions. First of all, it was considered whether a study focuses on the outcome level, or whether it predominantly deals with outputs and process-related issues in policy-making. The studies were subsequently assessed on what methods were used to evaluate the results relevant to the horizontal policy objective, as well as the extent to which the choice of the methods was motivated and also transparently analysed in terms of strengths and weaknesses. In preparation of the analysis of the studies, an initial classification of methods was made based on a review of several documents²². This classification relates to two dimensions of the used data, namely whether the data source is primary or secondary, and whether the nature of the data is qualitative or quantitative. The following table lists possible methods using this classification.

	Qualitative Data	Quantitative Data
Primary Source Data	<ul style="list-style-type: none"> • Interviews (structured, semi-structured, non-structured) • Group Interviews & Panel Discussions • Direct observation 	<ul style="list-style-type: none"> • Statistical surveys • Direct measurement • Cost-Benefit Analysis • Cost-Effectiveness Analysis
Secondary Source Data	<ul style="list-style-type: none"> • Document analysis 	<ul style="list-style-type: none"> • Descriptive statistics • Econometric models using existing data • Quantitative modelling

Based on the data collected, the study was to make a more detailed inquiry into the key strengths and weaknesses of the methods as used. Palenberg (2011) assessed different evaluation methods based on document reviews, interviews and own analysis based on the following two categories:

- How powerful is a method, i.e. how useful and reliable are analysis results? This leads to methods being related to one of three levels of analysis that they can support:
 - Level 2 analysis can assess the interventions so that they can be compared with alternatives or benchmarks.
 - Level 1 analysis is capable of identifying the potential for improving the interventions assessed.
 - Level 0 analysis is entirely descriptive and can usually not produce well-founded recommendations.
- What are data, resource and skill requirements?
 - Data requirements are assessed both by the type (qualitative and quantitative, the latter subdivided in financial and non-financial) and the origin (what level of an intervention's result chain the data stems from) of data.
 - Time requirements are measured both in terms of working times for both the evaluator and stakeholders.
 - Finally, skill requirements indicate whether skills needed for the analysis exceed what is considered as basic evaluation skills.

As per the limited time and resources available for conducting this study, it seemed straightforward to analyse the findings from the analysis of the studies and to contrast this with Palenberg's findings and resulting matrix of evaluation methods. The result is an adapted matrix for the evaluation of coherence, based on the findings of the analysis of studies, which is presented in section 4.

²² E.g. Palenberg (2011), IOB (2009) and OECD (2009b)

The third question considers the level at which the studies examine coherence. It should be noted here that definitions of macro, meso and micro level may differ between different studies. In some cases, the level refers to a geographical dimension (e.g. community, village, district, province, country), while in other cases it could refer to different dimensions (e.g. firm, sector).

The fourth question was approached through a grid by means of adapting the IOB's own grid for assessing the quality of evaluations. As per the focus of this existing grid, the quality of evaluations was mainly assessed based on the validity and reliability of the research methods used.

4. Study findings

4.1. Overview of the studies

Following earlier exploratory research and discussions with the IOB in The Hague, it was agreed to select and analyse studies in three main thematic ‘fields’. These fields were the following, with the number of studies reviewed mentioned between brackets:

- Development cooperation (n=3),
- Environmental safety and improvement of the environment (n=15):
 - Biodiversity (n=3)
 - Environmental protection (n=4)
 - Conservation and deforestation (n=7)
 - Energy and biofuels (n=1)
- External action towards fragile states and the 3D approach (n=4).

From the list above, the most obvious thing to note is the relative abundance of studies relating to the environment. To some extent this was to be expected, given that the environment field has engaged with issues related to coherence for a much longer time²³, even if the term coherence has not been used that much. The 15 studies covering the environment have been divided into smaller groups so as to provide a clearer overview of the various sub-themes addressed.²⁴

4.2. Main findings in relation to the four key questions

This section presents the main findings from the 22 studies that were reviewed. Annex 4 presents a review grid that was used to capture relevant parameters, information and key findings from the studies, which were systematised in a separate spreadsheet to allow comparison of findings as a basis for the analysis. The following four sub-sections describe the main findings in relation to the four questions that were introduced in section 3.2.

4.2.1. How was coherence defined and operationalised in the studies?

Out of the 22 reviewed studies, seven studies actually used the specific term coherence: all three reviewed studies related to the field of development cooperation, three studies on environmental issues, and one on security. The way in which coherence was defined and operationalised varied:

- Out of the seven studies, five studies defined the term, while the other two did not.
- Two of the five studies that did define coherence clearly emphasize the aspect of mutually reinforcing policies that contribute to creating synergies.
- One study, on the other hand, focused exclusively on the consistency aspect, defining coherence as the absence of contradictions between policies.
- Another two studies adopted a definition incorporating both these aspects.
- In three of the five studies that did define coherence, the theoretical discussion of coherence was limited to the definition alone. The other two studies offered a slightly more extensive analysis of coherence, discussing its value and potential problems and challenges.

²³ The requirement to integrate the needs of the environment into other sectoral policies has been included in the EU Treaties since 1987. Some background information can be found here: <http://www.ieep.eu/work-areas/governance/environmental-policy-integration/>

²⁴ It should be noted that the sub-categories are not completely mutually exclusive. Studies counted under ‘environmental protection’ dealt with a range of environmental issues, ranging from the emission of greenhouse gases, to pollution, to biodiversity and conservation. Furthermore, some of the studies counted under ‘conservation and deforestation’ also dealt in part, but not exclusively, with biodiversity.

While the remaining 15 studies did not use the term coherence, they did engage with issues related to coherence. Looking across all 22 studies, we can discern three broad approaches to conceptualizing coherence.

1. 'side-effect' perspective (n = 2)
2. 'horizontal objective' perspective (n = 9)
3. 'trade-off' perspective (n = 11)

These three approaches are not mutually exclusive and some studies combined different elements of all three approaches. However, each reviewed study could be linked to either one of these categories, based on which perspective dominated. Some general findings are presented in relation to each of these three approaches.

A **'side-effect' perspective** on coherence does not represent a specific perspective on or conceptualisation of coherence as such. Rather, studies adopting this perspective assess the effects of policies on factors that the policy was not designed to target and what is considered an (unintended) 'side-effect'. Of the reviewed studies, two were predominantly based on this perspective. Alliance Environnement (2007), in an evaluation for the European Commission, examines the environmental impacts of Common Agricultural Policy measures related to the beef and veal and the milk sectors. These measures, mostly subsidies, were designed with a particular policy objective in mind, for instance the supporting of European farmers, but can have 'side-effects' on the environment, which is what this study seeks to measure.

Similarly, Van Beers et al. (2002) investigate the environmental impacts of a range of indirect subsidies in the agricultural, energy, transport and tourism sectors. Again, these subsidies were devised for a certain purpose, but were found to have 'side-effects' on the environment. Through this 'side-effect' approach, these studies can certainly come to some conclusions on the need to improve coherence as a means to avoid these side-effects – providing decision makers agree that these are undesired. By demonstrating that the policies have negative effects on the quality of the environment, and by pointing out that the European Union also spends significant resources on environmental policies that are aimed at raising the quality of the environment, decision makers could conclude that there is an undesired inconsistency between the two policies.

The limitation of the 'side-effect' perspective on coherence is that these studies are not able to progress much further in terms of evaluating or improving coherence and providing operational recommendations to that end. After all, by removing the subsidies that were found to have harmful environmental effects, the achievement of the policy objectives that those subsidies were designed for would be compromised. This perspective is however often used in the absence of a clear overall political consensus on the direction of such policies, i.e. an intervention logic, which would be a prerequisite for evaluating the effects of policies beyond side-effects.²⁵

The **'horizontal objective' perspective** focuses the analysis on the effects of the policies concerned on one overarching policy objective that is not 'owned' by a particular sector or Ministry, i.e. the 'horizontal objective'. Such a horizontal objective can be very broad, such as economic development or environmental protection, or more specific, such as achieving peace & security, or it can consist of a combination of several objectives. These studies examine whether and how other policies contribute to the achievement of the horizontal policy objective as defined. In addition to assessing the achievement of the objective concerned, this perspective also allows to evaluate to what extent the

²⁵ Compared to evaluations, academic studies can take a more flexible approach in theorising or otherwise 'reconstructing' the intervention logic.

(lack of) achievement can be related to the degree of integration of the horizontal policy objective into other policies.

Out of the 22 reviewed studies, nine studies were found to be predominantly based on this perspective. The breakdown of these nine studies in terms of the policy field they relate to yields a very interesting picture, as they include all three studies on development cooperation, as well as all four studies on security and the link between development, defence and diplomacy. Finally, two studies on conservation also adopted this horizontal perspective. In the security studies, the coherent approach of having various government departments dealing with diverse policy fields working towards the same horizontal objective is commonly referred to as a 'whole-of-government' approach. Since the horizontal perspective is inherently linked to how various government agencies work together, studies operating from this perspective are more likely to have a process-oriented focus than studies having a 'side-effect' or 'trade-off' perspective which may often 'argue backwards' from the results as found.

The 'horizontal' perspective is more comprehensive than the 'side-effect' perspective, mainly because it considers the interplay of several policies in achieving one objective. As a result, such studies can comment on the observed trade-offs or, more frequently, the synergies between the different policies. However, at the end of the day this perspective faces the same challenge as the 'side-effect' perspective. When there are not only conflicting policies, but also conflicting horizontal objectives, a study adopting the 'horizontal' perspective runs into difficulties when trying to properly assess overall coherence in relation to one of them. In the case of conflicting overall objectives, the question which one takes precedence over the other inevitably presents itself at some point. This is illustrated in the conclusions reached by Patrick et al. (2007), based on a review of approaches to engaging with fragile states used by seven donor countries. While all the seven governments '*regard fragile states as both a developmental and security challenge, donor capitals differ in the weight they give these two considerations*'. Furthermore, development agencies advocate '*policy coherence for development*', while '*foreign and defense ministries tend to be more preoccupied with achieving what might be termed policy coherence for national security*'. In other words, the question is whether defence policies should be used for promoting development or whether development policies should be used for promoting security. Studies adopting a 'horizontal' perspective are not always in a position to answer this type of question, especially when the precise nature of the overall objective is not clarified in key policy documents. Instead, these studies may represent more of a 'consistency' approach and often do not make explicit or otherwise operationalise the overall objective concerned.

Finally, a group of studies adopted a '**trade-off**' perspective. These studies focus on the extent to which two different policy objectives are compatible with one another. They examine the trade-offs between different objectives, and investigate whether these trade-offs are ubiquitous, or whether there are conditions under which the objectives can be reconciled and both goals can be achieved simultaneously, resulting in win-win situations. Of the 22 reviewed studies, a total of 11 studies predominantly adopt such a trade-off perspective. Interestingly, all these studies deal with environmental questions. Furthermore, most of them are academic studies. Compared to the 'side-effect' and 'horizontal' perspectives, some of these academic studies with a 'trade-off' perspective have a different starting point. They depart from theoretically defined objectives as opposed to from the (public) policies as adopted. Thus, Dasgupta et al. (2005) investigate poverty data and environmental data and, on the basis of that analysis, put forward recommendations regarding policy integration. Illukpitiya et al. (2010) investigate the relation between deforestation and agricultural yields, and find that non-timber forest product extraction is a decreasing function of agricultural efficiency in forest peripheries. Therefore, they argue, there is scope for win-win situations when agricultural policies and forest conservation policy are combined.

Another issue that characterises this perspective is, as its name suggests, that it explicitly considers the trade-offs between two different objectives and may be less suitable for an analysis that covers different levels of coherence. For instance, a number of studies examine the impacts of protected areas. In doing so, they assess the effects on both environmental and economic factors (e.g. Ferraro et al. 2011, Gjertsen 2005, Sims 2010, Sandker 2009). Studies from a 'horizontal' perspective tend to restrict their analysis to one type of outcome only. For instance, studies on international development look at the impact of donor policies on developing countries, but do not really consider the effect of those policies in the donor countries themselves. As such, potential trade-offs between objectives regarding international development and donor countries' objectives on domestic issues are not explicitly juxtaposed.

Thirdly, many of the 'trade-off' studies focus on the conditions under which policies may alleviate trade-offs or promote synergies. In other words, they investigate whether certain factors and characteristics change the dynamics between different objectives. In the context of the effect of protected areas, such factors can include the suitability of the land to agriculture, distance to major cities, the size and composition of the labour force, and others.

Again, it should be noted that some of the studies on development cooperation and security also argue in terms of trade-offs and synergies, but their dominant focus is not on measuring the trade-offs between different objectives. Similarly, the 11 trade-off studies also display elements of a horizontal perspective.

4.2.2. What methods have been used to collect and analyse the information to evaluate progress?

This section has two aims. First of all, it will provide an overview of the methods that have been used to assess coherence in the reviewed studies, and show whether there are preferences to combine methods in certain ways in an approach to triangulate the data as gathered. Secondly, it will link the level of the analysis of the studies to the three levels used by Palenberg (2011).

The following table presents an overview of the various methods that were found.

	Development (3 studies)	Environment (15 studies)	Security (4 studies)	Total
Interviews	3	7	4	14
Group interviews			1	1
Direct Observation		2		2
Surveys		4		4
Document Analysis²⁶	1	2	3	6
Descriptive²⁷ Statistics	1	5		6
Regression Analysis		5		5
Quantitative Modelling		5		5
Multi-Criteria Analysis		1		1

²⁶ ²⁷ Note that studies were only counted as using document analysis or descriptive statistics when this was an explicit part of the study's methodology, i.e. not when such documents or statistics are mainly used for purposes of contextualisation, introduction, etc.

The totals in this table add up to 44, meaning that the average number of methods used by one study is exactly two. The standard deviation is around 0.85, indicating that fluctuation around the mean of two is not very large. While the table disaggregates for the three areas of studies reviewed, the low number of studies reviewed in the first and third category makes that it has insufficient analytical value to make cross-comparisons. Table 6 below however looks further into the combination of methods used in the studies, linked to the level of analysis of each study. Another relevant aspect linked to the methods that were used is what type of indicators were used in analysing coherence. Table 5 presents a selection of the indicators used in a number of the reviewed studies.

Table 5: Selected Indicators Used in Reviewed Studies		
Some of the more qualitative studies that were reviewed did not translate the gathered data into indicators. However, quantitative studies always work with some type of indicators. This box presents a selected overview of indicators that were used in some of the studies.		
Study	Coherence issue	Operationalisation and examples of indicators
Van Zeijts et al. (2005)	Agriculture and Biodiversity	Farmland biodiversity indicator
		Total grassland area
		Nitrogen application, as approximate indicator for farming intensity
Alliance Environnement (2007)	Agriculture and Environment	Agricultural subsidies
		Total number of livestock
		Livestock densities
Van Beers et al. (2002)	Economy and Environment	Subsidies to agriculture, tourism, energy and transport
		Production levels
		Greenhouse gas emissions
Dasgupta et al. (2005)	Environment and Development	Poverty rates
		Deforestation rates
		Percentage of agricultural land that is steeply sloped, as indicator of soil fragility
		Number of people without access to clean water, as indicator of water pollution
Hengsdijk et al. (2007)	Biodiversity and Development	Per capita income of agricultural population
		Ratio between active labour force and total available agricultural labour force, as indicator of underemployment
		Nitrogen surplus in agriculture
		Ratio of agricultural land to forest
Brown et al. (2001)	Conservation and Development	Willingness to Pay for conservation, with data derived from surveys
		Tourism revenue
		Water nutrient concentration, as indicator of water quality
Ferraro et al. (2010)	Conservation and Development	Protected area (binary indicator)
		Poverty level from census data
		Forest cover
		Land use capacity, based on soil quality, precipitation, climate and slope, as indicator of suitability for cultivation

The second step in the analysis for this research question involved looking at the levels of analysis that the methods used in the studies allow for. Palenberg assesses a wide range of approaches to evaluating the efficiency of aid interventions and, for each approach, determines whether it either constitutes purely descriptive analysis, a level 1, or level 2 analysis. The purpose for the use of these three levels here was to find out whether the methods used by the assessed studies show general differences in the level of analysis that they allow. To Palenberg, descriptive analyses can provide certain insights into efficiency-related aspects of interventions, but cannot reliably compare the findings to anything else. Level 1 analysis is capable of identifying the potential for improvements within an intervention, while level 2 analysis can compare interventions with alternatives.

Some important issues need to be considered when trying to apply this concept of levels of analysis to coherence rather than efficiency, and to the application of methods in actual studies rather than to the 'inherent' potential of methods. Palenberg classifies methods as being of a certain level of analysis according to the method's potential based on a desk study and semi-structured interviews. He does not consider to what extent the quality of actual implementation of a method could affect its level of analysis. Thus, cost-benefit analysis is classified as a level 2 method, but it seems questionable that a poorly executed cost-benefit analysis should be classified as such. Relatedly, there is not much discussion in Palenberg on how the effective use of baselines and counterfactuals fits into the approach of thinking in terms of levels of analysis. Another issue to consider is that, for the purpose of this study, we are only interested in how coherence has been assessed. Several of the studies deal partially with coherence, but sometimes only within the context of assessing a certain policy. In these cases, the level of analysis in assessing the efficiency of that intervention is not necessarily the same as the level of analysis in assessing coherence.

For the purpose of this study we assign a certain level of analysis of coherence to each reviewed study, and not to the individual methods as used. This was done by the following procedure. A study would be of a level 2 analysis if it can actually demonstrate the added value of coherence. This would imply being able to attribute results to coherent policy-making (or lack thereof) and compare these outcomes to what the outcome would have been in the absence (or presence) of coherent policy-making. None of the reviewed studies was effectively able to do this, so no study was classified as being of level 2 analysis. For determining whether a study corresponded to descriptive or level 1 analysis, we mostly looked at whether there was any effective use of counterfactuals and baselines. In other words, it was assessed whether the study could compare, in relation to coherence, the examined policy to other policies, or across space or time, in order to come to conclusions about how the coherence aspects of the policy might be improved. For instance, a number of the environmental trade-off studies were classified as level 1 analysis. Mostly by comparing policies or regions, they could make recommendations about how a better design of the policy could alleviate the trade-off concerned. These were not classified as level 2 analysis, because they did not consider whether separate policies dealing with the two sectors involved in that particular trade-off separately, could have attained the same, or even better outcomes.

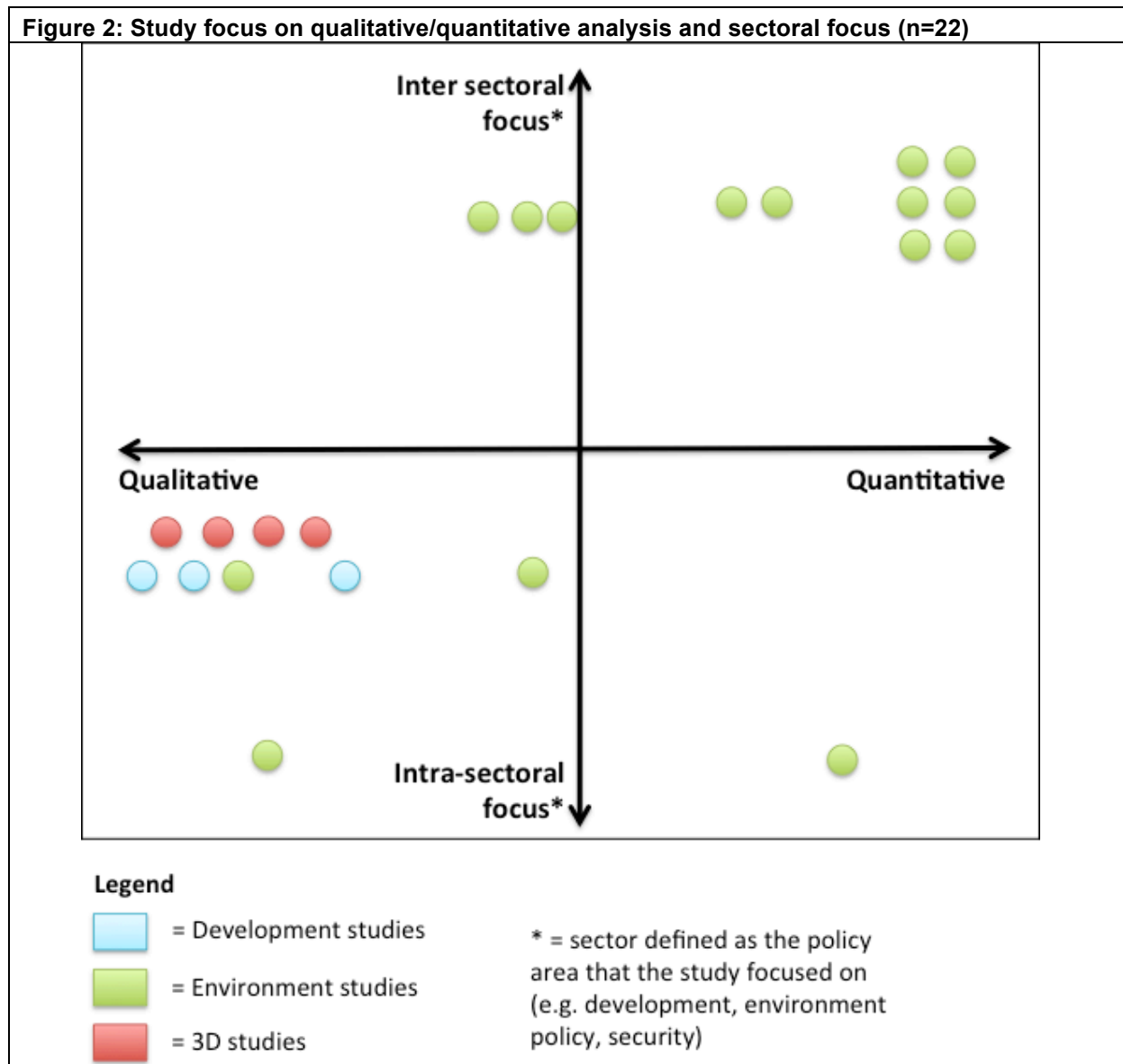
Table 6 presents the results of the examination of the level of analysis of the reviewed studies, broken down by combination of methods. The overview shows that none of the 22 studies featured level 2 analysis. A total of 13 studies were classified as level 1 analysis, while nine studies were descriptive in their analysis of coherence.

Table 6: Combination of methods used and the level of analysis made				
Combination of methods	2nd level	1st level	Descriptive	Sub- total
Descriptive statistics, interviews (Action Aid 2003, Alliance Environnement 2007)		1	1	2
Direct observation, document research, interviews (Schut et al. 2010)			1	1
Document research, interviews (Wilms et al. 2012, IOB 2008, Norwegian MFA 2004, Van Beijnum et al. 2006)		2	2	4
Interviews (Fair Politics 2010, Patrick et al. 2007)		1	1	2
Descriptive statistics, interviews, survey (McElwee 2012)			1	1
Document research, interviews, focus groups (Fishtein et al. 2012)		1		1
Descriptive statistics, direct observation, interviews, regression analysis (Gjertsen 2005)		1		1
Interviews, Multi-criteria analysis, survey (Brown et al. 2001)		1		1
Quantitative modelling (OECD 2008, Van Beers et al. 2002, Hengsdijk et al. 2007)		1	2	3
Regression analysis (Sims 2010, Ferraro et al. 2011)		2		2
Regression analysis, surveys (Illukpitiya et al. 2010)		1		1
Descriptive statistics, regression analysis (Dasgupta et al. 2005)			1	1
Descriptive statistics, quantitative modelling (Van Zeijts et al. 2011)		1		1
Interviews, quantitative modelling, survey (Sandker et al. 2009)		1		1
Sub-total	0	13	9	22

In addition to the overview of the level of analysis that the combinations of methods have been used for, some further observations can be made based on the preferences for methods by the different studies:

- Only six studies used more than two different methods, while seven studies exclusively relied on one approach to collecting and analysing data. This indicates that opportunities for triangulation of data were often limited.
- Furthermore, in most of the nine studies that used two different methods, one of those methods was the use of document research or descriptive statistics. In some of those cases, the use of that method was rather limited, thus effectively making that those studies also primarily relied on one method.
- The studies either predominantly used qualitative or quantitative methods. Out of the 10 studies that used regression analysis or quantitative modelling, only three also used methods of a more qualitative nature such as interviews, direct observation or surveys.
- Studies covering different policy field also differed in terms of the preferred methods to assess coherence. All ten studies that used regression analysis or quantitative modelling were related to the environment. This meant that none of the studies on development cooperation or security

- used these quantitative techniques for analysis.
- Finally, we can combine the analysis of the methods that were used with the analysis presented in section 4.2.1 of which perspective to coherence was adopted. Figure 2 gives a visual representation of the relationship between these two dimensions. It shows that, in general, studies adopting a trade-off perspective, which corresponds to a highly inter-sectoral approach, tended to use much more quantitative methods than studies with a more intra-sectoral perspective. Furthermore, the figure shows the exclusive presence of environmental studies among the former category.



These findings point to apparent differences in preferences between different ‘sectoral’ policy evaluation communities when it comes to certain approaches to data collection. While the development evaluations that were looked at mostly relied on qualitative analysis through documents and interviews, the environmental evaluations showed a much more ‘hard’ approach that put the use of modelling and analysis of quantitative secondary data as central to their methodology. Such choices can partly be related to the ‘tradition’ and (disciplinary) preferences in the respective evaluation communities (influencing evaluators’ expertise), but also with the availability and quality of data. It moreover seems that the quality and availability of data (and related budgetary considerations) combined with the specialisms of the evaluation community would mostly influence the choice of

methods in evaluating coherence in international cooperation.

4.2.3. At what level(s) did the evaluation focus, i.e. at micro (households), meso (district/provinces) or macro (countries or regions)?

Table 6 indicates at which ‘spatial’ levels the reviewed studies focused their analysis. On the one hand it was considered whether studies evaluated coherence at micro, meso and/or macro level(s). While these distinctions are not always easily made, in general micro refers to village or community level, meso to district or province level, and macro to national or continental level. Some studies do not have such a geographic focus. Where appropriate, micro was alternatively defined as referring to firm-level, meso to sector-level and macro to economy-level. These studies appear in Table 6 under ‘other or multiple levels’. What the overview shows is that most studies focus their analysis at the meso level, whereas in policy discussions there seems much interest to an analysis at a larger scale, e.g. on the effects of a particular EU policy on several developing countries.²⁸

Studies were also assessed on whether they investigated coherence at the level of process results, outputs or outcomes. While some studies are not confined to any one of these levels, it was found that, of the 22 studies that were examined, four studies had a focus on process results, while two studies mostly looked at the level of outputs. The other 16 studies examined coherence exclusively in relation to results at outcome level (i.e. actual effects). Some of these, however, investigated effects on proxy variables rather than the actual variables that the horizontal policy relates to.

	Input/process level	Output level	Net- and/or Gross-effects
Local level	1		4
Meso level (district)		1	6
Macro level (country or region)	2	1	4
Other or multiple levels	1		2

4.2.4. What is the quality of the analysis, and what are the strengths and weaknesses of the applied methods?

Only five out of the 22 studies offered much reflection on the benefits and limitations of the applied methods. The group of studies reviewed shows that those studies that have a predominantly qualitative, interview-based approach contain much information, but leave a lot to be desired in terms of analytical quality. In these studies there is often no systematic and transparent way of analysing and interpreting the gathered data. Obviously, the analytical quality differs between the different studies. Fishtein et al. (2012) is a good example of a study predominantly based on interviews that does provide analytical depth, as well as detailed discussion of the interview techniques used. The two studies by Fair Politics (2010, 2012) on the other hand do refer extensively to data gathered through the interviews, but are less successful in systematically analysing these. Action Aid (2003), meanwhile, states that interviews were a dominant aspect of the methodology, but fails to actually refer to the interviews in its analysis, thus not providing any transparency regarding which part of the analysis was based on interviews. While such studies have an important role in a policy debate on coherence between policies that lacks a strong evidence base, future studies undertaken by non-

²⁸ While analysing these policy debates is not within the remit of this study, an overview is provided in Keijzer 2010 and in Keijzer et al 2012.

governmental organisations could gain in validity and reliability by further investing in providing a transparent description of the methodology used.

The studies adopting a predominantly quantitative approach tend to have more rigorous analysis of the findings. However, these studies often neglect possibly important dimensions of the subject matter at hand because of restrictive assumptions needed to make the quantitative models work in an elegant manner. Some of these studies, such as Gjertsen (2004) and OECD (2008) do not reflect on such limitations and do not provide a frank and clear discussion of the assumptions that were made. Others are more transparent and display a certain level of self-criticism. Van Zeijts et al. (2011), for instance, use a quantitative technique to investigate the trade-offs between increased agricultural production and biodiversity conservation, but admit that '*potentially important factors are excluded because they do not fit the models*', including the use of crop rotation and payments to small farmers. Similarly, Van Beers et al. (2002), who investigate the environmental impacts of several indirect subsidies, warn that their method is not well-suited in very complex policy environments, with many different types of policies interacting.

Another issue that emerged from the reviewed studies when it comes to the quality of their analysis, is that many studies lack a proper baseline or counterfactual. In eight studies, there was neither a baseline nor a counterfactual. In a number of the remaining studies, there was only a limited descriptive presentation of a baseline, against which the measurement of progress is very difficult. Some of the environmental studies seeking to measure trade-offs did use counterfactuals in determining, for instance, the effect of conservation efforts on poverty levels. Especially for the studies that adopted a 'horizontal' perspective on coherence, which include all studies related to development cooperation and security, the use of a counterfactual is very complicated in view of the subject-matter and the available data. Partly for that reason, Wilms et al. (2012) admit that it is very difficult to assess whether increased coherence contributed to outcomes. Other studies also place question marks around the ability to establish causality between coherence as observed at the process or output level and outcomes. IOB (2008) finds that attribution is too difficult, since there are too many different factors contributing to observed outcomes. The Norwegian Ministry of Foreign Affairs (2004) also find severe problems in establishing causality, and more radically proposes that impact assessments on coherence should be dropped altogether, and that evaluations should instead not measure beyond the level of outputs.

Finally, in many of the studies, the (horizontal) objectives are not defined very clearly. Often the objectives remain very broad and unspecified, such as 'increased development' or 'poverty reduction'. However, such goals can be very ambiguous. For instance, there could be increased development if this is defined in terms of Gross National Income, but perhaps this is achieved at the cost of rising inequalities. If poverty reduction is the goal, it should be clearly specified whether that is taken to be achieved only when a certain income threshold is reached, or even when poverty is reduced only marginally. Such ambiguity about policy objectives can undermine the quality of the analysis, since it effectively obscures the criteria for coherence. If it is not completely clear what different policies are seeking to achieve, then it will prove impossible to determine whether and to what extent they are coherent with each other.

4.3. Other findings and observations from the studies

Although the purpose of the analysis of studies was to gather evidence in relation to the second and third study questions, i.e. on the extent to which coherence can be evaluated and the merits of different methods, some of the studies also provided interesting conceptual and theoretical approaches on the subjects they covered which can reinforce or alternatively interrogate the theoretical analysis presented in chapter 2. This section will present some of these studies individually and thus presents additional ideas that feed into the study's overall conclusions as set out in chapter 5.

4.3.1. Selected findings from the studies

Frequent lack of intervention logic for promoting coherence

In many cases, promoting coherence or integrated policy-making is simply assumed to be beneficial. An intervention logic, or policy theory, is often lacking. Some studies highlight that such an intervention logic is lacking in the policy programmes they evaluate, while other studies are themselves failing to provide it. This lack of policy theory is particularly apparent with respect to coherence approached from a horizontal perspective. Wilms et al. (2012) evaluate a policy programme on promoting coherence in relation to the horizontal objective of biodiversity, and find that the policy does not have a strong intervention logic. It is not made clear, neither by the policy documents nor by the policy actors, how increased coherence and inter-ministerial and inter-departmental cooperation are actually supposed to translate into improved outcomes.

A similar point is made by Patrick et al. (2007), who assess whole-of-government approaches to fragile states in seven donor countries. They find that in most donor countries, there is no consensus on the rationale for a whole-of-government approach. There is substantial commitment to promoting coherence, but it is not made clear what problem coherence could actually solve, and the authors conclude that donors are essentially 'flying blind' when it comes to coherence and fragile states.

The studies on the environment that depart from a trade-off perspective are somewhat less exposed to this problem. They tend to investigate the linkages between two factors, such as biodiversity and poverty, and on that basis advocate integrated policy-making so that trade-offs can be minimized and win-win situations exploited. In many of these studies, however, the policy theory remains rather shallow. Dasgupta et al. (2005), for instance, measure to what extent environmental and poverty problems overlap geographically in South-East Asia. On the basis of that exercise, they advocate integrated policies dealing jointly with both the environment and poverty in those areas where the spatial overlap is strong. As such, their rationale for coherent policies is solely based on the fact that two policy objectives are promoted in the same geographical area. There is not much of a theoretical analysis to demonstrate why integrated policies would be more effective than different policies treating the environment and poverty separately.

Intervention logic and solid diagnosis of underlying problems are crucial

Related to the point that an intervention logic is often lacking, a number of studies show that a solid diagnosis of the underlying problem is crucial. If this diagnosis is incorrect or incomplete, interventions aimed at promoting coherence are seriously undermined. Fishtein et al. (2012) argue that the whole-of-government approach to fragile states is predicated on the assumption that conflict is primarily driven by economic factors. At the heart of the approach lies the nexus between security and development, stating that insecurity undermines development, while a lack of development prevents security. This gives rise to a type of vicious circle, and a whole-of-government approach is then theorized to be able to break that vicious circle. Fishtein et al. (2012), however, argue that political factors are much more important than economic factors in conflict dynamics in Afghanistan. If looking

through the lens of that theory, the whole-of-government approach appears to lose some of its appeal. The authors even hold that aid can undermine security by reinforcing uneven and oppressing power relationships and by providing additional valuable resources over which to fight. Whether one accepts their theory of the underlying drivers of the conflict or not, it is clear that, for both designing and reviewing a whole-of-government approach, it is of critical importance to make explicit which theoretical perspective one adopts.

In similar vein, McElwee (2012) argues that policies around Payments for Environmental Services (PES) are only effective if the underlying problem is the presence of an externality. The externality would be that the market does not reward the sustainable management of environmental resources, while there is a definite social benefit. If that is the underlying and dominating problem, then PES can offer a solution. However, McElwee (2012) analyses many other potential drivers of deforestation or environmental degradation, such as political dynamics, incomplete property rights, capacity constraints, or lack of access to capital markets. Based on the analysis of these drivers and their origins, she argues that PES can do little to address them. Thus, if deforestation is driven by such factors, then PES cannot be expected to resolve it.

Better coherence at process- or output-level does not guarantee results at outcome-level

Changes in mechanisms, processes or policies are ultimately aimed at improving performance at the outcome-level. Several of the reviewed studies emphasize that this link from process or policy to outcome is by no means guaranteed. Wilms et al. (2012) illustrate that an optimal institutional set-up for ensuring coherence does not necessarily lead to improved coherence at outcome-level, in their case due to a lack of political will and performance-related targets. Fair Politics (2010), in a study looking at the impacts of EU policies in Ghana, finds that policies that seem more coherent on paper may fail to deliver actual results on the ground, since this is dependent on many other factors. While this is a seemingly obvious point, explicitly signalling it serves to reinforce the importance of and need for assessments of coherence at outcome-level and, in relation to the previous remarks, of interrogating the intervention logic.

One-size-fits-all approaches to promoting horizontal objectives are inappropriate

Policy outcomes are highly context-specific. As such, it may not be so surprising that many of the studies stress that the interplay between different policies and the relationships between different policy objectives are also highly context-specific. Van Zeijts et al. (2011) investigate the integration of biodiversity objectives into European agricultural policy and find that the outcomes display significant differences across regions. The introduction of a mandatory ecological set-aside of 5% of farmlands plays out very differently in different regions, depending on whether the region has intensive or extensive farming systems. Such a measure could significantly increase biodiversity in intensively farmed regions, but would do little in extensively farmed regions. In the latter regions it could be more appropriate to use agricultural policies to encourage the conservation of the current biodiversity level, rather than to aim to increase it through ecological set-asides.

Similarly, Alliance Environnement (2007) stresses that agri-environmental measures should be fine-tuned to local circumstances. They show, for instance, that conditions on stocking densities attached to extensification payments should not be set at the same threshold level across the EU, because that neglects the specific environmental and farming conditions in different regions. The EU-wide threshold was set at too high a level for the situation in several Spanish regions, thus generating perverse incentives and the actual encouragement of overstocking, ultimately resulting in soil erosion and water pollution.

Schut et al. (2010) assess the environmental and socio-economic effects of biofuel projects in

Mozambique and how certification schemes can be used in reducing negative effects, thus heightening coherence. The biofuel sector is highly heterogeneous, with small-scale projects on the one hand and large-scale commercial projects on the other, and the study finds that any certification scheme has to take account of this heterogeneity. Failing to do so risks excluding smallholders from access to the biofuels market. Solutions for promoting coherence on the ground must be grounded in local realities.

Political economy dynamics exert a major influence

Building further upon the last point, a number of studies stress the importance of political economy dynamics in particular. Promoting coherence should not be seen as a technical issue to be resolved, but rather as a political project. McElwee (2012) assesses whether payments for environmental services (PES) can be a win-win for both conservation and development objectives. She finds that this highly depends on political realities, both at a local and national level. At the local level, the effect of PES are dependent on the ability of the PES programme to be flexible and adapt to the local political settlement, which is determined by factors such as the presence of powerful landholders and tenure security arrangements. In addressing PES programmes, one must also take account of national political economy dynamics and their history. For Vietnam, McElwee shows that the operation of PES programmes at a local level must be considered within the broader context of relations between the state and market in that particular economy.

Sandker et al. (2009) used a participatory modelling approach to explore the trade-offs between conservation and development in a region of Cameroon, and in particular tried to assess the effectiveness of integrated interventions dealing with both conservation and development objectives. They find that local economic and environment governance is of prime importance. In their model, the only strategy that can reconcile conservation and development objectives in the long-term is one where local governance of forest taxes and wildlife royalties takes center stage.

Coherence often involves non-linear relationships

Linked to the observations made on the intervention logics as described or implied in the studies reviewed, some studies noted that coherence is not a 'simple' matter, partially because coherence is not often associated with straightforward, linear relationships. Regarding other evaluation criteria such as efficiency or effectiveness, in most cases one could say 'the more the better' (apart from a certain need to experiment and even make mistakes along the way that one can learn from). When it comes to coherence, however, there are more situations in which more coherence is not necessarily better. Since absolute coherence between all public policies in a complex political system with many different interest groups influencing those policies is simply not possible, and probably not desirable, it follows that more coherence is not always necessarily better. For instance in a situation of political volatility, the main challenge is not to find the technically optimal set of policies in the sense of maximised coherence, but rather to find a set of policies that preserves the political settlement. If heightening coherence comes at the cost of alienating powerful groups in society, the political settlement could be endangered (see Khan 2010).

Furthermore, non-linearities can also characterize the relationship between trade-offs and synergies. Ferraro et al. (2011) investigate trade-offs and synergies between deforestation and poverty alleviation, and find that the outcome really depends on specific area characteristics, thus also highlighting the importance of context-specificity discussed above. Interestingly, looking at the effects of forest protection, they find that area characteristics associated with the most avoided deforestation are also those associated with the least poverty alleviation. This suggests that "(...) *win-win efforts to protect ecosystems and alleviate poverty may be possible when policymakers are satisfied with low levels of each outcome, but tradeoffs exist when more of either outcome is desired*".

Promoting coherence in one sense may undermine coherence in another

Promoting coherence should not be seen as a one-dimensional and 'singular' challenge. One cannot establish that coherence between a set of policies or objectives has been increased and then automatically conclude that overall coherence has been taken to a higher level or assume that this has any direct implication in terms of achieving outcomes. One additional dimension that has to be taken into account is the time dimension. Schut et al. (2010), investigating certification schemes in making biofuel policies more coherent with environmental and socio-economic objectives warn that increasing short-term coherence through more strict certification standards can undermine the long-term viability of smallholder actors in the biofuel sector, if they are not helped in meeting the high standards. This could then negatively impact both environmental and socio-economic objectives in the long term. As such, short-term coherence can undermine long-term coherence.

Alternatively, promoting coherence between several policies or objectives could undermine coherence in relation to third objectives. For instance, Patrick et al. (2007) warn that increasing coherence between security and development objectives through whole-of-government approaches may actually undermine other objectives such as local ownership and the harmonization of policies across different donors. This can happen because promoting coherence between the policies of one donor tends to reinforce top-down strategic decision-making directed from donor capitals. In the field, this can have the effect of reducing the flexibility to respond to demands made by local stakeholders or other donors.

4.3.2. Implications of the findings for the analysis in chapters 1 and 2

The findings presented above show a number of relevant contrasts when compared with the conceptual framework for this study as presented in chapter 2.

Firstly, the conceptual framework developed for this study seeks to align itself to existing evaluation concepts and guidelines, as well as to an overall results chain approach. This reflects the focus of the wider policy evaluation community to assess to what extent the intended results of public or private policies have been realised. Efforts to evaluate coherence in the field of international cooperation will often be confronted by a lack of explicit objectives and an absence of stated intentions in relation to how different policy objectives should interrelate, at least beyond recommendations on the process for reaching such decisions. This is why the majority of the studies assessed either used a 'side-effect' or 'trade-off' perspective that reasons back from the results as observed, and fewer studies used a more 'theory-based' perspective that starts from the policies as defined (i.e. a 'horizontal objective' perspective).²⁹

As a second observation, the conceptual framework emphasised the need to distinguish between gross and net outcomes and the possibility to promote coherence at five different levels of governance. The studies analysed here however leave the question of how to attribute, contribute or otherwise associate effects to the intervention assessed largely unanswered, and a few exceptions notwithstanding focus on relations between two actions or policies (of which many at a micro level). In that sense the level of ambition adopted by the different studies is lower than what the conceptual framework implies, and where a higher level of ambition was applied this often resulted in studies with a descriptive level of analysis.

²⁹ A related observation is that by the conceptual framework chosen, it was implicitly assumed that the concept of coherence can be 'accommodated' in the existing evaluation criteria – criteria which typically apply to evaluations that often evaluate the effects of interventions carried out in a single policy area. A wider debate that this paper does not specifically look into is whether a more globalised and more direct interactions between different policies and different levels of governance might not merit a more fundamental look at how the practice of evaluation itself is conceived and operationalised.

A third observation relates to the association of coherence with the existing evaluation criteria. The theoretical discussion hypothesized that coherence is most closely related to effectiveness and efficiency. This was to a large extent reflected in the analysed studies in the sense that they stressed that policies have an influence on whether, how and at what costs other policies succeed in meeting their own objectives. This was most obviously the case in studies with a side-effect perspective on coherence. The theoretical discussion also made a distinction between transformation and optimization efficiency. In relation to that distinction, it can be remarked that the side-effect perspective is mostly associated with transformation efficiency, for its analysis has implications for the translation of inputs into results in the field in which the side-effect occurs. The trade-off perspective, on the other hand, relates more to optimization efficiency in the sense of looking for the policies that most efficiently realize a range of diverse objectives. Finally, the horizontal objective perspective could be said to represent somewhat of a hybrid in this respect. On the one hand, it assesses the translation of inputs into results, for instance in security studies that investigate whether development aid can enhance the realization of security objectives. On the other hand, it raises the question as to what set of policies is most efficient in realizing a certain horizontal objective. Which of the perspectives proves most useful for further investments in evaluating coherence in international cooperation will depend on the progress made at the level of policy debates, particularly on the extent to which explicit objectives for coherence are defined that go beyond process and that specify desired results.

5. Analysis of findings, conclusions and recommendations

5.1. Introduction

This section analyses the findings in relation to the four research questions that were presented in section 4. On this basis the overall research question as presented in section 1 is answered and recommendations are made with particular focus on the international cooperation evaluation community.

5.2. Analysis of findings in relation to the four research questions

1) In what ways and to what extent can ‘coherence’ be defined and operationalised for evaluation purposes?

There is presently no widely accepted definition of coherence in the development evaluation community. This is related to the absence of a well-developed practice in this area, as well as by differences in overall evaluation policies and definitions of key concepts between OECD members. A conceptual exploration made by this study shows that coherence is a multi-layered concept that is not easily captured in its full complexity by a single definition. Recent political discussions on ‘Policy Coherence for Development’ in the EU, OECD and UN nonetheless provide some direction by clarifying that the focus should be on the contribution of different public policies to development outcomes, but at the same time give rise to a multitude of conceptual and methodological questions including the definition of the desired development outcomes.

Based on an analysis of 22 studies, three ‘schools of thought’ are identified that guide efforts to investigate coherence:

1. The **side-effects perspective**: this perspective refers to a general felt need to ensure that a policy does not undermine the achievement of the objectives of other policies. In terms of the operationalisation of coherence for evaluation purposes, from this perspective it would actually seem that the evaluation of coherence may not be fundamentally different from regular evaluation of one single policy. Essentially, one is just measuring the effect of A on B, which in this case is the effect of policy X on variable B, which is a variable related to the objective of policy Y.
2. The **horizontal objective perspective**: here coherence is defined as the extent to which policies alone or together contribute to the achievement of one horizontal objective. The focus is then on how policies relate to each other. Do certain policies contradict each other when it comes to that horizontal objective; can certain policies reinforce each other with respect to the horizontal objective? For evaluation purposes, this type of definition is certainly more challenging compared to a side-effect perspective. One would have to analyse the inter-linkages and interdependencies between several different policies. Thus, whereas the side-effect perspective evaluates the effect of policy X on variable B, the horizontal perspective minimally demands the evaluation of how policy X influences the effect that policy Y itself has on variables A and B.
3. The **trade-off perspective**: this approach places the focus not only on different policies, but also on different objectives. The main question asked by studies approaching coherence in this way is whether and to what extent different objectives are compatible with one another. Among the studies reviewed for this paper, the studies in this category all deal with the relation between environmental and developmental objectives, or specific sub-sets such as biodiversity

conservation and poverty reduction. In terms of evaluating policies on coherence, this approach implies investigating how a policy or combination of policies affects the relationship between objectives. Thus, it investigates to what extent policies X and Y influence each other in terms of the achievement of their respective objective. Do policies succeed in reconciling two diverse objectives ('win-win'), or in alleviating the trade-off, or do they not alleviate the trade-off at all?

Another finding that emerged from the reviewed studies is that in order to operationalise coherence for evaluation purposes, the most important thing is to clearly define all the terms in the equation. Thus, it is not only crucial to define coherence itself, but also to define the different policies and especially their objectives. Leaving any ambiguity in this respect can reduce the possibility to properly assess coherence.

2) To what extent can the relation between coherence and effectiveness/efficiency be evaluated (i.e. is coherence additional or complementary)?

This second study question examined the relation of coherence to other evaluation criteria, and sought to find out whether coherent policy making actually leads to better outcomes. Is it actually worthwhile to invest time and resources in terms of establishing linkages between different policy fields and stimulating policy integration, or is it in fact more efficient to have each policy exclusively focus on its own realm?

Based on the studies reviewed, this study could not give a satisfactory answer to this question. One specific effort was to look into the overall 'level of analysis' of each study, based on Palenberg 2011. The highest level of analysis, level 2 analysis, would correspond to relating observed outcomes to coherent (or incoherent) policies, and being able to say what the outcomes would have been in the absence (or presence) of coherent policies. Thus, to qualify as having 'level 2 analysis', a study would have to be able to really indicate the cost of incoherence or otherwise describe the value-added of coherence. It was found that none of the reviewed studies was effectively able to reach such conclusions, so no studies were classified as 'level 2 analysis'. A total of 13 studies were classified as 'level 1 analysis', while the other nine were classified as 'descriptive analysis'.

The absence of level 2 analysis in relation to coherence among the reviewed studies does not imply that this level of analysis is not in reach for evaluations. It does indicate, however, that it is not straightforward. To properly evaluate the relation between coherence on the one hand and effectiveness and efficiency on the other hand, it is required to have extensive data on many policies and variables combined with a valid and reliable research design. Furthermore, ideally one would have access to such data through time and across space, in order to be able to use baselines and counterfactuals.

3) What methods have been used in past studies and evaluations that look into coherence inside or between policies, at what levels (micro, meso, macro), and what are their respective strengths and weaknesses?

A variety of methods were used by the 22 studies that were reviewed. Most frequently used were interviews, document analysis and descriptive statistics. A separate body of studies principally relied on a quantitative approach, using modelling techniques and regression analysis.

The studies were relatively evenly divided over the micro-, meso- and macro-levels. Further, it was found that the studies each used two methods on average. Few studies used more than three different methods. Finally, there were few cases of studies that combined quantitative and qualitative

approaches; studies were either predominantly quantitative or qualitative in their analysis.

Regarding strengths and weaknesses, it should be noted that it is difficult to come to definite conclusions about the respective strengths and weaknesses of different methods. First of all, few if any studies offered much reflection on the benefits and limitations of their methodological approaches. Secondly, how useful a certain method appeared to be in a certain study is obviously closely related to the way in which the method was applied. If some studies were not able to use interviews effectively and draw conclusions on coherence based on interviews, this is not necessarily a limitation intrinsic to the 'interview' method as such. It may just be that these studies implemented the method poorly, for instance that they interviewed the wrong people, not enough people, or did not ask the right questions.

Bearing this in mind, it does appear that studies with a qualitative, interview-based approach generally were able to present a wealth of information, but were not very good at systematically analysing and presenting those data. Conversely, studies using modelling and regression techniques displayed more analytical rigor, but were seen to be constrained in terms of the assumptions required for these models, so that possibly important aspects of the equation were not considered at all. In other words, qualitative studies seemed to prioritise validity at the expense of reliability, while quantitative studies seemed to prioritise reliability at the expense of validity.

4) Based on the answers to the first three questions, what practical and methodological dilemma's can be observed with regard to improve the evaluation of policy coherence in the specific field of policies on international cooperation?

One major dilemma for improving the evaluation of coherence in international development is to set the appropriate level of ambition and determine appropriate levels of investment in relation to that ambition. All policies affect the behaviour of societal actors in some way, which in turn always has certain knock-on effects. However, it is simply not feasible to evaluate all inter-linkages, even including the smallest ones, between all policies. One inevitably has to adopt a certain focus and draw a line.

A large number of the studies reviewed adopt a trade-off perspective and as such examine the relations between different policy objectives. If adopting this perspective in evaluating coherence in international development, evaluations would need to look not only at effects of policies in developing countries, but also consider the effect and purpose of donor country policies in their own domestic economy and society. This could then be combined with the analysis of the effects of these same policies on developing countries. Only then could the trade-off between the different policy objectives (domestic and international development) really be evaluated.

However, one could also decide to take the horizontal objective of international development as a starting point and then evaluate the coherence of policies within that framework. This corresponds to the horizontal perspective discussed in this paper. Although this might seem to represent a lower level of ambition, it is certainly not inferior to seeking to evaluate the trade-offs between objectives. On a practical level, it could be a superior approach, since it can bring to light policies that are at odds with each other. When such contradicting policies actually share the same overall objective, it is clear that the incoherence needs to be resolved. If the contradicting policies have very different objectives, it is important to properly evaluate their respective impacts as a basis for trade-off decision at the political level.

5.3. Conclusion and recommendations

The findings of this study as presented in relation to its four main question indicate that the evaluation of coherence is still in an early and nascent stage, especially when compared with 'main stream' development evaluation in relation to other criteria such as effectiveness, efficiency and impact. This is not surprising considering the limited amount of investment that has been made to evaluate coherence in international cooperation. While virtually every aid intervention today is evaluated, sometimes with substantial investment such as in the case of single modalities like budget support, this paper has shown that the number of serious evidence-based studies of coherence remains very limited. This observation is also made in more policy-oriented studies, which note that the level of ambition in terms of promoting coherence as expressed at the political level is not accompanied by levels of investment in financial and human resources that one could expect such an ambition to legitimate (Keijzer et al 2012).

The findings of the study provide no basis to conclude that rigorous evaluation of coherence is not possible. A start has certainly been made in evaluating the impact of donor country 'non-aid' policies on developing countries, and the effect of globalisation in terms of blurring the lines between internal and external policies as well as political pressures to allocate part of the ODA budget to line ministries may motivate further efforts. In taking on such an endeavour, clues can be taken from how coherence has been approached in other policy fields, most notably the one related to environmental issues. Some reflections are put forwarded relating to this overall recommendation:

- One important caveat is that most ex-post assessments of coherence discussed in this paper focus on the micro-level. Coherence at macro-level, which would be of more relevance to the field of international development cooperation in view of what assessments are called for in current policy discussions, is especially under-researched.
- When it comes to coherence at macro-level, particularly in the environment field there does seem to be more experience with ex-ante studies. These tend to have a strong emphasis on the use of quantitative models using secondary data. Such secondary data will often be available in developing countries but are often not sufficiently systematised and available for direct use from a distance. Many EU member states do not have a strong focus on doing such detailed ex-ante assessments of proposed new policies or reforms of existing policies. The EU does have this practice, but although overall guidelines require such assessments to look into the potential effects in developing countries this requirement is rarely met, with lack of data being an important reason.
- Another issue is that the trade-off perspective encountered in many of the environmental studies examined for this paper has not really been considered in the field of international development, where there has been a more implicit preference for horizontal objectives (or by default for side-effects where these horizontal objectives are insufficiently specified).

This study thus offers a basis to **conclude** that coherence *can* be evaluated in the field of international cooperation, predominantly based on the evidence that coherence *is* evaluated in other policy areas. In line with this conclusion and in view of existing political commitments to take this further, this study puts forward four recommendations on how this could be done.

Recommendation 1 – Manage expectations and identify feasible steps forward: in view of the limited investments and evidence base in the field of international cooperation, the study's findings call for further 'management of expectations' in this area and for a more focused discussion on how the increasing call for evaluating coherence can be translated into a feasible path to further developing this emerging field of work. It would not be realistic to assume that evaluations of entire whole-of-government approaches to development will soon approach the quality and reliability of current

evaluations of the impact of single development cooperation interventions. Similarly, the use of evaluations to measure the trade-offs between domestic and international development objectives at macro-level and from there determine whether policies deliver win-win situations or affect the trade-offs does not appear a realistic prospect for the time being.

In other words, one should learn how to walk before attempting to run, i.e. first investing in exploratory and pilot research (see also recommendation 2) in a way that seeks to ensure methodological quality and diversity to give a solid base for comparing the assessed interventions to alternatives and benchmarks (i.e. level 2 analysis). Such investment is needed to further establish the influence of coherence on development outcomes and possibly allow for better approximating the costs of incoherence.

Recommendation 2 – Support and invest in pilot studies to reduce data and methodological deficits: one element of an incremental approach could be to undertake more exploratory studies from a ‘side-effect’ perspective, which could gain further insights on whether assumptions on (in)coherence can be confirmed at the level of development outcomes, including by support discussions in the DAC to undertake pilot studies in this field.³⁰ In such studies, it will be important to take account of some of the substance-related findings in this paper. As such, they will need to consider country heterogeneity, as well as heterogeneity within countries. The impacts of policies will vary largely between different groups of developing countries. Similarly, local political economy dynamics in developing countries are likely to have a large influence on the impact that policies will have on different groups within those countries.

Recommendation 3 – Explore joint action while seeking flexibility in dealing with mandate limitations: in order to stimulate such evaluations, it will be important to provide evaluation offices of development ministries the flexibility to invest in them. In many EU member states, development evaluation units have retained a rather narrow mandate up until today, which means that they can only really invest in evaluations of the aid policies implemented by their own ministry. In view of international commitments to take further the evaluation of coherence (e.g. Busan 2011) it seems important to look for creative ways to agree to undertake such efforts collectively at the DAC or EU, despite the fact that not all members’ individual mandates would allow for this. Should this not result in the desired unanimity or critical mass not be achieved, then this may nonetheless give a more legitimate basis for a ‘coalition of the willing’ to act as a first mover.

Recommendation 4 – Challenge the international community of evaluation experts: although evaluating coherence requires a strong political mandate, also to make intended coherence results more explicit as a basis for improved intervention logics, a stronger base for giving shape to such a mandate can be created through involving the community of independent evaluation specialists. Government evaluation offices should be pro-active in inviting such specialists to participate in carrying out evaluations. Similar efforts were made in making the shift from project evaluations to evaluations of programme-based approaches, including through discussion papers and specialist workshops. Such efforts may be needed again now that a stronger focus is emerging on evaluating development cooperation in relation to other policies.

³⁰ For more information on these pilots, see:

[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DCD/DAC\(2011\)32&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DCD/DAC(2011)32&docLanguage=En)

Bibliography

- Den Hertog, L. and Stroß, S. (2011). *Policy Coherence in the EU System. Concepts and Legal Rooting of an Ambitious Term*. Conference Paper, 'The EU as a Global Player', Madrid, 7-8 April 2011
- ECDPM and ICEI (2005) EU mechanisms that promote policy coherence for development – a scoping study. Studies in European Development Co-operation Evaluation No 2. Amsterdam: Aksant Academic Publishers
- ECDPM (2008) Evaluating Co-ordination, Complementarity and Coherence in EU development policy: a synthesis. Studies in European Development Co-operation Evaluation No 8. Amsterdam: Aksant Academic Publishers
- IOB (2009). *Evaluation Policy and Guidelines*. Policy and Operations Evaluation Department, Dutch Ministry of Foreign Affairs.
- De Lange, P. et al (2011) Evaluation of the Dutch support to capacity development: Facilitating resourcefulness. The Hague: Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs (IOB).
- Keijzer, N. (2010) Policy Coherence for Development: from moving the goalposts to result-based management? Maastricht: ECDPM Discussion Paper 101
- Keijzer, N., King, M., Spierings, E., Matthews, A. (2012) Measuring Policy Coherence for Development. Study commissioned by The Netherlands Ministry of Foreign Affairs and the German Federal Ministry for Economic Cooperation and Development. Maastricht: ECDPM
- Khan, M. (2010). 'Political Settlements and the Governance of Growth-Enhancing Institutions'. SOAS Working Paper. London: SOAS
- Mackie, J. et al (2007). Evaluation Study on the EU Institutions & member States' mechanisms for Promoting Policy Coherence for Development. Studies in European Development Co-operation Evaluation 7. Amsterdam: Aksant Academic Publishers. Available at <http://www.three-cs.net>
- Ministerie van Financiën (2006) Regeling van de Minister van Financiën van 18 april 2006 inzake het periodieke evaluatieonderzoek en de beleidsinformatie van de rijksoverheid (Regeling periodiek evaluatieonderzoek en beleids-informatie 2006)
- OECD (2002). Glossary of Key Terms in Evaluation and Results Based Management.
- OECD (2009a). *Approaches to Assessing Policy Coherence for Development: A Summary of National and Expert Views. Meeting of the National Focal Points for Policy Coherence for Development*. SG/PCD(2009)3. Paris: OECD
- OECD (2009b). *Framework and Assessment Methodology for Policy Coherence for Development*. SG/PCD(2009)4. Paris: OECD
- Palenberg, M. (2011). *Tools and Methods for Evaluating the Efficiency of Development Interventions*. BMZ Evaluation Division: Evaluation Working Papers.
- Picciotto, R. (2005). 'The Evaluation of Policy Coherence for Development'. *Evaluation*, Vol. 11(3): 311-330
- Ugla, F. (2007) Mainstreaming at Sida: a synthesis report. Sida Studies in Evaluation 2007:05. Stockholm: Swedish International Development Agency

Annex 1: definition of research methods listed in table 3³¹

Cost-Benefit Analysis

Expresses all costs and benefits of the evaluated intervention in monetary terms. While attaching a money-value to costs is usually relatively straightforward, the monetization of benefits may present big challenges. Benefits are often approximated by 'willingness to pay', but this is problematic when there are no perfect markets, or when there is no market at all for the benefit.

Cost-Effectiveness Analysis

Measures the outcome an in intervention in its natural units, and compares it with the intervention's costs. As such, it is similar to Cost-Benefit Analysis, but it circumvents the challenge of monetizing benefits. This method is less applicable to interventions with multiple outcomes.

Descriptive Statistics

Using available data and indicators to track changes on the outcome-level, and investigate whether any observed changes may be related to the evaluated intervention. Attribution is usually to hard to establish with certainty.

Direct Measurement

Quantifiable effects may be simply measured directly. However, in many cases this could be too complex, for instance when the effect is not easily observable or when it manifests itself in multiple locations. Then, one is forced to rely on secondary data.

Direct Observation

Direct observation techniques allow for a systematic and structured process of observing what is actually taking place in the 'field', where the impacts are assessed. Observations can relate to physical outputs as well as to certain processes.

Document Analysis

Systematically reviewing the available written sources on the subject matter, including policy-related documents and academic articles.

Econometric Analysis

Applying mathematical and statistical tools to economic data in order to test whether a theorized relationship of cause and effect has any empirical basis.

Group Interview

Similar to normal interviews, but in a group interview the interaction between the interviewees may yield additional information. Group interviews could be useful to get a clear view of the differential impact the evaluated issue may have on different stakeholders, as well as to identify points of controversy.

Interview

Used to collect analyses, opinions, points of view and suggestions, which may not be easily obtained through other methods. The main difference with surveys is that interviews are not standardized, and are therefore more qualitative. The advantage is that data come directly from the cultural, political

³¹ The definitions are based on IOB (2009), Palenberg (2011), and the Evaluation website of Europe Aid (http://ec.europa.eu/europeaid/evaluation/methodology/tools/too_en.htm)

and/or institutional context, while the disadvantage is that different interviews do not lend themselves to easy comparisons. Three different types of interview may be distinguished:

1. **Unstructured interview:** there is no predetermined set of questions, so the interviewee has complete freedom to bring up issues the evaluator had not considered before.
2. **Semi-structured interview:** The evaluator has a rough interview guide, but will generate new questions depending on the points brought up by the interviewee.
3. **Structured Interview:** The evaluator strictly follows a set of predetermined questions. Structured interviews generate data of a more quantitative type.

Quantitative Modelling

Quantitative models are an analytical tool to quantify relationships between different processes and variables. Econometric modelling is also a form of quantitative modelling, but quantitative modelling does not necessarily involve economic variables. Furthermore, quantitative modelling may rely on non-stochastic processes, whereas econometric models are mostly stochastic.

Statistical Survey

In a survey, a large number of people are asked to fill in the same predetermined questionnaire. The questionnaire can be devised so as to generate standardized data that are suitable for subsequent statistical analysis.

Annex 2: Search terms used and studies analysed

2.1 Search terms used

The databases IngentaConnect and ScienceDirect were used to search for relevant studies, as these included a larger number of relevant journals compared to JSTOR.

Further to the initial results from the evaluations as assessed, which focused mostly on the ‘promoting synergies’ aspect of coherence, it was agreed to use search terms that explicitly reflect the conflicts or trade-offs between policies in relation to horizontal objectives. The following key search terms were chosen (see 2.3 for the results of an investigation of a possible bias in these terms for particular methods/disciplines):

- biodiversity AND impact AND poverty AND trade-off
- environment AND impact AND development AND economic*
- forestry AND impact AND sustain* AND poverty

2.2 Search results

The search results of different combinations of the above search terms were as follows:

EconLit

Biodiversity AND impact AND poverty	39 results
Biodiversity AND impact AND trade-off	23 results
Biodiversity AND impact AND economic development	264 results
Environment AND sustain* AND impact	1529 results
Agriculture AND trade-off AND impact	67 results
Cattle AND poverty AND impact	3 results
Deforestation AND impact AND agriculture	154 results
Fish* AND poverty AND impact:	43 results

Science Direct

Biodiversity AND poverty AND trade-off	349 results
Biodiversity AND poverty AND tradeoff above).	309 results (mostly different from the ones above).
Biodiversity AND economic development AND trade-off	399 results
Biodiversity AND impact AND poverty	3638 results
Forest* AND impact AND poverty AND trade-off	552 results
Forest* AND impact AND environment AND tradeoff	1692 results
Fisheries AND trade-off	23 results

Web of Science

Biodiversity AND impact AND trade-off	38 results
Biodiversity AND poverty AND trade-off	3 results
Biodiversity AND impact AND economic development	85 results
Forest AND impact AND trade-off	83 results
Fisheries AND impact AND trade-off	17 results

The results identified concerned articles that appeared in the following journals:

1. World Development
2. Agricultural Systems
3. Journal of Environmental Economics and Management
4. Ecological Economics
5. Environmental Resource Economics

While the first two journals are multidisciplinary in focus, the last three have a focus on economics.

2.3 Alternative search terms used

When analysing those studies that met the four selection criteria, it was decided to do an additional search on terms other than 'trade-offs' and with a specific focus on security studies to look into a possible bias of the search terms used to select studies from a certain academic discipline – or using particular methods.

The search results presented below however were however considered as not sufficiently useful for this study as per the four criteria (see box 1 in 3.2) for one or more of the following three reasons: (1) there was insufficient information on methodology, (2) the analysis was exclusively focused on the macro level or (3) the research did not sufficiently base itself on the analysis of data but was primarily theory-based.

Science Direct

"conflicting policies" AND biodiversity	32 results
biodiversity AND development AND policy AND conflict	6065 results (in all text)
biodiversity AND development AND policy AND conflict	27 results (title, keyword, abstract)
biodiversity AND development AND synergies	2029 results
deforestation AND development AND policy AND conflict	2369 results

Web of Science

biodiversity AND development AND policy AND conflict	95 results (in 'topic')
deforestation AND development AND policy AND conflict	16 results
biodiversity AND development AND synergies	42 results

Annex 3: Studies and evaluations analysed

- ActionAid (2003). 'Policy (In)Coherence in European Union Support to Developing Countries: A Three Country Case Study'
- Alliance Environnement, (2007) Evaluation of the environmental impacts of Common Agricultural Policy measures related to the beef and veal sector and the milk sector.
- Brown, K. et al. (2001). 'Trade-off Analysis for Marine Protected Area Management'. *Ecological Economics*, Vol. 37: 417-434
- Dasgupta, S. et al. (2005). 'Where is the Poverty-Environment Nexus? Evidence from Cambodia, Lao PDR, and Vietnam'. *World Development*, Vol.33 (4): 617-638
- Fair Politics (2010). 'Ghana's Traders, Lumberjacks and Fortune Hunters'. Evert Vermeer Foundation: Amsterdam & Fair Politics (2012). 'The EU Raw Materials Policy and Mining in Rwanda'. Evert Vermeer Foundation: Amsterdam (N.B. these two studies had big similarities in scope and approach, hence they were assessed jointly)
- Ferraro, P. and M.Hanauer (2011). 'Protecting Ecosystems and Alleviating Poverty with Parks and Reserves: 'Win-Win' or Tradeoffs?'. *Environmental Resource Economics*, Vol. 48: 269-286
- Fishtein, P. and A.Wilder (2012). 'Winning Hearts and Minds? Examining the Relationship between Aid and Security in Afghanistan'. Feldstein International Center, Tufts University.
- Gjertsen, H. (2005). 'Can Habitat Protection Lead to Improvements in Human Well-being? Evidence from Marine Protected Areas in the Philippines'. *World Development*, Vol.33 (3): 199-217
- Hengsdijk, H. et al. (2007). 'Poverty and Biodiversity Trade-offs in Rural Development: A Case study for Puijiang County, China'. *Agricultural Systems*, Vol. 94: 851-61
- Illukpitiya, P. and J.F. Yanagida (2010). 'Farming vs Forests: Trade-Off Between Agriculture and the Extraction of Non-Timber Forest Products'. *Ecological Economics*, Vol.69: 1952-1963
- IOB (2008). 'Het Nederlandse Afrikabeleid 1998-2006. Evaluatie van de Bilaterale Samenwerking'. IOB Evaluatie, No.308.
- McElwee, P. (2012). 'Payments for Environmental Services as Neoliberal Market-Based Forest Conservation in Vietnam: Panacea or Problem?'. *Geoforum*, Vol. 43: 412-426
- Norwegian Ministry of Foreign Affairs (2004). 'Towards a Strategic Framework for Peacebuilding: Getting Their Act Together: Overview Report of the Joint Utstein Study of Peacebuilding'. Evaluation Report 1/2004
- OECD (2008). 'Reconciling Development and Environment Goals. Measuring the Impact of Policies'. The Development Dimension.
- Patrick, S. and K. Brown (2007). 'Greater than the Sum of its Parts? Assessing 'Whole of Government' Approaches to Fragile States'. New York: International Peace Academy
- Sandker, M. et al. (2009). 'Exploring the Effectiveness of Integrated Conservation and Development Interventions in a Central African Forest Landscape'. *Biodiversity and Conservation*, Vol. 18 (11): 2875-2892
- Schut, M. et al. (2010). 'Working Towards Sustainability: Learning Experiences for Sustainable Biofuel Strategies in Mozambique'. Wageningen University.
- Sims, K.R.E. (2010). 'Conservation and Development: Evidence from Thai Protected Areas'. *Journal of Environmental Economic and Management*, Vol. 60: 94-114
- Van Beers, C., et al. (2002). 'Environmental Impact of Indirect Subsidies. Development and Application of a Policy Oriented Method'. TU Delft and RIVM.
- Van Beijnum, M. and L. van de Goor (2006). 'The Netherlands and its Whole of Government Approach on Fragile States: Case Study Sudan.' Clingendael Institute: The Hague

Van Zeijts, H. et al. (2011). 'Greening the Common Agricultural Policy: Impacts on Farmland Biodiversity on an EU Scale'. PBL Netherlands Environmental Assessment Agency.

Wilms, P. et al. (2012). 'Evaluatie Beleidsprogramme 'Biodiversiteit Werkt''. Ape rapport nr. 994. Agentschap NL

Annex 4: Template used for analysing the studies

DATA COLLECTION GRID: study on the evaluation of coherence in international cooperation

Part 1: general information on the study and (for overall data grid)

Full reference of the study and URL if available online	Author (year), title, place of publication: publisher
Which actor (e.g. ministry) commissioned the study, and what policy areas does it cover?	
What is the objective of the evaluation, and how is coherence (or the horizontal objective) defined and positioned in that context?	
Does the evaluation address coherence in relation to process results, at the output level (i.e. policies) or outcome level (i.e. actual effects)?	
At what level(s) is/are effects measured by the evaluation?	E.g. effects on households, on specific actors, on specific geographic areas (e.g. large cities), on countries or regions, ...
Did the study involve collection of primary data, and how was the quality thereof judged?	
Did the study use existing secondary data, and how was the quality thereof judged?	
What methods for data collection and/or analysis have been used?	
Does the study include some kind of baseline measurement in relation to coherence/the horizontal objective?	
Any other remarks?	

Part 2: observations on validity and reliability³²

1	REMARKS ON VALIDITY (does the research meet its purpose)
1.1	Problem definition: Problem statement is clear and visible in the research questions
1.2	Object of study: well defined and delineated
1.3	Policy theory: intervention theory referred to and operationalised to some degree
1.4	Analysis: sources of information referred to, systematic linkages between findings, conclusions and recommendations
Remarks (strengths and/or weaknesses)	

2	REMARKS ON RELIABILITY (dependability of research results)
2.1	Research methods: specification of used methods and information on their perceived relevance. Methods used to verify/triangulate
2.2	Reach: representativeness of samples or case studies, reference to limitations of the research
2.3	Known independence: of both the used secondary sources and the authors of the study towards the institution who commissioned the study
2.4	Quality control: reflection on course of research, internal or external quality control
Remarks (strengths and/or weaknesses)	

Part 3: selected key findings and other remarks (for overall data grid)

SELECTED RELEVANT KEY FINDINGS IN RELATION TO COHERENCE OR APPLICABLE HORIZONTAL OBJECTIVES		
Findings (quote in italics, otherwise summarised)	Page nr.	Derived through what methods?
OTHER RELEVANT REMARKS IN RELATION TO THE DOCUMENT		
(optional to refer here to specific content considered useful for different purposes e.g. specific financial data on EU aid provided, useful graphs/conceptual ideas)		

³² Note: the third criterion of usefulness was not deemed relevant for this study.

About ECDPM

ECDPM was established in 1986 as an independent foundation to improve European cooperation with the group of African, Caribbean and Pacific countries (ACP). Its main goal today is to broker effective partnerships between the European Union and the developing world, especially Africa. ECDPM promotes inclusive forms of development and cooperates with public and private sector organisations to better manage international relations. It also supports the reform of policies and institutions in both Europe and the developing world. One of ECDPM's key strengths is its extensive network of relations in developing countries, including emerging economies. Among its partners are multilateral institutions, international centres of excellence and a broad range of state and non-state organisations.

Thematic priorities

ECDPM organises its work around four themes:

- Reconciling values and interests in the external action of the EU and other international players
- Promoting economic governance and trade for inclusive and sustainable growth
- Supporting societal dynamics of change related to democracy and governance in developing countries, particularly Africa
- Addressing food security as a global public good through information and support to regional integration, markets and agriculture

Approach

ECDPM is a "think and do tank". It links policies and practice using a mix of roles and methods. ECDPM organises and facilitates policy dialogues, provides tailor-made analysis and advice, participates in South-North networks and does policy-oriented research with partners from the South.

ECDPM also assists with the implementation of policies and has a strong track record in evaluating policy impact. ECDPM's activities are largely designed to support institutions in the developing world to define their own agendas. ECDPM brings a frank and independent perspective to its activities, entering partnerships with an open mind and a clear focus on results.

For more information please visit www.ecdpm.org

ECDPM Discussion Papers

ECDPM Discussion Papers present initial findings of work-in-progress at the Centre to facilitate meaningful and substantive exchange on key policy questions. The aim is to stimulate broader reflection and informed debate on EU external action, with a focus on relations with countries in the South.

This publication benefits from the generous support of ECDPM's core and institutional funders: The Netherlands, Belgium, Finland, Ireland, Luxemburg, Portugal, Sweden and Switzerland.

ISSN 1571-7577

European Centre for Development
Policy Management

ecdpm

HEAD OFFICE SIÈGE

Onze Lieve Vrouweplein 21
6211 HE Maastricht
The Netherlands *Pays Bas*
Tel +31 (0)43 350 29 00
Fax +31 (0)43 350 29 02

BRUSSELS OFFICE BUREAU DE BRUXELLES

Rue Archimède 5
1000 Brussels *Bruxelles*
Belgium *Belgique*
Tel +32 (0)2 237 43 10
Fax +32 (0)2 237 43 19

info@ecdpm.org
www.ecdpm.org
KvK 41077447



Printed on FSC certified paper.